RECURSIVE ESTIMATION OF K–DISTRIBUTION PARAMETERS

Pei–Jung Chung

Department of Electrical and
Computer Engineering
Carnegie Mellon University
Pittsburgh, PA, 15213
pjchung@ece.cmu.edu

William J.J. Roberts

Atlantic Coast Technologies Inc.
11499 Columbia Pike
Suite 200
Silver Spring, MD, 20904
robertsb@acoast.com

ABSTRACT

We address the problem of estimating parameters of K–distribution. A recursive procedure based on the recursive EM algorithm is derived to find the ML estimates. Recursive EM is a stochastic approximation procedure with a gain matrix derived from the augmented data. Under mild conditions estimates generated by such procedure are characterized by strong consistency and asymptotic normality. Because of the simple structure of the augmented data, the proposed algorithm has a simple implementation. Numerical results show that the proposed approach performs well for various parameter sets.

1. INTRODUCTION

The K-distribution [4] is commonly used for clutter modeling in synthetic aperture radar (SAR) applications [8]. The parameters of the K-distribution can be estimated by applying the maximum likelihood (ML) criterion using the expectation-maximization (EM) approach [6].

Recursive estimation of the parameters of the K-distribution is of practical importance in many radar related applications. Often data arrives piecemeal and recursive estimation allows the parameter estimates to be updated as new observations become available. The algorithm in [6] cannot be used for recursive estimation as it requires the entire observation sequence in each iteration. In this work we develop a recursive estimation scheme using the recursive EM algorithm [7]. In our derivation the parameters are assumed to be constant. However, with proper choice of the step size the proposed algorithm can be easily extended to the time–varying case.

The recursive EM algorithm is a stochastic approximation procedure with a specialized gain matrix derived from the augmented data of EM. Under mild conditions, the estimates generated by such procedures are strongly consistent and asymptotic normally distributed [2] [7]. Because of the simple structure of the augmented data, the augmented information matrix is usually very easy to compute and invert. Consequently, the resulted recursion has a very simple implementation.

The remainder of this paper is organized as follows. In Section 2 we give the problem formulation. In Section 3 we provide some details of the recursive estimation techniques and in Section 4 we apply it to the K-distribution. In Section 5 we provide simulation results and in Section 6 we give some conclusions.

2. PROBLEM FORMULATION

Let $y_1, y_2, \dots, y_T$ be a sequence of independent observations of the K–distributed random variable Y. The underlying probability density function (p.d.f.) $f_Y(y|\underline{\vartheta})$ may be derived by considering a generalized Rayleigh distribution with a gamma–distributed mean parameter. We may write

$$f_Y(y|\underline{\vartheta}) = \int_0^\infty f_{Y|W}(y|w) f(w|\underline{\vartheta}) dw \; . \tag{1}$$

The K–distribution results when

$$f_{Y|W}(y|w) = \frac{2y^{N-1}}{(2w)^{N/2}\Gamma(N/2)} \exp(\frac{-y^2}{2w}) \tag{2}$$

i.e., the generalized Rayleigh distribution resulting from the square root of the sum of the squares of N independent and identically distributed (iid) Gaussian scalar random variables with zero mean and variance w. The gamma function is denoted by $\Gamma(\cdot)$. The variance w is an observation of the gamma distributed random variable W with the parameters $\underline{\vartheta} = [\sigma, \alpha]$, i.e.,

$$f_W(w|\underline{\vartheta}) = \frac{\sigma^\alpha w^{\alpha-1}}{\Gamma(\alpha)} \exp(\sigma w). \tag{3}$$

Substituting (3) and (2) into (1) and performing the integration using eq. (3.471(9)), p. 340 of [3] yields

$$f_Y(y|\underline{\vartheta}) = \frac{y^{\alpha-1+N/2}2^{2-N/4-\alpha/2}\sigma^{N/4+\alpha/2}}{\Gamma(N/2)\Gamma(\alpha)} K_{\alpha-N/2}(\sqrt{2\sigma}\, y) \tag{4}$$

where $K_\eta(\cdot)$ is given by eq. (8.406), p. 952 of [3] and is known as the modified Bessel function of the second kind of order η. Equation (4) is the K–distribution with the parameter vector $\underline{\vartheta} = [\sigma, \alpha]$. As in other studies, we do not consider estimation of N as it corresponds to the number of looks in SAR applications which is generally known. The problem of interest is to find the ML estimate $\hat{\underline{\vartheta}}$ based on the data set $\{y_t\}_{t=1}^T$.

