A Linear Least Squares Algorithm for Bearings-Only Target Motion Analysis*

Roy L. Streit
Naval Undersea Warfare Center
Code 2002, Building 1171/3
Newport, RI 02841
streitrl@csd.npt.nuwc.navy.mil

Michael J. Walsh
Naval Undersea Warfare Center
Code 2211, Building 1171/1
Newport, RI 02841
walshmj@csd.npt.nuwc.navy.mil

Abstract—Traditional bearings-only target motion analysis (TMA) statistical models assume *a priori* that measurements are independent when conditioned on the target. This paper presents a novel track-before-detect "empirical" maximum *a posteriori* (EMAP) approach in which measurements are assumed independent prior to the detection decision. The EMAP estimators proposed here are joint detection/estimation methods whose intended use is target tracking. A limiting case of the EMAP formulation is shown to be equivalent to the traditional maximum likelihood (ML) formulation. Triangulation and constant velocity target examples are presented. The EMAP algorithm is an iteratively re-weighted linear least squares algorithm for these problems, and has significantly lower computational complexity than the standard ML estimator.

TABLE OF CONTENTS

1. INTRODUCTION

Two formulations of the bearings-only target motion analysis (TMA) problem are discussed in this paper. The formulations use nearly indistinguishable statistical probability density functions (PDF's) of the measured bearings, but they differ significantly in their statistical interpretation. The first formulation based on the ML method is standard because it assumes measurements are statistically independent, conditioned on target state. ML target state estimates are defined using a specified target motion model and can be obtained numerically by gradient ascent procedures, which require computing the derivatives of the likelihood function with respect to target state variables, or by search procedures (e.g., genetic algorithms) which perform extensive enumeration of possible target state variables.

The second formulation is novel in that it assumes the measurements determine a sequence of statistically independent pre-detection "empirical" PDF's on target state. Multiplying these densities gives the overall empirical density on target state. EMAP target state estimates are defined by evaluating the empirical density for a specified class of parametric target motion models. EMAP estimates are computed numerically — without taking the gradient of the likelihood function — by the new expectation-maximization (EM) based algorithm derived in this paper. In contrast to the ML likelihood function, the EMAP likelihood function is a product of range marginals over a parameterized target density, called herein the geometric kernel. It will be seen that adjusting the down-range variance parameter of the kernel is an intuitive method for speeding up solution convergence and controlling solution accuracy.

The main contributions of this paper are a novel mathematical treatment of the observability issue for passive sonar TMA, first reported in [1], and the derivation of an iteratively re-weighted linear least squares target state estimation algorithm using the method of EM. Unlike the standard ML algorithm, this new algorithm does not require computing the gradient of the likelihood function in most cases of practical interest.

Section 2 develops notation and basic probabilistic structures used throughout the paper. Section 3 discusses the ML formulation of the bearings-only TMA problem for constant velocity targets. The EMAP formulation for constant velocity targets is presented in Section 4. Derivation of the EMAP estimation algorithm is given in Section 5. The asymptotic equivalence of the EMAP and ML approaches to the bearings-only TMA problem is shown in Section 6. Examples are given in Section 7. Summary and concluding remarks are given in Section 8.

2. NOTATION AND DEFINITIONS

The radiated sound field of a single point target impinges upon a sensor array, and the sensor signal processor generates an estimate of arrival angle at the sensor location. Arrival angles are estimates of azimuthal bearings, so the observable target coordinates lie in a horizontal plane. Let (x, y, z) denote target position at an arbitrary, but fixed, time. Azimuthal angles are measured in the x-y plane counter-clockwise from the positive x axis. Let $Z = \theta$ denote the measured bearing of the target when sensor position is $X^o = (x^o, y^o, z^o)$.

Both ML and EMAP formulations of the bearings-only TMA problem use a marginalization approach to avoid estimating target depth. Let the conditional PDF of bearing to the target be denoted by $p_{\theta|\mathbf{xyzx}^o\mathbf{y}^o\mathbf{z}^o}(\theta|x, y, z, x^o, y^o, z^o)$. Marginalizing over target depth gives

$$\int_{z_{\min}}^{z_{\max}} p_{\theta|\mathbf{xyzx}^o\mathbf{y}^o\mathbf{z}^o}(\theta|x, y, z, x^o, y^o, z^o)\, p_{\mathbf{z}}(z)\, dz \equiv p_{\theta|\mathbf{xyx}^o\mathbf{y}^o\mathbf{z}^o}(\theta|x, y, x^o, y^o, z^o), \quad (2.1)$$

where the second term in thc integrand denotes the *a priori* target depth PDF, which is zero outside the water column $[z_{\min}, z_{\max}]$ and is assumed known. If the target has a known fixed depth, then the *a priori* target depth density is simply the Dirac delta function located at the given target depth.

The marginal PDF (2.1) is called simply the bearing density throughout the sequel. In general, the bearing density is written

$$p_{\mathbf{z}|\mathbf{XX}^o}(Z|X, X^o) \equiv p_{\theta|\mathbf{xyx}^o\mathbf{y}^o\mathbf{z}^o}(\theta|x, y, x^o, y^o, z^o), \quad (2.2)$$

where target state is defined by $X = (x, y, s)$.

For later reference, the down-range marginal density of a PDF $f_{\mathbf{xy}}(x, y)$ with respect to the (one-dimensional) Cartesian variables x and y is defined by

$$f_\phi(\phi) \equiv \int_0^\infty f_{\mathbf{xy}}(\rho\cos\phi, \rho\sin\phi)\, \rho\, d\rho, \qquad -\pi < \phi \le \pi, \quad (2.3)$$

where (ρ, ϕ) are the polar coordinates of the target. The PDF $f_{\mathbf{xy}}(x, y)$ is said to be diffuse down-range if

$$f_{\mathbf{xy}}(\rho\cos\phi, \rho\sin\phi)\, \rho \propto \begin{cases} f_\phi(\phi), & \rho > 0, \\ 0, & \rho \le 0. \end{cases} \quad (2.4)$$

It follows from (2.4) that a diffuse down-range density is not necessarily diffuse in x and y.

3. Maximum Likelihood Formulation

Let $Z \equiv \{Z_n\}_{n=1}^N \equiv \{\theta_n\}_{n=1}^N$ denote a sequence of independent bearing measurements on the same target obtained at sensor locations $X^o \equiv \{X_n^o\}_{n=1}^N \equiv \{x_n^o, y_n^o, z_n^o\}_{n=1}^N$ and at times $\{t_n\}_{n=1}^N$. The measurements are not rcquired to be identically distributed. Let $X \equiv \{X_n\}_{n=1}^N \equiv \{x_n, y_n\}_{n=1}^N$ denote the sequence of target states at the measurement times $\{t_n\}_{n=1}^N$. Without loss of generality, it is supposed that $\{t_n\}_{n=1}^N$ are listed in increasing order, that is, $t_n \le t_{n+1}, n = 1, \ldots, N-1$.

As is typical in TMA problems, the measurements Z are assumed independent, conditioned on target state. The conditional likelihood function of Z is then given by

$$\mathcal{L}_{\mathbf{Z}|\mathbf{XX}^o}(Z\,|X, X^o) = \prod_{n=1}^N p_{\mathbf{Z}_n|\mathbf{X}_n\mathbf{X}_n^o}(Z_n\,|X_n, X_n^o). \quad (3.1)$$

A deterministic target motion model is specified, so that $x_n = x(t_n)$ and $y_n = y(t_n)$. The standard target motion model for bearings-only TMA is constant velocity, so

$$\begin{aligned} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= \frac{t_N - t}{t_N - t_1}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \frac{t - t_1}{t_N - t_1}\begin{bmatrix} x_N \\ y_N \end{bmatrix} \\ &\equiv \alpha(t)\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \beta(t)\begin{bmatrix} x_N \\ y_N \end{bmatrix} \end{aligned} \quad (3.2)$$

for $t_1 \le t \le t_N$. The model (3.2) uses end-point parameterization because the position parameters $\{x_1, y_1, x_N, y_N\}$ are dimensionally commensurate (a useful feature if Cramer-Rao lower bounds on estimation error are compared); however, the end-point model is mathematically equivalent to the more common position-velocity model. Target state is therefore fully parameterized by $\lambda = \{x_1, y_1, x_N, y_N\}$, so that $X = X(\lambda) = \{x(\lambda), y(\lambda)\}$. Hence, the target state estimate is determined from the ML parameter estimate

$$\hat{\lambda}_{\mathrm{ML}} = \arg\max_\lambda \mathcal{L}_{\mathbf{Z}|\mathbf{XX}^o}(Z\,|X(\lambda), X^o). \quad (3.3)$$

Taking the gradient of the conditional likelihood function with respect to λ and setting the result to zero gives the necessary conditions to be solved by appropriate numerical procedures.

A potentially serious difficulty with using the necessary ML conditions is that the gradient of the likelihood function is required with respect to the target parameter vector λ. Because the gradient is typically required at each iteration of a numerical procedure, ML estimates are often difficult and time consuming to compute. Also, the use of gradients may lead to unreliable search directions under weak observability conditions.

