FULL SPACE-TIME CLUTTER COVARIANCE ESTIMATION

Frank C. Robey[1] *and Edward J. Baranoski*[2]
MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02173,
E-mail: [1]fcr@ll.mit.edu, [2]ejb@ll.mit.edu

ABSTRACT

For an airborne radar array, beamforming and detection are problems in both space and time. To null clutter, it is necessary to exploit both dimensions, and to do so optimally requires knowledge of the full space-time covariance matrix and the array steering vectors. This paper derives an Expectation-Maximization (EM) algorithm for the estimation of full space-time covariance matrices while simultaneously estimating array steering vectors. The EM approach iterates between estimating the spatial steering vectors and power associated with clutter scattering from different angles and the formation of a full space-time covariance matrix. The final result is an estimate of the set of array steering vectors and an estimate of the space-time covariance matrix. In practice, one would never need to form this covariance since all calculations could be performed using the SVD of the appropriately weighted clutter space-time steering vector matrix. The technique is capable of providing a positive definite estimate of the space-time covariance and complete array calibration with only a single space-time data sample.

1. INTRODUCTION

For an airborne radar array with multiple pulses, beamforming and detection are problems in both space and time [1, 2]. The radar returns from a target are functions of velocity and position, with a Doppler frequency based upon the scattering objects velocity relative to the aircraft. Clutter returns share this space-time relationship but the Doppler frequency is a function of only platform velocity and altitude and the clutter azimuth. on altitude can be neglected. For an adaptive radar to detect target returns while rejecting clutter, both the spatial and temporal dimensions must be exploited.

In an airborne adaptive radar array, there are two major problems that we would like to address. The first is that of poor array calibration. Flexure of the airframe and mutual coupling between antenna elements can induce uncertainties in the array calibration. The second problem is estimating a space-time clutter covariance matrix. Because of the motion of the aircraft, uncertainties of the array calibration and different clutter environments, the space-time clutter covariance is unknown and can change over a time period on the order of a few coherent processing intervals. The full dimension of the space-time adaptive problem is the number of spatial elements times the number of pulses which can be very large and can make it impossible to obtain sufficient training samples for estimating a full space-time covariance matrix using unstructured techniques.

There have been algorithms proposed that solve in varied ways either the array calibration problem or the covariance estimation problem. Steinberg [3] proposed a technique for calibrating an array when a dominate scatterer can be found in the data, and Attia [4] has proposed a technique making use of the spatial correlation properties of the radar clutter. The first of these is not applicable if a dominate scatterer can not be found such as might occur over water, and the second would not be applicable if the complex gain of each of the array elements varies independently as a function of azimuth.

Fuhrmann [5] and Robey [6] have shown that, with knowledge of the array manifold, high performance adaptive detection and beamforming can be accomplished by modeling the interference environment even when the number of samples used to estimate the covariance matrix are small. The success of these structured covariance techniques has led us to explore simultaneously calibrating the antenna array while estimating the clutter covariance matrix.

This paper explores an Expectation-Maximization (EM) algorithm [7] for the estimation of full space-time covariance matrices for airborne radar applications while simultaneously determining the radar array calibration. This algorithm exploits the Doppler phase progression that clutter sources incur relative to the moving platform and the comparatively low rank of the clutter subspace in the full space-time covariance.

2. THEORY

In this section we will explore the use of the EM algorithm for determining the space-time clutter subspace. We will use the following assumptions:

- Complex Gaussian independent identically distributed zero-mean receiver noise, both spatially and temporally, with known variance. Jamming is neglected for simplicity in this paper. To accommodate jamming and spatially colored noise, the interference covariance may be estimated before the coherent processing interval. Using this covariance, a whitening transformation can be calculated and the data transformed into this whitened space. All the processing outlined in this paper and any further processing of the data can take place in this whitened space.
- A clutter distribution such that the characteristics of the clutter scattering shows up as unknown parameters in the likelihood equation.
- A spatial array of uniformly polarized sensors.