3. RECURSIVE EM ALGORITHM

We give a brief description of the recursive EM algorithm in its most general form. The observation can be either a vector or a scalar. Suppose $\underline{y}_1, \underline{y}_2, \ldots$ are independent observations, each with the underlying p.d.f. $f_{\underline{Y}}(\underline{y}|\underline{\vartheta})$, where $\underline{\vartheta}$ denotes an unknown parameter vector. The corresponding augmented data $\underline{z}_1, \underline{z}_2, \ldots$ are characterized by the p.d.f. $f_{\underline{Z}}(\underline{z}|\underline{\vartheta})$. Let $\underline{\vartheta}^t$ denote the estimate after t observations. The following procedure is aimed at finding the extremum $\underline{\vartheta}^*$ of $\log f_{\underline{Y}}(\underline{y}|\underline{\vartheta})$ which would coincide with the maximum likelihood estimate,

$$\underline{\vartheta}^{t+1} = \Pi_\Theta[\underline{\vartheta}^t + at^{-\beta}\, \mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)^{-1}\underline{\gamma}(\underline{y}_t, \underline{\vartheta}^t)], \tag{5}$$

where Π_Θ is the projection onto the constraint set Θ, $a > 0$ is a constant and

$$\underline{\gamma}(\underline{y}_t, \underline{\vartheta}^t) = \underline{\nabla}_{\underline{\vartheta}} \log f_{\underline{Y}}(\underline{y}_t|\underline{\vartheta})|_{\underline{\vartheta}=\underline{\vartheta}^t}, \tag{6}$$

$$\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t) = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}}\underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{Z}}(\underline{z}|\underline{\vartheta})|\underline{y}_t, \underline{\vartheta}\right]|_{\underline{\vartheta}=\underline{\vartheta}^t} \tag{7}$$

represent the gradient vector and the augmented information matrix calculated at $\underline{\vartheta}^t$, respectively. $\underline{\nabla}_{\underline{\vartheta}}$ is a column gradient operator with respect to $\underline{\vartheta}$. $(\cdot)^T$ represents the transpose of a vector. The constraint set can be chosen as a hyper-rectangle $[\underline{a}, \underline{b}]$ defined by the lower bound $\underline{a}$ and the upper bound $\underline{b}$.

A proper choice of β depends on the following matrix

$$\mathbf{D}_{\mathrm{EM}}(\underline{\vartheta}) = \frac{1}{2}\mathbf{I} - a\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta})^{-1}\mathcal{I}(\underline{\vartheta}), \tag{8}$$

where

$$\mathcal{I}(\underline{\vartheta}) = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}}\underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{Y}}(\underline{y}|\underline{\vartheta})\right] \tag{9}$$

denotes the Fisher information matrix corresponding to one observation. $\mathbf{I}$ is an identity matrix with the dimension of $\underline{\vartheta}$. Use $\beta = 1$ if $\mathbf{D}_{\mathrm{EM}}(\underline{\vartheta})$ is a stable matrix. Otherwise $1/2 < \beta < 1$. A matrix is called stable if all eigenvalues have negative real parts.