A maximum *a posteriori* (MAP) formulation can be obtained from the ML formulation by incorporating an appropriate *a priori* density on the end-point parameters of the constant velocity target model. Alternatively, a target process noise model can be included to compensate for target maneuvers. MAP estimators are not pursued further in this paper; however, the EMAP formulation presented in the next section is readily adaptcd to eithcr of these approaches to MAP estimation.

The traditional bearings-only TMA problem is obtained from (3.1) by assuming straight line propagation in the azimuthal plane. In this case, simple geometry and additive Gaussian noise assumptions give

$$p_{\theta_n|\mathbf{X}_n\mathbf{X}_n^o}(\theta_n\,|X_n, X_n^o) = \frac{1}{\sqrt{2\pi}\sigma_n} \times \exp\left\{\frac{-1}{2\sigma_n^2}\left[\theta_n - \tan^{-1}\left(\frac{y_n - y_n^o}{x_n - x_n^o}\right)\right]^2\right\}, \quad (3.4)$$

where σ_n denotes the standard deviation of θ_n. Substituting (3.4) into (3.1) and taking the natural logarithm gives the usual $1/\sigma_n^2$-weighted nonlinear least squares problem of traditional bearings-only TMA.

4. Empirical MAP Formulation

Traditional TMA statistical models are post-detection models, that is, they assume *a priori* that the measurements Z belong to a common target with motion of a specified parametric form (e.g., Equation (3.2)). Post-detection tracking implies that measurements are independent if they are conditioned on the target. ML estimators thus answer the question "Given data generated from a target track, which parameterized track best fits the data?"

In contrast, the EMAP estimators proposed here differ fundamentally from traditional post-detection TMA because they are joint detection/estimation methods which seek to answer the alternative question "Does a target track of the specified parametric form fit the data?" A generalized likelihood ratio test (GLRT) in which track paramcters are estimated and substituted into a likelihood ratio is the EMAP answer to the question; however, it is the estimated track — and not the GLRT detector — which is the object of interest in this paper.

The data $\{(Z_n, X_n^o)\}_{n=1}^N$ are assumed statistically independent because measurements are not specified *a priori* to belong to the same track. Independence implies that

$$p_{\mathbf{Z}\mathbf{X}^o}(Z, X^o) = \prod_{n=1}^{N} p_{\mathbf{Z}_n\mathbf{X}_n^o}(Z_n, X_n^o). \tag{4.1}$$

The data $\{(Z_n, X_n^o)\}_{n=1}^N$ contribute independent probability density assessments of "potential" target position that are valid at the times at which the measurements are obtained. Let $X^\Omega \equiv \{X_n^\Omega\}_{n=1}^N \equiv \{x_n^\Omega, y_n^\Omega\}_{n=1}^N$ denote so-called "empirical" random variables associated with potential locations. Empirical random variables are assumed indcpcndent when conditioned on their corresponding mcasurements and sensor locations; hencc, the empirical target location PDF for the full data set is

$$p_{\mathbf{X}^\Omega|\mathbf{Z}\mathbf{X}^o}(X^\Omega | Z, X^o) = \prod_{n=1}^{N} p_{\mathbf{X}_n^\Omega|\mathbf{Z}_n\mathbf{X}_n^o}(X_n^\Omega | Z_n, X_n^o). \tag{4.2}$$

The empirical target likelihood function (4.2) is evaluated for specified parametric target motion models, once the conditional density of X_n^Ω is defined.

Full target state is not observable from a single bearing mcasurement. Consequently, the dummy random variable r_n is introduced to model the "missing" sensor range measurement corresponding to θ_n, and the dcnsity of the empirical variable X_n^Ω is expressed as a marginal density over r_n. Using Bayes Theorem, thc marginal density is written in the form

$$p_{\mathbf{X}_n^\Omega|\mathbf{Z}_n\mathbf{X}_n^o}(X_n^\Omega | Z_n, X_n^o) = \int_0^\infty p_{\mathbf{x}_n^\Omega\mathbf{y}_n^\Omega|\mathbf{r}_n\mathbf{Z}_n\mathbf{X}_n^o}(x_n^\Omega, y_n^\Omega | r_n, Z_n, X_n^o) \times p_{\mathbf{r}_n|\mathbf{Z}_n\mathbf{X}_n^o}(r_n | Z_n, X_n^o)\, dr_n. \tag{4.3}$$

The statistical relationship between the missing range measurement r_n and the other random variables must be defined. Substituting the integral (4.3) into (4.2) gives the overall empirical PDF as a product of integrals.

Each term in the integrand of the integral representation (4.3) has a meaningful physical interpretation. The first term, called the geometric kernel, is a density on empirical target position, and it is conditioned on range, bearing, and sensor position. For bearings-only TMA problems, the geometric kernel is assumed to be a bivariate Gaussian whose mean vector and covariance matrix are determined by the conditioning variables. The kernel's mean vector is determined by range, bearing, and sensor location; the kernel's covariance matrix is a joint function of all the conditioning variables. Purely geometric considerations (see Appendix A of [1] for thc special case when the sensor lies at the origin) gives the kernel in the form

$$p_{\mathbf{x}_n^\Omega\mathbf{y}_n^\Omega|\mathbf{r}_n\mathbf{Z}_n\mathbf{X}_n^o}(x_n^\Omega, y_n^\Omega | r_n, Z_n, X_n^o) = \frac{1}{2\pi\sigma_n\kappa_n r_n^2} \times \exp\left\{ \frac{-1}{2r_n^2} \begin{bmatrix} x_n^\Omega - x_n^o - r_n\cos\theta_n \\ y_n^\Omega - y_n^o - r_n\sin\theta_n \end{bmatrix}' \Lambda(\theta_n, \sigma_n, \kappa_n) \times \begin{bmatrix} x_n^\Omega - x_n^o - r_n\cos\theta_n \\ y_n^\Omega - y_n^o - r_n\sin\theta_n \end{bmatrix} \right\}, \tag{4.4}$$

where $\Lambda(\theta_n, \sigma_n, \kappa_n)$ is the inverse covariance matrix. Loosely speaking, the cross-range and down-range variances of the geometric kernel are determined by the sensor and a free design parameter, respectively, together with square law azimuthal dispersion of empirical target location.

The standard deviation of the measured bearing θ_n is denoted by $\sigma_n \equiv \sigma(\theta_n)$, and it is determined by sensor signal processing considerations. The bearing variance $\sigma^2(\cdot)$ may be constant, but in general it depends parametrically on bearing because of beamwidth equalization issues (i.e., some beams may be narrower than others).

The standard deviation of the missing range measurement r_n, denoted in (4.4) by $\kappa_n \equiv \kappa(\theta_n)$, is a new and potentially useful free design parameter which can be used to control solution accuracy and rate of convergence. Using a down-range variance greater than κ_n^2 will increase the smearing of the down-range mixture components and, conversely, using a smaller variance will decrease the smearing. This opens the possibility in practice of using what may be called a "monotone" simulated annealing schcme in which the down-range variance is initially made too large to speed up convergence, and is then monotonically reduced in stages to the desired level of accuracy, namely κ_n^2.

The second density in the integrand of (4.3) specifies the sensor measurement window for the missing range measurement. It is derived by assuming that the *a priori* joint density of the measurement pair (r_n, θ_n) corresponds to a uniformly distributed point over a feasible region $\mathcal{R}(X_n^o)$ of the x-y plane whose inner and outer radii are given by the radial functions

$r_{\min}(\theta_n, X_n^o)$ and $r_{\max}(\theta_n, X_n^o)$, respectively. The outer radius may be interpreted as the maximum range at which signals of specified (maximum) source level are detected with specified probability P_d. Similarly, the inner radius may be interpreted as the near-field limit of the sensor, i.e., the minimum range at which the sensor's beamformer reliably estimates bearings. Thus, in polar coordinates, the joint density of (r_n, θ_n) is given by

$$p_{\mathbf{r}_n \theta_n | \mathbf{X}_n^o}(r_n, \theta_n | X_n^o) = \begin{cases} \mathcal{A}(X_n^o)\, r_n, & \text{for } (r_n, \theta_n) \in \mathcal{R}(X_n^o), \\ 0, & \text{otherwise,} \end{cases} \tag{4.5}$$

where $\mathcal{A}(X_n^o)$ is the reciprocal of the area of the feasible region $\mathcal{R}(X_n^o)$. Conditioning on bearing as well as sensor location gives, using Bayes Theorem,

$$p_{\mathbf{r}_n | \theta_n \mathbf{X}_n^o}(r_n | \theta_n, X_n^o) = \begin{cases} \varepsilon(\theta_n, X_n^o)\, r_n, & \text{for } r_{\min}(\theta_n, X_n^o) \le r_n \\ & \quad \le r_{\max}(\theta_n, X_n^o), \\ 0, & \text{otherwise,} \end{cases} \tag{4.6}$$

where the normalization constant is $\varepsilon(\theta_n, X_n^o) = 2/(r_{\max}^2(\theta_n, X_n^o) - r_{\min}^2(\theta_n, X_n^o))$. The *a priori* bearing density $p_{\theta_n | \mathbf{X}_n^o}(\theta_n | X_n^o)$ is uniformly distributed if and only if the normalization constant $\varepsilon(\theta_n, X_n^o)$ is independent of θ_n.