This work was supported in part by Air Force contract F19628-95-C-0002.

0-7803-2431-5/95 $4.00 © 1995 IEEE

- Clutter signals of uniform polarization: the calibration is in the polarization plane of the clutter signals.
- Narrow band radar, so array dispersion caused by the propagation delay across the array can be neglected.
- K complex i.i.d. zero-mean samples (no target signals are present). These possibly multiple samples can come, for example, from nearby range gates where the clutter is expected to be homogeneous.
- High PRF radar so that clutter is unambiguous in Doppler frequency, or a transmit beam shape such that ambiguous clutter scatterers are sufficiently suppressed below the receiver noise level. Backlobes of the transmit beam are also assumed to be suitably suppressed.
- All vectors are assumed to be unit 2-norm column vectors (unless otherwise noted).
- Internal clutter motion is assumed to be negligible.

Let us consider a pulsed Doppler radar system with N antenna elements and M pulses. The Doppler frequency $f_d(\theta)$ of a clutter patch at angle θ is due to the apparent motion of the clutter relative to the moving airborne platform. This frequency is dependent on the azimuth of the clutter as

$$f_d(\theta) = \frac{2v}{\lambda}\cos\theta \tag{1}$$

where v is the velocity of the radar platform, λ is the wavelength of the radar carrier frequency, and θ is the angle of the clutter patch relative to the velocity vector of the radar platform. We can construct a full NM-element space-time steering vector $\widetilde{\mathbf{v}}(\theta)$ for the clutter patch at angle θ as

$$\widetilde{\mathbf{v}}(\theta) = \mathbf{d}(\theta) \otimes \mathbf{v}(\theta) \tag{2}$$

where $\mathbf{d}(\theta) = \frac{1}{\sqrt{M}}\left[\begin{array}{ccc} e^{j2\pi f_d(\theta)} & \cdots & e^{j2\pi M f_d(\theta)}\end{array}\right]^T$ is the Doppler phase vector for clutter from angle θ, $\mathbf{v}(\theta)$ is the spatial steering vector to the clutter patch, and $\otimes$ represents the Kronecker product such that we could alternatively write $\widetilde{\mathbf{v}}(\theta)$ as

$$\widetilde{\mathbf{v}}(\theta) = \frac{1}{\sqrt{M}}\left[\begin{array}{c} e^{j2\pi f_d(\theta)}\mathbf{v}(\theta) \\ \vdots \\ e^{j2\pi M f_d(\theta)}\mathbf{v}(\theta)\end{array}\right]. \tag{3}$$

For a given range, the space-time clutter return is the integral over the angular space Θ

$$\widetilde{\mathbf{x}}_c(\rho) = \int_\Theta a(\theta,\rho)\widetilde{\mathbf{v}}(\theta)d\theta, \tag{4}$$

where $a(\theta,\rho)$ is the amplitude of the clutter return from angle θ and range ρ. We will assume that the ranges of the clutter patches considered are close enough to ignore any possible range dependence on the steering vectors or the Doppler frequency coming from specific azimuths.

The data received by the array will be denoted $\mathbf{x}$ and is

$$\mathbf{x} = \widetilde{\mathbf{x}}_c(\rho) + \mathbf{n} \tag{5}$$

with $\mathrm{E}\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma^2\mathbf{I}_M \otimes \mathbf{I}_N = \sigma^2\mathbf{I}$. We will have K of these data samples. The space-time covariance of these received data vectors is

$$\mathbf{R}_x(\rho) = \sigma^2\mathbf{I} + \int_\Theta |a(\theta,\rho)|^2\,\widetilde{\mathbf{v}}(\theta)\widetilde{\mathbf{v}}^H(\theta)d\theta, \tag{6}$$

where σ^2 is the noise which is assumed to be independent from element to element and pulse to pulse. Due to the limited spatial and temporal resolution, the continuous clutter distribution can be modeled as a discrete summation of clutter patches. The covariance matrix can then be rewritten as

$$\mathbf{R}_x(\rho) = \sigma^2\mathbf{I} + \sum_{i=1}^{L} |a(\theta,\rho)|^2\,\widetilde{\mathbf{v}}_i\widetilde{\mathbf{v}}_i^H, \tag{7}$$

where $\widetilde{\mathbf{v}}_i$ is shorthand for $\widetilde{\mathbf{v}}(\theta_i)$.