If the initial estimate $\underline{\vartheta}^0$ is close enough to $\underline{\vartheta}^*$ and the conditions (a) $\mathsf{E}\|\underline{\gamma}(\underline{y}_t, \underline{\vartheta}^t)\|^2 < \infty$ and (b) $\mathbf{0} < \mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t) < \infty$ are satisfied, $\underline{\vartheta}^t$ converges with probability one to $\underline{\vartheta}^*$ [2]. For large t, the normalized error $t^{\beta/2}(\underline{\vartheta}^t - \underline{\vartheta}^*)$ is normally distributed with zero mean and a covariance matrix which can be obtained by solving a matrix equation. As pointed out by Titterington [7], recursion (5) will not lead to asymptotic efficiency. Asymptotic efficiency implies at same time the best convergence rate achievable by such procedures. However, compared to an optimal procedure

$$\underline{\vartheta}^{t+1} = \underline{\vartheta}^t + t^{-1}\, \mathcal{I}(\underline{\vartheta}^t)^{-1}\underline{\gamma}(\underline{y}_t, \underline{\vartheta}^t), \tag{10}$$

the practical advantage of recursion (5) is that $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)^{-1}$ will usually be much easier to compute than $\mathcal{I}(\underline{\vartheta}^t)^{-1}$. Besides, the optimal convergence rate can also be achieved by recursive EM if an additional averaging step proposed by Polyak [5] is undertaken.

4. RECURSIVE ESTIMATION

In this section we apply the recursive EM algorithm to estimate the parameters of K–distribution. We obtain the gradient vector (6) by taking the first derivative of eq. (4). The first and second component are given by

$$\frac{\partial}{\partial\sigma}\log f_Y(y_t|\underline{\vartheta})_{\underline{\vartheta}=\underline{\vartheta}^t} = \frac{1}{\sigma^t}\left(\frac{N}{4} + \frac{\alpha^t}{2}\right)$$

$$-\frac{y_t[K_{\alpha^t-\frac{N}{2}-1}(\sqrt{2\sigma^t}y_t) + K_{\alpha^t-\frac{N}{2}+1}(\sqrt{2\sigma^t}y_t)]}{2\sqrt{2\sigma^t}K_{\alpha^t-\frac{N}{2}}(\sqrt{2\sigma^t}y_t)} \tag{11}$$

and

$$\frac{\partial}{\partial\alpha}\log f_Y(y_t|\underline{\vartheta})|_{\underline{\vartheta}=\underline{\vartheta}^t} = \log y_t - \frac{\log 2}{2} + \frac{\log\sigma^t}{2} - \Psi(\alpha^t)$$

$$+\frac{1}{K_{\alpha^t-\frac{N}{2}}(\sqrt{2\sigma^t}y_t)}\left\{[\frac{\partial}{\partial\alpha}K_{\alpha-\frac{N}{2}}(\sqrt{2\sigma}\, y_t)]|_{\underline{\vartheta}=\underline{\vartheta}^t}\right\}, \tag{12}$$

respectively. Note that $\Psi(\cdot)$ is the digamma function defined as the derivative of the log of the gamma function [1].

The matrix $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)$ is derived from the augmentation scheme proposed by Roberts [6]

$$\underline{z} = [y, w]^T. \tag{13}$$

The augmented data specified above yields the corresponding p.d.f.

$$f_{\underline{Z}}(\underline{z}|\underline{\vartheta}) = f_{Y|W}(y|w) f_W(w|\underline{\vartheta}). \tag{14}$$

Equivalently,

$$\log f_{\underline{Z}}(\underline{z}|\underline{\vartheta}) = \log f_{Y|W}(y|w) + \log f_W(w|\underline{\vartheta}). \tag{15}$$

Clearly, $\log f_{\underline{Z}}(\underline{z}|\underline{\vartheta})$ depends on $\underline{\vartheta}$ only through the second term on the right hand side of eq. (15). The resulting augmented information matrix

$$\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t) = \frac{1}{\sigma^t} \begin{bmatrix} \alpha^t/\sigma^t & -1 \\ -1 & \Psi'(\alpha^t)\sigma^t \end{bmatrix} \tag{16}$$

can be implemented easily. The trigamma function $\Psi'(\cdot)$ is the second derivative of the log of the gamma function [1].