The integral representation appropriate for the bearings-only TMA problem is obtained from the general expression (4.3) by substituting the specific forms (4.4) and (4.6). The result is

$$p_{\mathbf{x}_n^\Omega \mathbf{y}_n^\Omega | \mathbf{Z}_n \mathbf{X}_n^o}(x_n^\Omega, y_n^\Omega | Z_n, X_n^o) = c(\theta_n, X_n^o) \times \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} \exp\left\{ \frac{-1}{2r_n^2} \begin{bmatrix} x_n^\Omega - x_n^o - r_n \cos\theta_n \\ y_n^\Omega - y_n^o - r_n \sin\theta_n \end{bmatrix}' \times \Lambda(\theta_n, \sigma_n, \kappa_n) \begin{bmatrix} x_n^\Omega - x_n^o - r_n \cos\theta_n \\ y_n^\Omega - y_n^o - r_n \sin\theta_n \end{bmatrix} \right\} \frac{dr_n}{r_n}, \tag{4.7}$$

where the proportionality factor is given by

$$c(\theta_n, X_n^o) = \frac{\varepsilon(\theta_n, X_n^o)}{2\pi\sigma(\theta_n)\kappa(\theta_n)}. \tag{4.8}$$

The integral representation (4.7) is fundamental to the formulation of the EMAP likelihood function.

Appendix B of [1] shows that the down-range marginal density of the integral (4.7) is closely approximated by the Gaussian distribution, provided the bearing measurement standard deviation σ_n is small, say on the order of several degrees or less. The result shows (with appropriate use of diffuse priors) that the traditional bearings-only TMA problem is recovered from (4.7) via down-range marginalization. The standard ML approach (see Section 3) to the bearings-only TMA problem is obtained asymptotically as $\kappa \to 0$, as shown in Section 6.

Let $\lambda = \{x_1, y_1, x_N, y_N\}$, as in the ML formulation of Section 3. Substituting the target model parameterization (3.2) into the integral representation (4.7), and then substituting the result into the likelihood function (4.2) gives

$$p_{\mathbf{x}^\Omega \mathbf{y}^\Omega | \mathbf{Z} \mathbf{X}^o}(x(\lambda), y(\lambda) | Z, X^o) = \left\{ \prod_{n=1}^{N} c(\theta_n, X_n^o) \right\} \times \left\{ \prod_{n=1}^{N} \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} \exp\left[\frac{-Q_n(r_n; \lambda)}{2r_n^2} \right] \frac{dr_n}{r_n} \right\}, \tag{4.9}$$

where the quadratic form $Q_n \equiv Q_n(r_n; \lambda)$ of the exponential term is

$$Q_n = \begin{bmatrix} \alpha(t_n)x_1 + \beta(t_n)x_N - x_n^o - r_n \cos\theta_n \\ \alpha(t_n)y_1 + \beta(t_n)y_N - y_n^o - r_n \sin\theta_n \end{bmatrix}' \times \Lambda(\theta_n, \sigma_n, \kappa_n) \times \begin{bmatrix} \alpha(t_n)x_1 + \beta(t_n)x_N - x_n^o - r_n \cos\theta_n \\ \alpha(t_n)y_1 + \beta(t_n)y_N - y_n^o - r_n \sin\theta_n \end{bmatrix}, \tag{4.10}$$

and where the bearing and down-range standard deviations of the geometric kernel are given by $\sigma_n \equiv \sigma(\theta_n)$ and $\kappa_n \equiv \kappa(\theta_n)$. The forms of $\sigma(\cdot)$ and $\kappa(\cdot)$ are derived from sensor signal processing considerations, and chosen for solution accuracy/rate of convergence criteria, respectively.

The likelihood function (4.9) yields the EMAP parameter estimate

$$\widehat{\lambda}_{\text{EMAP}} = \arg\max_{\lambda}\, p_{\mathbf{x}^\Omega \mathbf{y}^\Omega | \mathbf{Z} \mathbf{X}^o}(x(\lambda), y(\lambda) | Z, X^o). \tag{4.11}$$

The standard necessary conditions for the EMAP estimate (4.11) are found by setting the gradient of the posterior likelihood (4.9) with respect to the position parameters λ to zero.

5. EMAP Algorithm Derivation and Statement

A general EMAP estimation algorithm is derived in this section using the method of EM. Familiarity with the method of EM is assumed. General discussions of the method are widely available; for a general introduction, see [2, 3, 4]. For applications of the EM method specifically to Gaussian mixtures, see [5, 6, 7]. The EMAP algorithm is derived for constant velocity target motion; however, the derivation is very general and is easily extended to more general models.

The EMAP algorithm may be derived by discretizing the integrals in the likelihood function (4.9); however, discretization needlessly obscures the discussion. Instead, integrals are retained in the following derivation. The objective of the E-step is to define the so-called auxiliary function of the EM method and to simplify it if possible. The auxiliary function depends on two sets of target end-point parameter vectors, $\lambda' = \{x_1', y_1', x_N', y_N'\}$ and $\lambda = \{x_1, y_1, x_N, y_N\}$, where λ' is an initial (given) estimate and λ is arbitrary. The terms of the auxiliary function that are functions of λ are, using the quadratic form $Q_n(r_n; \lambda)$ defined by equation (4.10),

$$\Psi(\lambda) = -\frac{1}{2}\Psi_{\text{MSE}}(\lambda), \tag{5.1}$$

where

$$\Psi_{\rm MSE}(\lambda) = \sum_{n=1}^{N} \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} w_{\theta_n}(r_n; \lambda')\, Q_n(r_n; \lambda)\, \frac{dr_n}{r_n^2} \tag{5.2}$$

is a weighted mean squared error, and where the weights in (5.2) are given by the (Bayesian) ratio

$$w_{\theta_n}(r_n; \lambda') = \frac{\exp\left\{\frac{-Q_n(r_n;\lambda')}{2r_n^2}\right\} \frac{1}{r_n}}{\int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} \exp\left\{\frac{-Q_n(r_n;\lambda')}{2r_n^2}\right\} \frac{dr_n}{r_n}}. \tag{5.3}$$

The details of this important but tedious step are given in the Appendix 8. The weights (5.3) are nonnegative, so it is evident from (5.2) that $\Psi_{\rm MSE}(\lambda)$ is a nonnegative quadratic function of λ.

The objective of the M-step is to maximize the auxiliary function as a function of the parameter vector λ. Let the time dependent matrix $H(t)$ be given by the 2×4 matrix

$$H(t) = \begin{bmatrix} \alpha(t) & \beta(t) & 0 & 0 \\ 0 & 0 & \alpha(t) & \beta(t) \end{bmatrix} \tag{5.4}$$

(*cf.* Equation (3.2)). The weights (5.3) depend on the initial parameter vector λ', but not on λ; hence, setting the gradient of $\Psi(\lambda)$ with respect to λ to zero and solving for λ gives the updated parameter estimate

$$\begin{aligned} \lambda^+ &\equiv \arg\max_\lambda \Psi(\lambda) \equiv \arg\min_\lambda \Psi_{\rm MSE}(\lambda) \\ &= \left(\sum_{n=1}^{N} A_n(\theta_n)\right)^{-1} \left(\sum_{n=1}^{N} b_n(\theta_n)\right), \end{aligned} \tag{5.5}$$

where $A_n(\theta_n)$ is the 4×4 matrix given by

$$A_n(\theta_n) \equiv H'(t_n)\Lambda(\theta_n, \sigma_n, \kappa_n)H(t_n) \times \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} w_{\theta_n}(r_n; \lambda')\, \frac{dr_n}{r_n^2}, \tag{5.6}$$

and $b_n(\theta_n)$ is a vector of length 4 given by

$$b_n(\theta_n) \equiv H'(t_n)\Lambda(\theta_n, \sigma_n, \kappa_n) \times \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} w_{\theta_n}(r_n; \lambda') \begin{bmatrix} x_n^o + r_n \cos\theta_n \\ y_n^o + r_n \sin\theta_n \end{bmatrix} \frac{dr_n}{r_n^2}. \tag{5.7}$$

It is evident that the matrices given in (5.6) have rank at most 2, so the matrix whose inverse is required in (5.5) attains full rank if only if the target is observable (in the statistical sense) from the measured data set. Observability questions are widely discussed in bearings-only TMA problems, but lie outside the intended scope of this paper.