3. DERIVATION OF ESTIMATOR

We will use the Expectation-Maximization (EM) algorithm [8] to perform the estimation of the full space-time covariance. The EM algorithm consists of the following two steps:

1. E-step: Given a complete data set $\mathbf{Y}$, that if known would uniquely determine the observed data, compute $\mathrm{E}\{l_{cd}(\mathbf{Y};\Theta)|\mathbf{Z},\Theta^p\}$. $l_{cd}(\mathbf{Y};\Theta)$ here is the complete data log-likelihood and Θ^p is the parameter estimates at step p.
2. M-step: Find $\Theta^{p+1} = \mathrm{argmax}_\Theta \mathrm{E}\{l_{cd}(\mathbf{Y};\Theta)|\mathbf{Z},\Theta^p\}$. These are then the parameter estimates at step $p+1$.

For our EM implementation we choose for the complete data space the full space-time signal from each patch of clutter in the presence of a small additive noise component. The rational for choosing this model as the complete data space (besides the physical significance) is that if we had such data the estimation of the spatial steering vector to each patch and the associated clutter power for that patch it would be a simple maximum likelihood problem to find the full space-time covariance matrix. Instead, we estimate the hypothetical complete data statistics from incomplete data measurements. The complete data is:

$$\{\mathbf{y}_{ik}\} : \mathbf{y}_{ik} = \widetilde{\mathbf{v}}_i s_{ik} + \mathbf{n}_{yik} = (\mathbf{d}_i \otimes \mathbf{v}_i)\, s_{ik} + \mathbf{n}_{yik}, \tag{8}$$

with the following characteristics:

$\mathbf{d}_i$,	the Doppler progression, is known for any particular clutter patch
$\mathbf{v}_i$	the array steering vector, is unknown
s_{ik}	the clutter signal from that patch, is unknown
$\mathbf{n}_{yik}$	an independent noise vector for each clutter patch.

and,

$$\mathrm{E}\left\{\mathbf{n}_{yik}\mathbf{n}_{yik}{}^H\right\} = \varepsilon_i\mathbf{I}, \quad \varepsilon_i \text{ is known} \tag{9}$$

$$\mathrm{E}\{s_{ik}s_{ik}^*\} = \gamma_i \tag{10}$$

The incomplete or measured data is simply the sum of the complete data;

$$\mathbf{x}_k = \sum_I \mathbf{y}_{ik}\ . \tag{11}$$

What we have done is to assign some portion of the receiver noise as an independent noise source for each of the clutter scatterers.

There are two likelihood functions that are needed for the EM algorithm. The first is the incomplete, or received data likelihood function. The second is the likelihood function of the complete data. We will deal with the complete data log-likelihood first, which dropping constant terms is proportional to [9]

$$l_{cd} = \sum_{i=1}^{L} \left(-\log|\mathbf{R}_{y_i}| - \operatorname{tr} \mathbf{R}_{y_i}^{-1}\mathbf{S}_i\right) = \sum_{i=1}^{L} l_{cd}^{(i)}, \qquad (12)$$