Occasionally, $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)$ may become nearly singular during recursion and lead to numerical instability. To overcome this problem, we suggest to check the conditional number c_t of $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)$. If c_t is larger than a pre–selected constant $c_{\max}$, $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^t)$ will be replaced by the augmented matrix $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^{t-1})$ of the previous recursion.

5. SIMULATION RESULTS

In this section, we apply the proposed algorithm to simulated data and study its performance. Three parameter sets, (1) $[\sigma_{\mathrm{ref}}, \alpha_{\mathrm{ref}}] = [1.0,\ 0.5]$, (2) $[2.0,\ 1.0]$ and (3) $[3.0,\ 1.5]$ are considered. The initial estimates are given by (1) $[\sigma_0, \alpha_0] = [1.3,\ 0.8]$, (2) $[2.3,\ 1.3]$ and (3) $[2.7,\ 1.8]$. All three experiments use the same step size $at^{-\beta} = 0.2t^{-0.6}$. We use a constraint set Θ with lower bound $L = 0.1$ and upper bound $U = 6$ for both parameters. The data length is $T = 1000$. Each experiment is run thourgh 200 Monte Carlo trials. The functions $\Psi(\cdot)$, $\Psi'(\cdot)$ and $\partial/\partial\alpha K_\alpha(\cdot)$ are implemented in the similar way as in [6].

The Mean Squared Error (MSE) versus number of recursions are displayed in fig. 1, fig. 2 and fig. 3. Upper part is the MSE of σ, lower part is the MSE of α. In general, the MSE decreases with increasing number of recursions. In the upper part of fig. 2 and fig. 3, one can observe that at the beginning of the recursion, the MSE may increase slightly. This can be seen as the transient behavior of the algorithm.

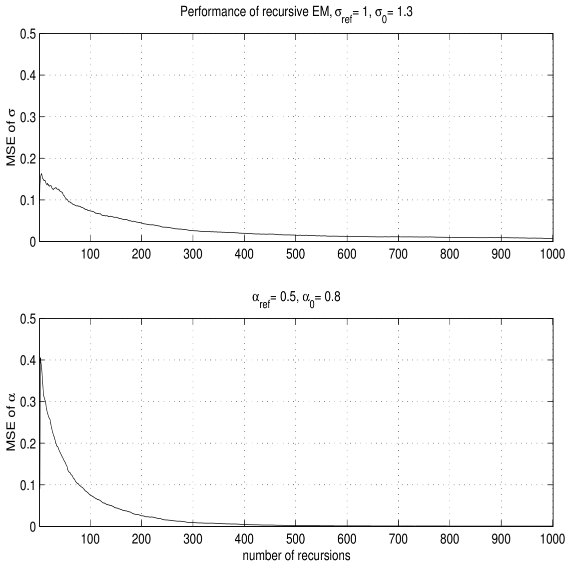

Fig. 1. MSE versus number of recursions. $\sigma_{\mathrm{ref}} = 1$, $\alpha_{\mathrm{ref}} = 0.5$. Upper: MSE of σ, lower: MSE of α.

In all three experiment, the MSEs of the σ estimates are higher than those of the α estimates, although the distances to the reference parameters are the same, i.e. 0.3. The sequence $\{\alpha^t\}$ generated by the recursive procedure has a better convergence rate than $\{\sigma^t\}$. Clearly, the convergence rates are also influenced by the reference parameters. For the same number of recursions, the first experiment has the lowest MSE. The third experiment has the largest MSE. The MSEs $[\sigma_{\mathrm{MSE}}, \alpha_{\mathrm{MSE}}]$ after 1000 recursions are (1) $[0.0072,\ 0.00065]$, (2) $[0.04,\ 0.0053]$ and (3) $[0.0827,\ 0.0164]$. This implies that the proposed algorithm provides quite accurate estimates.