The EMAP algorithm is an iteratively re-weighted linear least squares algorithm. Explicitly, the algorithm takes the following recursive form:

1. Let $\sigma_n \equiv \sigma(\theta_n)$ and $\kappa_n \equiv \kappa(\theta_n)$ for $1 \le n \le N$.
2. Initialize the target end-point parameter

   $$\lambda^{(0)} = \left(x_1^{(0)}, y_1^{(0)}, x_N^{(0)}, y_N^{(0)}\right),$$

   and set $k = 0$.
3. For $k \ge 0$, define the unnormalized weight function

   $$\varpi_{\theta_n}(r_n; \lambda^{(k)}) = \exp\left\{\frac{-Q_n(r_n; \lambda^{(k)})}{2r_n^2}\right\}, \tag{5.8}$$

   where the quadratic form is defined by (4.10).
4. Compute the $3N$ one-dimensional integrals

   $$d_n^{(k)}(\ell) = \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} \varpi_{\theta_n}(r_n; \lambda^{(k)})\, \frac{dr_n}{r_n^\ell}, \quad 1 \le n \le N, \quad \ell = 1,2,3. \tag{5.9}$$

5. Using the integrals (5.9), compute the 4×4 matrix

   $$A^{(k)} \equiv \sum_{n=1}^{N} \frac{d_n^{(k)}(3)}{d_n^{(k)}(1)} H'(t_n)\Lambda(\theta_n, \sigma_n, \kappa_n)H(t_n) \tag{5.10}$$

   and the length 4 vector

   $$b^{(k)} \equiv \sum_{n=1}^{N} H'(t_n)\Lambda(\theta_n, \sigma_n, \kappa_n) \times \left(\frac{d_n^{(k)}(3)}{d_n^{(k)}(1)} \begin{bmatrix} x_n^o \\ y_n^o \end{bmatrix} + \frac{d_n^{(k)}(2)}{d_n^{(k)}(1)} \begin{bmatrix} \cos\theta_n \\ \sin\theta_n \end{bmatrix}\right). \tag{5.11}$$

6. Finally, compute the updated parameter vector

   $$\lambda^{(k+1)} = \left(A^{(k)}\right)^{-1} b^{(k)}. \tag{5.12}$$

Linear least squares problems such as (5.5) are in practice best solved by reliable methods of numerical matrix analysis, instead of using the normal equations (5.12). Details are left to the reader.

6. ML Approach as the Asymptotic Limit of EMAP

Laplace's method for obtaining asymptotic expansions of integrals is used in this section to reveal the close relationship between the EMAP integral representation (4.7) and the standard

Gaussian bearing error expression (3.4). Consider the generic Laplace-type integral

$$I(\lambda) = \int_a^b e^{-\lambda\phi(t)} f(t)\, dt, \tag{6.1}$$

where $\phi(t)$ is such that its absolute minimum on the interval $[a, b]$ occurs at the point $t = t_0$, where $a < t_0 < b$, $\phi'(t_0) = 0$, and $\phi''(t_0) > 0$. It is assumed that $f(t)$ has at least 2 and $\phi(t)$ has at least 4 continuous derivatives on $[a, b]$. The classical Laplace asymptotic expression, given by

$$I(\lambda) = \sqrt{\frac{2\pi}{\lambda\phi''(t_0)}}\, f(t_0)\, e^{-\lambda\phi(t_0)} + o\left(\frac{e^{-\lambda\phi(t_0)}}{\lambda^{\frac{3}{2}}}\right), \tag{6.2}$$

holds as $\lambda \to \infty$. It is derived rigorously using Watson's lemma in [8]. Expression (6.2) is merely the first term in an asymptotic expansion of the integral (6.1). The next term is not given in [8], but can be derived following their method (after fixing a minor numerical error in a coefficient). Omitting the tedious details, the result is

$$I(\lambda) = e^{-\lambda\phi(t_0)} \left\{\sqrt{2\pi}\, f(t_0)\, (\lambda\phi''(t_0))^{-\frac{1}{2}} + \sqrt{\pi}\, C\, (\lambda\phi''(t_0))^{-\frac{3}{2}}\right\} + o\left(\frac{e^{-\lambda\phi(t_0)}}{\lambda^{\frac{5}{2}}}\right), \tag{6.3}$$

where the coefficient C is given by

$$C = f(t_0)\left[\frac{1}{2}\left(\frac{\phi'''(t_0)}{\phi''(t_0)}\right)^2 - \frac{1}{3}\frac{\phi^{\text{IV}}(t_0)}{\phi''(t_0)}\right] - f'(t_0)\frac{\phi'''(t_0)}{\phi''(t_0)} + f''(t_0). \tag{6.4}$$

These asymptotic results are applied to integrals of the form (4.7) and (5.9).

The particular integrals of interest in this paper take the form

$$J_\ell(\kappa) = \frac{\gamma}{2\pi\sigma\kappa}\int_{r_1}^{r_2} r^{-\ell} \exp\left\{-\frac{a(r)}{\kappa^2} - \frac{b(r)}{\sigma^2}\right\} dr, \quad \ell = 1, 2, 3, \tag{6.5}$$

where

$$\begin{aligned} a(r) &= \frac{1}{2}\left[1 - \frac{x\cos\theta + y\sin\theta}{r}\right]^2, \\ b(r) &= \frac{1}{2}\left[\frac{x\sin\theta - y\cos\theta}{r}\right]^2. \end{aligned} \tag{6.6}$$

The special case $J_1(\kappa)$ is equivalent to (4.7), as is seen by letting

$$\begin{aligned} r &= r_n \\ r_1 &= r_{\min}(\theta_n, X_n^o) \\ r_2 &= r_{\max}(\theta_n, X_n^o) \\ x &= x_n^\Omega - x_n^o \\ y &= y_n^\Omega - y_n^o \\ \kappa &= \kappa_n = \kappa(\theta_n) \\ \sigma &= \sigma_n = \sigma(\theta_n) \\ \gamma &= 2\pi\sigma(\theta_n)\kappa(\theta_n)c(\theta_n, X_n^o) \end{aligned}$$

and by algebraically manipulating the quadratic form. The asymptotic form of (6.5) is sought as $\kappa \to 0$; hence, κ^{-2} plays the role of λ and $a(r)$ plays the role of $\phi(t)$ in (6.1). The necessary condition for the minimum of $a(r)$ is that its derivative be zero. The unique root of $a'(r) = 0$ is

$$r_0 = x\cos\theta + y\sin\theta. \tag{6.7}$$

It is assumed that r_0 is interior to the range of integration in (6.5). The second order condition is also satisfied, that is,

$$\left.\frac{d^2a(r)}{dr^2}\right|_{r=r_0} = \frac{1}{(x\cos\theta + y\sin\theta)^2} > 0. \tag{6.8}$$

The function

$$g(r) = \frac{\gamma}{2\pi\sigma\kappa} r^{-\ell} \exp\left\{-\frac{b(r)}{\sigma^2}\right\}$$

plays the role of $f(t)$ in (6.1). Using (6.2) and the fact that $a(r_0) = 0$ gives the asymptotic result

$$J_\ell(\kappa) = g(r_0)\sqrt{\frac{2\pi}{\kappa^{-2}a''(r_0)}}, \quad \kappa \to 0. \tag{6.9}$$

Substituting (6.7) and (6.8) into (6.9) and simplifying gives

$$J_\ell(\kappa) = \frac{\gamma}{\sqrt{2\pi}\sigma}\frac{1}{|x\cos\theta + y\sin\theta|^{\ell-1}} \times \exp\left\{\frac{-1}{2\sigma^2}\left(\frac{x\sin\theta - y\cos\theta}{x\cos\theta + y\sin\theta}\right)^2\right\}, \quad \kappa \to 0. \tag{6.10}$$

Because $(x\cos\theta + y\sin\theta, x\sin\theta - y\cos\theta)$ are the coordinates of the point (x, y) after rotating the coordinate system by θ,

$$\frac{x\sin\theta - y\cos\theta}{x\cos\theta + y\sin\theta} = \tan\left(\tan^{-1}\left(\frac{y}{x}\right) - \theta\right) \cong \tan^{-1}\left(\frac{y}{x}\right) - \theta, \tag{6.11}$$

where the approximation in (6.11) is valid for small bearing measurement errors. Substituting (6.11) into (6.10) gives the approximation

$$J_\ell(\kappa) \cong \frac{\gamma}{\sqrt{2\pi}\sigma}\frac{p(x\cos\theta + y\sin\theta)}{|x\cos\theta + y\sin\theta|^{\ell-1}} \times \exp\left\{\frac{-1}{2\sigma^2}\left(\theta - \tan^{-1}\left(\frac{y}{x}\right)\right)^2\right\}. \tag{6.12}$$

For $\ell = 1$, (6.12) is

$$J_1(\kappa) \cong \frac{\gamma}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-1}{2\sigma^2} \left(\theta - \tan^{-1}\left(\frac{y}{x}\right) \right)^2 \right\}. \quad (6.13)$$

The result (6.13) is important because it shows that the classical ML formulation of the bearings-only TMA problem is recovered in the limit as $\kappa \to 0$. Greater accuracy may be sought by using the next term in the asymptotic expansion.

Small measurement errors are typical in applications; however, if measurement errors are sufficiently large that the approximation (6.13) is inadequate, one may use the alternative density

$$J_1(\kappa) \cong \frac{\gamma}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-1}{2\sigma^2} \left(\frac{x \sin\theta - y\cos\theta}{x\cos\theta + y\sin\theta} \right)^2 \right\} \quad (6.14)$$

in the ML formulation. The model (6.14) is closely related to the pseudo-linear approximations used for traditional bearings-only TMA. The earliest references to approximations of this type are [9] and [10]; pseudo-linear methods are discussed extensively in [11], where an extensive bibliography is also given.