This equation was originally a double summation over the different clutter patches and the number of samples K at each clutter patch however, the second summation was subsumed into $\mathbf{S}_i$, the sample covariance of the hypothetical i-th complete data vector, $\frac{1}{K}\sum_{k=1}^{K} \mathbf{y}_{ik}\mathbf{y}_{ik}^H$. $\mathbf{R}_{y_i}$ is the true covariance matrix for the i-th complete data vector

$$\mathbf{R}_{y_i} = \varepsilon_i\mathbf{I} + \gamma_i\mathbf{v}_i\mathbf{v}_i^H, \qquad (13)$$

which is rank-1 plus diagonal and its singular values are $\gamma_i + \varepsilon_i$ and an $(N-1)$ multiplicity of ε_i. Because of the independence of the complete data, we can deal with each complete data vector separately, with the log-likelihood for the i-th complete data vector denoted by $l_{cd}^{(i)}$.

The first step in the derivation of the EM algorithm is taking the conditional expected value of the log-likelihood function, conditioned on the incomplete data and the current parameter set $\Theta = \{\mathbf{x}, \{\hat{\mathbf{v}}^p\}, \{\gamma_i^p\}\}$ where $\mathbf{x}$ is the set of array measurements, $\{\hat{\mathbf{v}}^p\}$ is the estimate of the unknown spatial steering vectors, and $\{\gamma_i^p\}$ is the set of signal power estimates. For the i-th complete data vector, this is

$$\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} = -\log|\mathbf{R}_{y_i}| - \operatorname{tr} \mathbf{R}_{y_i}^{-1}\mathrm{E}\{\mathbf{S}_i|\Theta\}. \qquad (14)$$

We will defer taking the conditional expectation on the right until after the second step is accomplished.

The second EM step is to find the unknown parameters in the complete data space that maximise (14). Neglecting constant terms, (13) can be used to write (14) as

$$\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} = -\log(\gamma_i + \varepsilon_i) - \operatorname{tr} \mathbf{R}_{y_i}^{-1}\mathrm{E}\{\mathbf{S}_i|\Theta\}. \qquad (15)$$

Here we have made use of the property that the determinant of the covariance matrix is the product of the singular values. The ε_i singular values that would appear alone in this equation have been dropped since they do not enter into the maximization.

We will now introduce a change in coordinates via a unitary transformation in order to assist in the maximization. The particular transformation that we will use is:

$$\mathbf{Q} = \left[\mathbf{d} \,\vdots\, \mathbf{D}_\perp\right] \otimes \mathbf{I}_N. \qquad (16)$$

This can be recognized as a Doppler filtering transformation. Here we are only concerned with the output of a single Doppler filter. The first column of the matrix entering into the Kronecker product is the phase progression, or a DFT vector; the remaining columns are an orthonormal basis for the orthogonal complement of $\mathbf{d}_i$. The orthonormal basis is arbitrary for the purposes of this simplification, however, in implementing the algorithm it would be expected that there would be significant computational savings if this basis were the set of vectors required to complete the DFT matrix, and if the other clutter patches were chosen such that their Doppler progressions corresponded to these column vectors.

Because the transformation is unitary, we can insert $\mathbf{Q}\mathbf{Q}^H$ in (14) in the full space-time dimensioned data where appropriate for simplification. We will then make use of the following,

$$\mathbf{Q}\left(\mathbf{d}\mathbf{d}^H \otimes \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\right)\mathbf{Q} = \left(\mathbf{e}\mathbf{e}^H\right) \otimes \left(\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\right) \qquad (17)$$

where $\mathbf{e}$ is the elementary vector $[1, 0, \cdots 0]^H$.