6. CONCLUSION

This work deals with recursive estimation of K–distribution parameters. A recursive procedure based on the recursive EM algorithm was developed to find the ML estimates. Because of the simple structure of the augmented data, the resulting algorithm has a simple implementation. By numerical results we show that this procedure has good convergence behavior and provides accurate estimates. Various parameter sets lead to different convergence rates.

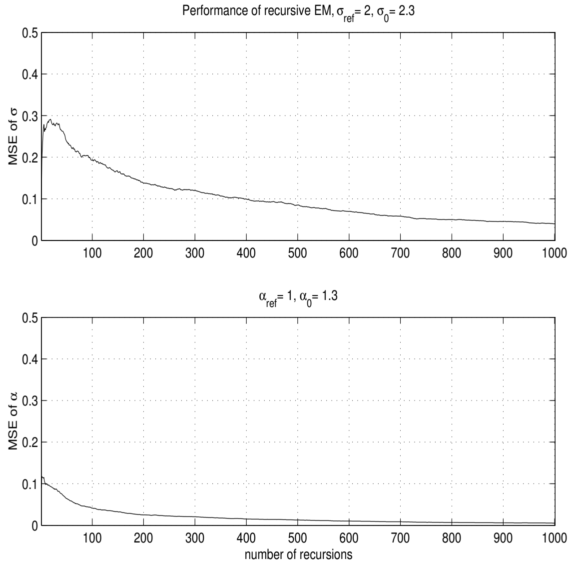

Fig. 2. MSE versus number of recursions. $\sigma_{\mathrm{ref}} = 2$, $\alpha_{\mathrm{ref}} = 1$.

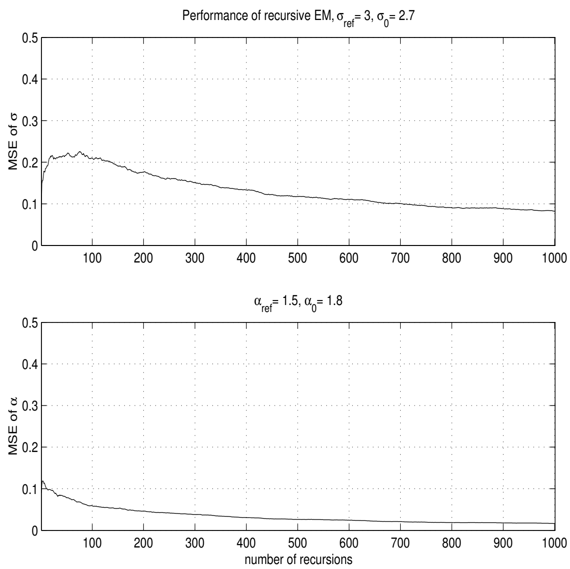

Fig. 3. MSE versus number of recursions. $\sigma_{\mathrm{ref}} = 3$, $\alpha_{\mathrm{ref}} = 1.5$.

7. REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications, 9th edition, 1972.

[2] Pei Jung Chung. *Fast Algorithms for Parameter Estimation of Sensor Array Signals*. Dr.–Ing. Dissertation, Faculty of Electrical Engineering and Information Sciences, Ruhr–Universität Bochum, Bochumer Universitätsverlag Bochum, 2002.

[3] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, New York, fourth edition, 1980.

[4] E. Jakeman and P.N. Pusey. A model for non-Rayleigh sea echo. *IEEE Trans. Antennas Propagat*, 24:806–814, November 1976.

[5] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30:838–855, 1992.

[6] W.J.J. Roberts and S. Furui. Maximum likelihood estimation of K-distribution parameters via the Expectation–Maximization algorithm. *IEEE Trans. Signal Processing*, 48(12):3303–3306, December 2000.

[7] D.M. Titterington. Recursive parameter estimation using incomplete data. *J. R. Statist. Soc. B*, 46(2):257–267, 1984.

[8] K.D. Ward. Compound representation of high resolution sea clutter. *Electron. Lett.*, 17(16):561–563, August 1981.