7. EXAMPLES

Triangulation

To illustrate application of the EMAP estimation algorithm of Section 5, consider first the example in Figure 1, which shows the sensor moving on a fixed course of 5° at a speed of 10 knots (≈ 5.14 meters/second) toward a fixed target initially 10,000 yards (9144 meters) away at a bearing of 0° (recall that all angles are measured counter-clockwise from the x axis). This is a classic triangulation problem in which the target parameters $\lambda = \{x, y\}$ are estimated from a series of azimuthal bearing measurements $Z = \{\theta_n\}_{n=1}^N$ taken at times $t = \{t_n\}_{n=1}^N$. In this example, the sensor takes bearing measurements at 1 minute intervals on a 10 minute leg for a total of 11 measurements.

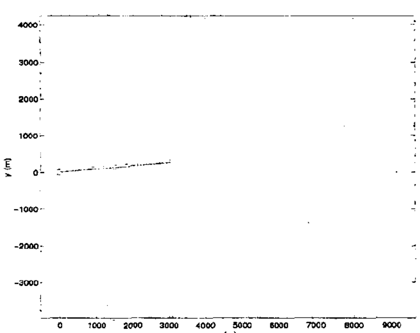


Figure 1: Classic triangulation example.

The E- and M-steps of the EMAP algorithm derived in the Appendix 8 are essentially unchanged for this reduced order triangulation problem. The target parameters to be estimated are the fixed locations $\lambda = \{x, y\}$. The quadratic terms $Q_{nk} \equiv Q_{nk}(r_{nk}; \lambda)$ in (A.3) become

$$Q_{nk} = \begin{bmatrix} x - x_n^o - r_{nk}\cos\theta_n \\ y - y_n^o - r_{nk}\sin\theta_n \end{bmatrix}' \Lambda(\theta_n, \sigma_n, \kappa_n) \times \begin{bmatrix} x - x_n^o - r_{nk}\cos\theta_n \\ y - y_n^o - r_{nk}\sin\theta_n \end{bmatrix}, \quad (7.1)$$

for the fixed target model. For triangulation, the matrix $H(t_n)$ defined in (A.22) reduces to the 2×2 identity matrix, the matrix (A.28) reduces to the 2×2 matrix

$$A_n(Z_n) \equiv \Lambda(\theta_n, \sigma_n, \kappa_n) \sum_{k=1}^{K_n} \frac{w_{Z_n}(r_{nk}; \lambda')}{r_{nk}^2}, \quad (7.2)$$

and the vector (A.29) reduces to the length 2 vector

$$b_n(Z_n) \equiv \Lambda(\theta_n, \sigma_n, \kappa_n) \sum_{k=1}^{K_n} \bar{v}_{nk}(r_{nk}) \frac{w_{Z_n}(r_{nk}; \lambda')}{r_{nk}^2}. \quad (7.3)$$

Figure 2 shows a plot of the likelihood function (A.2) in decibels referenced to the maximum likelihood value for various values of the target location parameters λ. For each n, $K_n = 79$ sample ranges, from 500 to 20,000 yards equispaced every 250 yards, were used for the summation in (A.2). The plot is generated with measurement standard deviations $\sigma = \sigma_0 r$ and $\kappa = \kappa_0 r$ (see Appendix A of [1]), with dimensionless standard deviations at $r = 1$ meter of $\sigma_0 = 0.0175$ and $\kappa_0 = 0.0873$ for each bearing measurement. The valuc of σ_0 corresponds to a bearing standard deviation of 1°. The values of σ and κ translate roughly to cross- and down-range standard deviations of 200 and 1000 yards, respectively, for a target 10,000 yards away.

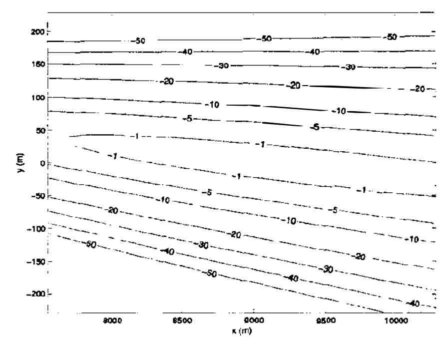


Figure 2: EMAP likelihood function (dB//max) for triangulation example.

The scatter plot of Figure 3 shows the estimation results for 250 Monte Carlo runs of the EMAP algorithm. For each run, new measurements $Z = \{\theta_n\}_{n=1}^{11}$ are generated, and the location parameters are initialized for a target at 15,000 yards with

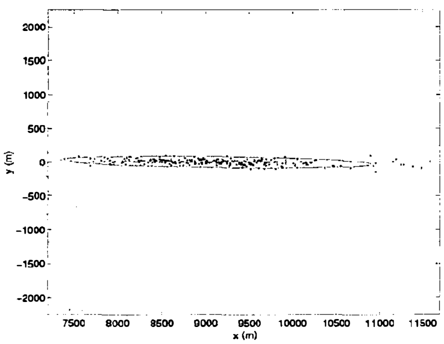


Figure 3: EMAP algorithm results for a 250 run Monte Carlo simulation. The EMAP and ML 90% containment ellipses are shown with solid and dashed lines, respectively.

a 45° bearing. The values $\sigma_0 = 0.0175$ and $\kappa_0 = 0.0873$ are used for the measurement standard deviations. An increase in the log-likelihood function (A.2) of less than 1×10^{-8} is used as the stopping criterion for the EMAP algorithm. The average number of iterations for each run is approximately 175. Included in the plot are the 90% containment ellipses based on the sample covariances of the EMAP and ML estimates, where the ML estimates (not plotted) are computed as described in Section 3. The two containment regions are indistinguishable. In fact, the EMAP estimates approach the ML estimates exactly as the down-range variance κ^2 is taken sufficiently small (see Section 6).

Constant Velocity Target

For a constant velocity target, the target parameters to be estimated are the end points of target motion, $\lambda = \{x_1, y_1, x_N, y_N\}$. Figure 4 shows the sensor moving on a fixed course of 5° at a speed of 10 knots on the first leg, followed by a fixed course of 95° at a speed of 10 knots on the second leg after a 1 minute simulated maneuver. The target starts 10,000

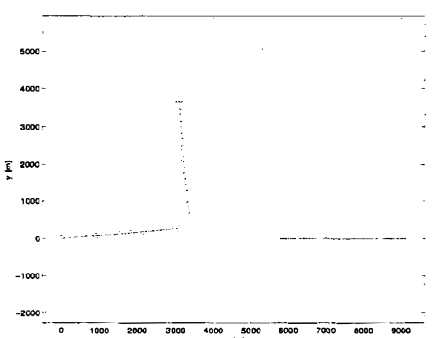
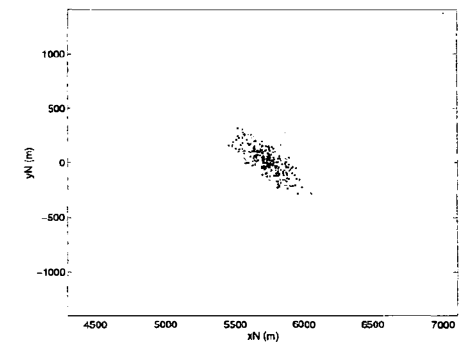


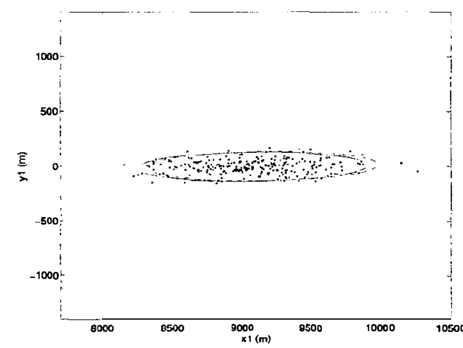
Figure 4: Constant velocity target example.

yards away at a bearing of 0°, and moves with a constant speed of 5 knots on a 180° course. The sensor takes bearing measurements at 1 minute intervals on both 10 minute legs for a total of 22 measurements.

Figure 5 shows the results of a 250 run Monte Carlo simulation of the EMAP algorithm for this problem using the same sample ranges, values for σ_0 and κ_0, and stopping criterion used for the triangulation example. For each run, new mea-



(a) Final location (x_N, y_N).

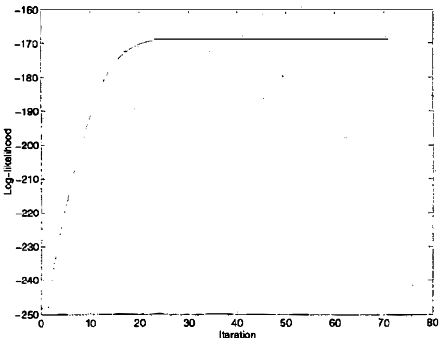


(b) Initial location (x_1, y_1).