We introduce the unitary transformation $\mathbf{Q}\mathbf{Q}^H$ both before and after $\mathrm{E}\{\mathbf{S}_i|\Theta\}$ in (15) and make use of the rotation property of the trace operator [10], to find

$$\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} = -\log(\gamma_i + \varepsilon_i) - \operatorname{tr}\left\{\mathbf{Q}^H\mathbf{R}_y^{-1}\mathbf{Q}\mathbf{Q}^H\mathrm{E}\{\mathbf{S}_i|\Theta\}\mathbf{Q}\right\}. \qquad (18)$$

In order to further simplify this we will make use of the Sherman-Morrison-Woodbury identity [11] to find the inverse of $\mathbf{R}_{y_i}$ as given by (13)

$$\mathbf{R}_{y_i}^{-1} = \frac{1}{\varepsilon_i}\mathbf{I} - \frac{\gamma_i}{\varepsilon_i(\varepsilon_i + \gamma_i)}\mathbf{v}\mathbf{v}^H. \qquad (19)$$

Substituting this into (15) and simplifying leads to

$$\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} = -\log(\gamma_i + \varepsilon_i) - \operatorname{tr}\left\{\left(\frac{1}{\varepsilon_i}\mathbf{I}_M \otimes \mathbf{I}_N - \frac{\gamma_i}{\varepsilon_i(\varepsilon_i + \gamma_i)}\mathbf{e}\mathbf{e}^H \otimes \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\right) \times \mathbf{Q}^H\mathrm{E}\{\mathbf{S}_i|\Theta\}\mathbf{Q}\right\}. \qquad (20)$$

We now examine the second portion of the term within the trace operator in order to perform the maximization with respect to the unknown parameter $\tilde{\mathbf{v}}$. The effect of the $\mathbf{e}\mathbf{e}^H$ term is that only the upper left $N \times N$ corner of $\mathbf{Q}^H\mathrm{E}\{\mathbf{S}_i|\Theta\}\mathbf{Q}$ is significant in the maximization. By the Rayleigh quotient theorem we can see that This equation will be maximized if $\tilde{\mathbf{v}}$ is colinear with the principal eigenvector of the upper left corner of this matrix.

The remaining parameter that we need to maximize the conditional expected value of the log-likelihood function is the clutter power γ_i. We will define a temporary variable α to simplify in this maximization,

$$\alpha = \operatorname{tr}\left\{\mathbf{e}\mathbf{e}^H \otimes \hat{\mathbf{v}}\hat{\mathbf{v}}^H\mathbf{Q}^H\mathrm{E}\{\mathbf{S}_i|\Theta\}\mathbf{Q}\right\} \qquad (21)$$

where $\hat{\mathbf{v}}$ is the maximizing vector. We now need to maximize

$$\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} = -\log(\varepsilon_i + \gamma_i) + \frac{\gamma_i}{\varepsilon_i(\varepsilon_i + \gamma_i)}\alpha \qquad (22)$$

where we have dropped the terms not dependent on γ_i. Taking the derivative of this with respect to γ_i and setting the result equal to zero we can solve for γ_i

$$\begin{aligned}\frac{\delta}{\delta\gamma_i}\mathrm{E}\left\{l_{cd}^{(i)}|\Theta\right\} &= -\frac{1}{(\varepsilon_i + \gamma_i)} + \frac{1}{(\varepsilon_i + \gamma_i)^2}\alpha = 0, \\ \gamma_i &= \alpha - \varepsilon_i \end{aligned} \qquad (23)$$

Since there is a positivity constraint on γ_i and the likelihood function is convex, the maximizing value of γ_i is

$$\gamma_i = \max(0, \alpha - \varepsilon_i). \qquad (24)$$

This completes the M-step of the EM derivation. We will now go back and determine the conditional expectation $E\{S_i|\Theta\}$ that we had deferred. We will need the following results from estimation theory, recalling that each complete data vector y_{ik} is independent from every other data vector:

$$\begin{aligned} E\{yx^H|\Theta\} &= R_{yx} \\ &= E\left\{ y_i \sum_i y_i^H \middle| \Theta \right\} \\ &= R_{y_i}. \end{aligned} \quad (25)$$

After some simplification, the conditional expectation $E\{S_i|\Theta\}$ is then [12]

$$\begin{aligned} &E\{S_i|\Theta\} \quad (26) \\ &= R^p_{y_i} - R^p_{y_i}\left((R^p_x)^{-1} - (R^p_x)^{-1}S_x(R^p_x)^{-1}\right)R^p_{y_i}{}^H \end{aligned}$$

and as stated earlier, we will need only the upper left of this matrix after Doppler processing. S_x is the sample covariance matrix of the incomplete data. Further simplifications of (26) are possible, but for the sake of brevity are not shown.

We now need to put the entire algorithm together. To form the estimate of the full space-time covariance matrix while also determining the array calibration, we perform the following steps:

1. Determine an initial estimate of the clutter scatterers steering vectors and power. This could be Vandermonde steering vectors assuming a uniform linear array with an *a priori* assumption about the scattering power (alternatively one could assume that the clutter energy is zero and the initial steering vectors would then be immaterial).
2. Form the incomplete data space-time sample covariance matrix using the collected data samples.
3. Form an estimate of the full space-time covariance matrix using the estimated steering vectors and clutter power.
4. Calculate the conditional expected value $E\{S_i|\Theta\}$, and transform this using Q, which will Doppler process this covariance according to the expected Doppler shifts of each of the clutter patches.
5. For each Doppler frequency, estimate the spatial steering vector as the principal component of the upper left portion of the Doppler processed complete data covariance.
6. Estimate the clutter power associated with each patch.
7. Iterate to step number 3 until acceptable convergence is obtained.

Once the steering vector $\hat{v}$ and power γ_i^p are estimated, the full space-time vector can be constructed by applying the Doppler sequence d. When this is performed for each Doppler frequency, we can construct the full complement of space-time vectors. The clutter covariance is simply the sum of the vector outer product of each space-time vector. Given that the clutter is low rank [2], the space-time vectors themselves are a more economical method of representing the full clutter information.

4. CONCLUSION

In an airborne adaptive array radar, there are two difficult problems that need to be addressed. These problems are the determination of the array calibration, and estimation of the received data covariance matrix. This paper has addressed both of these two problems in a unified approach. The resulting estimator will provide both array calibration and an estimate of the clutter scattering power when only a limited number of data samples are available. The array calibration could greatly reduce the amount of flight time needed to characterize an antenna array and could provide a means of calibrating in flight without cooperative sources or knowledge of the clutter scattering. This method shows great promise in reducing the number of samples required when one needs to form a full space-time covariance matrix for the received data.

REFERENCES

[1] R. Klemm, "Adaptive airborne MTI: An auxiliary channel approach," *IEE Proceedings*, vol. 134, pp. 269–276, June 1987.

[2] J. Ward, "Space-time adaptive processing for airborne radar," Tech. Rep. 1015, MIT Lincoln Laboratory, December 1994.

[3] B. Steinberg, "Radar imaging from a distorted array: The radio camera algorithm and experiments," *IEEE Transactions on Antennas and Propagation*, vol. AP-29, no. 5, 1981.

[4] E. Attia and B. Steinberg, "Self-cohering large antenna arrays using the spatial correlation properties of radar clutter," *IEEE Transactions on Antennas and Propagation*, vol. AP-37, no. 1, 1989.

[5] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Sig. Proc.*, vol. 39, no. 10, pp. 2194–2198, 1991.

[6] F. C. Robey, "A covariance modeling approach to adaptive beamforming and detection," Tech. Rep. 918, MIT Lincoln Laboratory, July 1991.

[7] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Royal Statistical Society.*, vol. B-39, pp. 1–37, 1977.

[8] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Royal Statistical Society.*, vol. B-39, pp. 1–37, 1977.

[9] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Annals of Mathematical Statistics*, vol. 34, pp. 152–177, March 1963.

[10] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Tech. Rep. 848, MIT Lincoln Laboratories, April 1989.

[11] G. H. Golub and C. F. V. Loan, *Matrix Computations (2^{ND} ed.)*. John Hopkins University Press, 1989.

[12] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison Wesley, 1991.