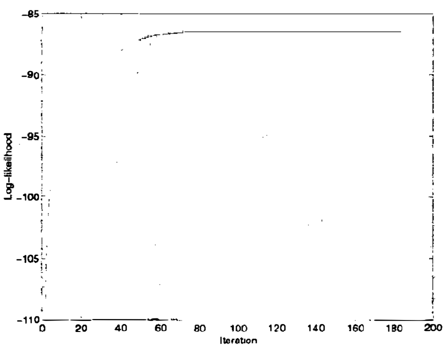
Figure 5: EMAP algorithm results for constant velocity target example. The EMAP and ML 90% containment ellipses are shown with solid and dashed lines, respectively.

surements $Z = \{\theta_n\}_{n=1}^{22}$ are generated, and the location parameters are initialized for a target moving from 15,000 yards away at a bearing of 45°, to a position 10,000 yards away at a 315° bearing, both with respect to the sensor's initial position. The average number of iterations for each run is approximately 70. The 90% containment ellipses based on the sample covariances of the EMAP estimates and the ML estimates (not plotted) are shown about the true target end points. As with the triangulation example, the EMAP and ML containment regions are nearly equivalent.

Figure 6 shows plots of the log-likelihood function (A.2) at each iteration for single runs of the EMAP algorithm for the constant velocity target and triangulation problems. The EM method guarantees that the log-likelihood will increase at each step. This property is an excellent consistency check in practice.



(a) Constant velocity target.



(b) Stationary target.

Figure 6: Log-likelihood functions for constant velocity target and triangulation examples.

Figure 7 shows plots of the weights $\{w_{\mathcal{Z}_1}(r_{1k};\lambda')\}_{k=1}^{79}$ down the first line of bearing at several iterations for a single run of the EMAP algorithm for the constant velocity target example. These weights describe the distribution of the components in the mixture for the first bearing. As the iterations increase, the peak of this distribution moves like a "breaking wave" closer to the component associated with the empirical target most likely to have generated the measurement. On the first iteration, the weight associated with the largest sampling range down the first line of bearing is essentially equal to 1 due to the poor initialization. The remaining weights are all nearly zero, as the sum of all the weights must add to 1 for each bearing (*cf.* Equation (A.12)). On the second iteration, the estimate of λ improves, and more mixture components contribute to the PDF of the target. The weight scales of the second and subsequent plots in Figure 7 are reduced to make the weight distributions more pronounced. Examination of the plot in Figure 7 for the final iteration of one run of the EMAP algorithm reveals that the standard deviation of the weight distribution is approximately 1000 yards, which is roughly the value of κ for the target range in this example. The down-range variance specifies a window over which the most significant components in

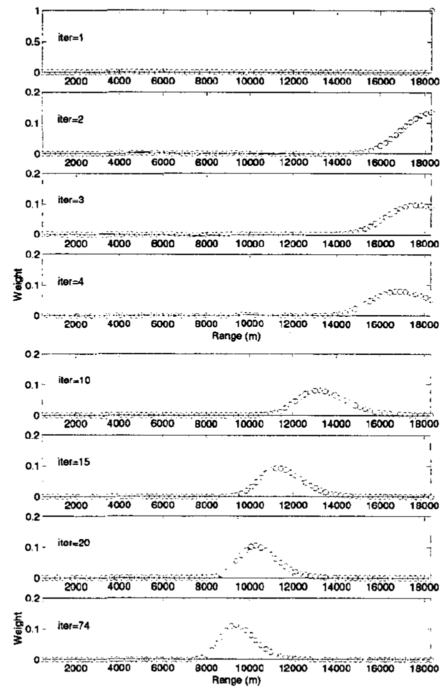


Figure 7: Down-range weights for first bearing at several iterations.

the mixture for each bearing are averaged.

Comments

The above examples are generalized for a maneuvering target by treating the maneuver time as an unknown parameter, and estimating it between EM iterations. This extension of the EM method is referred to as the generalized EM, or GEM, method, and is discussed in [4]. Maneuver time is estimated in the GEM framework by conducting a one dimensional search on maneuver time to increase the value of the auxiliary function between iterations. That is, maneuver time is chosen to maximize the auxiliary function over its value at the current target positional estimates, rather than to maximize the auxiliary function over the whole parameters space simultaneously. It is shown in [4] that the likelihood function does not decrease after an iteration of GEM, and that GEM converges if the likelihood function is bounded above. Though this approach to estimating maneuver time presents no great theoretical difficulties, the EMAP algorithm is more than just a simple iteratively re-weighted linear least squares algorithm in this case, and is not pursued further in this paper.

In the above examples, no attempt was made to speed up the EMAP algorithm convergence rate. The same "monotone" simulated annealing scheme suggested in Section 4 would speed up convergence at the beginning of the EM iterations, where initial estimates may be poor due to a bad initialization. Using non-equispaced sampling ranges and decreasing the number of ranges would also speed up convergence. Increasing the spacing down-range is justified, as the down-range standard deviation κ increases in proportion to range.

8. SUMMARY

Two formulations of the bearings-only TMA problem have been presented. One formulation is based on maximum likelihood, and it is classical in that it is a post-detection tracking approach in which measurements are conditioned on the target state. The other formulation is novel in that it is based on an empirical MAP method in which measurements are unconditionally independent because they are pre-detection measurements; that is, measurements are assumed unconditionally independent until proven conditionally independent by a detection decision. The EMAP approach is analogous to a GLRT method for simultaneous detection and track estimation.

An algorithm for solving the bearings-only TMA problem using the EMAP formulation is derived by the method of EM. The general EMAP algorithm is an iteratively re-weighted linear least squares algorithm, provided the target motion model is linear in the target motion parameters. Thus, the EMAP algorithm is a linear algorithm in most cases of practical interest. In the most general case, however, the EMAP algorithm is a nonlinear penalized least squares algorithm whose potential value in the application remains unexplored.

The empirical approach leads naturally to an integral representation of the measurement density in which uncertainty in range is compensated by adjusting the down-range variance κ^2 of the geometric kernel of the integrand. In effect, the proposed compensation averages the mixture components in a weighted sense over a sliding Gaussian window whose size, both down-range and cross-range, increases linearly with the down-range direction. It is shown that in the limit, as the size of the averaging window goes to zero, the EMAP approach is equivalent to the standard ML approach.

Triangulation and constant velocity target examples are presented to illustrate application of the EMAP algorithm. The EMAP algorithm estimates are nearly equivalent to the ML algorithm estimates for a reasonably sized value of the down-range variance κ^2. The EMAP algorithm is generalized to a maneuvering target by using the generalized EM method to estimate the maneuver time between EM iterations.

Different sensor types lead to different expressions for the geometric kernel and, hence, to different integral representations. For example, for linear arrays the angular measurement is conical angle, not azimuthal bearing; therefore, the geometric kernel generalizes to a trivariate Gaussian in (x, y, z) with one fixed variance corresponding to the cone angle measurement and two free variances. In this case, for unbounded isovelocity ocean models, the integral of the representation is over the surface of a cone with vertex at the acoustic center of the array, axis along the array, and half angle equal to the measured conical angle, instead of a line integral over a ray as for simple azimuthal bearings. For bounded non-isovelocity ocean models, the rays comprising the locus of the cone are distorted by internal refraction and boundary reflections into a manifold whose detailed structure is determined by the acoustic propagation model. Generalizing the geometric density term used in the EMAP formulation to other sensors with limited observability (in the statistical sense) would seem to present few intrinsic conceptual difficulties.

APPENDIX: EMAP AUXILIARY FUNCTION DERIVATION AND MAXIMIZATION

The EMAP auxiliary function is derived in this appendix. The derivation follows closely that of the EM based algorithm derived in the appendix of [7]. The EM method has two steps, the expectation step, or E-step, and the maximization step, or M-step. The E-step includes the so called "missing" data (unobserved range measurements in this application) in the likelihood function, and computes the expected value of the extended log-likelihood function with respect to this data to obtain the auxiliary function Ψ. The M-step maximizes Ψ over the parameters to be estimated. Successive application of these two steps is shown in [2, 3] to converge to a local maximum of the original likelihood function.

The end-point parameters $\lambda = \{x_1, y_1, x_N, y_N\}$ used in the constant velocity motion model (3.2) are the target parameters to be estimated. The likelihood function of interest is defined by

$$\mathcal{L}(Z, X^o | X(\lambda)) = \prod_{n=1}^{N} \int_{r_{\min}(\theta_n, X_n^o)}^{r_{\max}(\theta_n, X_n^o)} \exp\left[\frac{-Q_n(r_n; \lambda)}{2r_n^2}\right] \frac{dr_n}{r_n} \tag{A.1}$$

(*cf.* Equation (4.9)), the product of the (independent) posterior PDF's on empirical target location in (4.7) for each measurement (Z_n, X_n^o), where $X(\lambda) = \{x(\lambda), y(\lambda)\}$. The quadratic form $Q_n \equiv Q_n(r_n; \lambda)$ in the argument of the exponential term in the integrand of (A.1) is defined by the target motion model (3.2) and is given by (4.10).

Though not required for application of the EM method, for the purpose of implementation, the integral in (A.1) is approximated in this appendix by a sum over discrete values of range. Discretizing the continuous mixtures in the product (A.1) using equispaced intervals for each mixture gives the following discrete mixture representation for the likelihood function,

$$\mathcal{L}(Z, X^o | X(\lambda)) = \prod_{n=1}^{N} \sum_{k=1}^{K_n} \exp\left[\frac{-Q_{nk}(r_{nk}; \lambda)}{2r_{nk}^2}\right] \frac{1}{r_{nk}}, \tag{A.2}$$

where $r_{n1} \equiv r_{\min}(\theta_n, X_n^o)$ and $r_{nK_n} \equiv r_{\max}(\theta_n, X_n^o)$ correspond to the integration limits for the n-th integral. Thus, the likelihood function for this problem is defined by a product of discrete mixtures of continuous PDF's. The number of components in each mixture need not be the same; there are K_n components for measurement (Z_n, X_n^o) in (A.2). The quadratic term $Q_{nk} \equiv Q_{nk}(r_{nk}; \lambda)$ in (A.2) is given by

$$Q_{nk} = \begin{bmatrix} \alpha(t_n)x_1 + \beta(t_n)x_N - x_n^o - r_{nk}\cos\theta_n \\ \alpha(t_n)y_1 + \beta(t_n)y_N - y_n^o - r_{nk}\sin\theta_n \end{bmatrix}' \times \Lambda(\theta_n, \sigma_n, \kappa_n) \times \begin{bmatrix} \alpha(t_n)x_1 + \beta(t_n)x_N - x_n^o - r_{nk}\cos\theta_n \\ \alpha(t_n)y_1 + \beta(t_n)y_N - y_n^o - r_{nk}\sin\theta_n \end{bmatrix} \quad (A.3)$$

(*cf.* Equation (4.10)), which can be written more compactly in terms of the residual vector $\tilde{v}_{nk} \equiv \tilde{v}_{nk}(r_{nk}; \lambda)$,

$$Q_{nk} = \tilde{v}_{nk}'(r_{nk}; \lambda)\Lambda(\theta_n, \sigma_n, \kappa_n)\tilde{v}_{nk}(r_{nk}; \lambda), \quad (A.4)$$

where $\tilde{v}_{nk}(r_{nk}; \lambda) = v(t_n; \lambda) - \bar{v}_{nk}(r_{nk})$, the mean vector $\bar{v}_{nk}(r_{nk})$ is given by

$$\bar{v}_{nk}(r_{nk}) = \begin{bmatrix} x_n^o + r_{nk}\cos\theta_n \\ y_n^o + r_{nk}\sin\theta_n \end{bmatrix}, \quad (A.5)$$

and the vector $v(t_n; \lambda) = (x(t_n; \lambda), y(t_n; \lambda))$ is given by the target motion model (3.2), with the parameterization on λ made explicit.

Let $k_n \in \{1, \ldots, K_n\}$ be the index of the (bivariate) component of the n-th mixture in (A.2) that generated measurement $Z_n = \theta_n$. The index k_n represents the "missing" data in this problem; it specifies *exactly* the mixture component that gave rise to the realization of empirical target position from which the exact measurement $Z_n = \theta_n$ was made. The "complete" data set for this problem is the union of the original data set, $Z = \{\theta_n\}_{n=1}^N$, and the missing data set $K = \{k_n\}_{n=1}^N$,

$$Z \cup K = \{(\theta_n, k_n)\}_{n=1}^N. \quad (A.6)$$

If k_n is known for all n, the joint conditional PDF on empirical target location for the complete data set is defined by

$$\mathcal{L}(Z, K, X^o | X(\lambda)) = \prod_{n=1}^N \exp\left[\frac{-Q_{nk_n}(r_{nk_n}; \lambda)}{2r_{nk_n}^2}\right]\frac{1}{r_{nk_n}}. \quad (A.7)$$

From Bayes Theorem, the PDF of K is given by

$$\mathcal{K}(K|Z, X^o, X(\lambda)) = \frac{\mathcal{L}(Z, K, X^o|X(\lambda))}{\mathcal{L}(Z, X^o|X(\lambda))}. \quad (A.8)$$

Substituting (A.7) and (A.2) in (A.8) gives the ratio of the extended and original likelihood functions,

$$\mathcal{K}(K|Z, X^o, X(\lambda)) = \prod_{n=1}^N \frac{\exp\left\{\frac{-Q_{nk_n}(r_{nk_n};\lambda)}{2r_{nk_n}^2}\right\}\frac{1}{r_{nk_n}}}{\sum_{j=1}^{K_n}\exp\left\{\frac{-Q_{nj}(r_{nj};\lambda)}{2r_{nj}^2}\right\}\frac{1}{r_{nj}}} \quad (A.9)$$

$$= \prod_{n=1}^N w_{Z_n}(r_{nk_n}; \lambda), \quad (A.10)$$

where $w_{Z_n}(r_{nk_n}; \lambda)$ are the Bayesian terms in the product of (A.9) (*cf.* Equation (5.3)). The density (A.9) defines a discrete probability space on the missing data K; that is, $\mathcal{K}(K|Z, X^o, X(\lambda))$ is a probability mass function (PMF) on the integers $\{1, \ldots, K_n\}$ for all n. Defining the summation over K by the N-fold sum

$$\sum_K = \sum_{k_1=1}^{K_1} \cdots \sum_{k_N=1}^{K_N}, \quad (A.11)$$

the following two identities are easily obtained from (A.9),

$$\sum_K \mathcal{K}(K|Z, X^o, X(\lambda)) = 1, \quad (A.12)$$

$$\sum_{K\backslash k_i} \mathcal{K}(K|Z, X^o, X(\lambda)) = w_{Z_i}(r_{ik_i}; \lambda), \quad (A.13)$$

where the sum over $K\backslash k_i$ is equivalent to (A.11), with the sum over k_i omitted.

The auxiliary function of the EM method is defined by the expectation of the log of the extended likelihood function with respect to the missing data, conditioned on the original data and parameterized by a given initial estimate for the parameters to be estimated. For this problem, the auxiliary function is

$$\Psi(X(\lambda)|X(\lambda')) = \mathrm{E}_K\{\log \mathcal{L}(Z, K, X^o|X(\lambda))|Z, X^\bullet, X(\lambda')\}, \quad (A.14)$$

where the expectation is taken with respect to the missing data K, and λ' is a given initial estimate of the target parameters λ. By definition of expectation,

$$\Psi(X(\lambda)|X(\lambda')) = \sum_K [\log \mathcal{L}(Z, K, X^o|X(\lambda))] \times \mathcal{K}(K|Z, X^o, X(\lambda')). \quad (A.15)$$

Substituting (A.7) in (A.15), interchanging the summations, and applying the identity (A.13) gives

$$\begin{aligned}\Psi(X(\lambda)|X(\lambda')) &= \sum_K \sum_{n=1}^N \log\left\{\exp\left[\frac{-Q_{nk_n}(r_{nk_n};\lambda)}{2r_{nk_n}^2}\right]\frac{1}{r_{nk_n}}\right\} \times \mathcal{K}(K|Z, X^o, X(\lambda')) \\ &= \sum_{n=1}^N \sum_{k_n=1}^{K_n} \log\left\{\exp\left[\frac{-Q_{nk_n}(r_{nk_n};\lambda)}{2r_{nk_n}^2}\right]\frac{1}{r_{nk_n}}\right\} \times \sum_{K\backslash k_n} \mathcal{K}(K|Z, X^o, X(\lambda')) \\ &= \sum_{n=1}^N \sum_{k=1}^{K_n} \log\left\{\exp\left[\frac{-Q_{nk}(r_{nk};\lambda)}{2r_{nk}^2}\right]\frac{1}{r_{nk}}\right\} \times w_{Z_n}(r_{nk}; \lambda'). \end{aligned} \quad (A.16)$$

Taking the logarithm in (A.16) gives the auxiliary function as two double sums,

$$\Psi(X(\lambda)|X(\lambda')) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K_n}\frac{Q_{nk}(r_{nk};\lambda)}{r_{nk}^2}w_{Z_n}(r_{nk};\lambda') - \sum_{n=1}^{N}\sum_{k=1}^{K_n}\log(r_{nk})w_{Z_n}(r_{nk};\lambda'). \quad (A.17)$$

The first term in (A.17) (*cf.* Equation (5.2)) is the weighted mean squared error term $\Psi_{\rm MSE}(\lambda)$ in the generalized auxiliary function (5.1); it is the only term that is a function of λ, and the only term relevant to the M-step of the EM method.

The M-step maximizes the auxiliary function Ψ with respect to the parameters to be estimated to obtain an updated estimate. Each EM step guarantees an increase in the likelihood function, so successive applications of EM converge to a local maximum. This property suggests an iterative scheme for likelihood function maximization, where the monotonic increase in the likelihood function is used as a check against implementation error, and as a stopping criteria on the maximization. Defining the weighted mean squared error auxiliary function as

$$\Psi_{\rm MSE}(X(\lambda)|X(\lambda')) = \sum_{n=1}^{N}\sum_{k=1}^{K_n} w_{Z_n}(r_{nk};\lambda') \times Q_{nk}(r_{nk};\lambda)\frac{1}{r_{nk}^2}, \quad (A.18)$$

the M-step becomes

$$\lambda^+ \equiv \arg\max_{\lambda}\Psi(\lambda) \equiv \arg\min_{\lambda}\Psi_{\rm MSE}(\lambda), \quad (A.19)$$

where λ^+ are the updated parameter estimates, $\Psi(\lambda) \equiv \Psi(X(\lambda)|X(\lambda'))$, and $\Psi_{\rm MSE}(\lambda) \equiv -\Psi(\lambda)$. Substituting (A.18) in (A.19) gives

$$\begin{aligned}\lambda^+ &= \arg\min_{\lambda}\sum_{n=1}^{N}\sum_{k=1}^{K_n} w_{Z_n}(r_{nk};\lambda')\,Q_{nk}(r_{nk};\lambda)\,\frac{1}{r_{nk}^2}\\ &= \arg\min_{\lambda}\sum_{n=1}^{N}\sum_{k=1}^{K_n}\tilde{v}'_{nk}(r_{nk};\lambda)\,\Lambda(\theta_n,\sigma_n,\kappa_n)\\ &\qquad\times\tilde{v}_{nk}(r_{nk};\lambda)\,\frac{w_{Z_n}(r_{nk};\lambda')}{r_{nk}^2},\end{aligned} \quad (A.20)$$

where the quadratic term $Q_{nk}(r_{nk};\lambda)$ is written in terms of its components. From (A.4)–(A.5) and the target motion model (3.2), $\tilde{v}_{nk}(r_{nk};\lambda)$ is written

$$\tilde{v}_{nk}(r_{nk};\lambda) = H(t_n)\lambda - \bar{v}_{nk}(r_{nk}), \quad (A.21)$$

where

$$H(t_n) = \begin{bmatrix} \alpha(t_n) & \beta(t_n) & 0 & 0 \\ 0 & 0 & \alpha(t_n) & \beta(t_n) \end{bmatrix} \quad (A.22)$$

(*cf.* Equation (5.4)). Differentiating $\Psi_{\rm MSE}(\lambda)$ with respect to λ and setting the resulting partial derivatives equal to zero gives a necessary condition for the minimization in (A.20). The partial derivatives of the double sum in (A.20) with respect to the parameters λ are given by the chain rule for vector derivatives [12],

$$\frac{\partial\Psi_{\rm MSE}(\lambda)}{\partial\lambda} = \frac{\partial\tilde{v}_{nk}(r_{nk};\lambda)}{\partial\lambda}\frac{\partial\Psi_{\rm MSE}(\lambda)}{\partial\tilde{v}_{nk}(r_{nk};\lambda)}, \quad (A.23)$$

where

$$\frac{\partial\tilde{v}_{nk}(r_{nk};\lambda)}{\partial\lambda} = H'(t_n) \quad (A.24)$$

and

$$\frac{\partial\Psi_{\rm MSE}(\lambda)}{\partial\tilde{v}_{nk}(r_{nk};\lambda)} = 2\sum_{n=1}^{N}\sum_{k=1}^{K_n}\Lambda(\theta_n,\sigma_n,\kappa_n) \times \tilde{v}_k(t_n;\lambda)\,\frac{w_{Z_n}(r_{nk};\lambda')}{r_{nk}^2}. \quad (A.25)$$

Note that (A.23) satisfies the sufficient condition for λ^+ to be a minimum; that is, the (Hessian) matrix of second partial derivatives of (A.18) with respect to λ is positive definite, since the product of (A.24) and (A.25) yields a function of λ that is a sum of weighted quadratics, where the normalized downrange weights $w_{Z_n}(r_{nk};\lambda')/r_{nk}^2$ are all positive (see (A.9)–(A.10)), and the precision matrix $\Lambda(\theta_n,\sigma_n,\kappa_n)$ is positive definite. Substituting (A.24) and (A.25) in (A.23), setting the result equal to zero, and using (A.21) gives

$$\sum_{n=1}^{N}\sum_{k=1}^{K_n}H'(t_n)\Lambda(\theta_n,\sigma_n,\kappa_n) \times [H(t_n)\lambda^+ - \bar{v}_{nk}(r_{nk})]\,\frac{w_{Z_n}(r_{nk};\lambda')}{r_{nk}^2} = 0, \quad (A.26)$$

which reduces to a system of 4 linear equations in the 4 unknown parameters λ^+,

$$\left(\sum_{n=1}^{N}A_n(Z_n)\right)\lambda^+ = \sum_{n=1}^{N}b_n(Z_n), \quad (A.27)$$

where

$$A_n(Z_n) \equiv H'(t_n)\,\Lambda(\theta_n,\sigma_n,\kappa_n)\,H(t_n)\sum_{k=1}^{K_n}\frac{w_{Z_n}(r_{nk};\lambda')}{r_{nk}^2} \quad (A.28)$$

and

$$b_n(Z_n) \equiv H'(t_n)\,\Lambda(\theta_n,\sigma_n,\kappa_n)\sum_{k=1}^{K_n}\bar{v}_{nk}(r_{nk})\,\frac{w_{Z_n}(r_{nk};\lambda')}{r_{nk}^2}, \quad (A.29)$$

(*cf.* Equations (5.6)–(5.7)). The linear system (A.27) is easily solved by appropriate numerical linear algebra techniques (e.g. LU and QR factorizations, etc.).

Discretization of the likelihood function (A.1) is generalized for non-equispaced quadrature by inserting the factor a_{nk} for the width of the k-th interval of the n-th sum in the product (A.2), such that

$$\sum_{k=1}^{K_n} a_{nk} = r_{\max}(\theta_n, X_n^o) - r_{\min}(\theta_n, X_n^o). \qquad \text{(A.30)}$$

This generalization effects the definition of the extended likelihood function (A.7), which must include the factor a_{nk_n} for component k_n of mixture n. In terms of maximizing the weighted mean squared error auxiliary function (A.18), these extra factors only effect the weights $w_{Z_n}(r_{nk}; \lambda')$ defined in the E-step, which become

$$w_{Z_n}(r_{nk}; \lambda') = \frac{a_{nk} \exp\left\{ \frac{-Q_{nk}(r_{nk};\lambda')}{2r_{nk}^2} \right\} \frac{1}{r_{nk}}}{\sum_{j=1}^{K_n} a_{nj} \exp\left\{ \frac{-Q_{nj}(r_{nj};\lambda')}{2r_{nj}^2} \right\} \frac{1}{r_{nj}}} \qquad \text{(A.31)}$$

for this problem (*cf.* Equation (A.10) and (A.16)).

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors thank Marcus Graham and John Baylog, both with the Naval Undersea Warfare Center, for their many stimulating conversations and comments.

REFERENCES

[1] R. L. Streit, "Bearings-Only Target Motion Analysis With Propagation Loss Models of Uncertain Fidelity," Proceedings of the Third International Conference on Theoretical and Computational Acoustics, Newark, New Jersey, 14–18 July, 1997, to appear.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin,"Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp.1–38, 1977.

[3] C. F. J. Wu, "On the Convergence Properties of the EM Algorithm," Annals of Statistics, 11, 1983, pp. 95–103.

[4] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley & Sons, 1997.

[5] R. A. Redner and H. F. Walker, "Mixture Densities, Maximum Likelihood, and the EM Algorithm," SIAM Review, 26(2), 1984, pp. 195–239.

[6] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley & Sons, New York, 1985.

[7] R. L. Streit and T. E. Luginbuhl, "Maximum Likelihood Training of Probabilistic Neural Networks," IEEE Trans. on Neural Networks, NN-5, 1994, pp. 764–783.

[8] N. Bleistein and R. A. Handelsman, *Asymptotic Expansion of Integrals*, Holt, Rinehart, and Winston, New York, 1975.

[9] R. G. Stansfield, "Statistical Theory of D.F. Fixing," *J. IEE, London*, pt. 3A, vol. 94, no. 15, 1947, pp. 762–770.

[10] C. J. Ancker, "Airborne Direction Finding – The Theory of Navigation Errors," *IRE Trans. on Aeronautical and Navigational Electronics*, 1958, pp. 199-210.

[11] A. Holtsberg, *A Statistical Analysis of Bearings-Only Tracking*, Doctoral Dissertation, Lund Institute of Technology, Dept. of Mathematical Statistics, 1992.

[12] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Wiley & Sons, New York, 1981.

Roy L. Streit *received the BS degree (Honors) in physics and mathematics in 1968 from East Texas State University, the MA degree in mathematics in 1970 from the University of Missouri (Columbia), and the Ph. D. in mathematics in 1978 from the University of Rhode Island. From 1981 to 1982 he was a Visiting Scholar in the Department of Operations Research at Stanford University, Stanford, California, and from 1987 to 1989 he was an Exchange Scientist at the Defence Science and Technology Organisation, Adelaide, Australia. He joined the U. S. Navy Underwater Sound Laboratory in New London, Connecticut, in 1970, and he is now with the Naval Undersea Warfare Center (NUWC) in Newport, Rhode Island. His previous research includes diverse investigations supporting NUWC acoustic array R&D programs, applications of hidden Markov models to frequency line detection and tracking, and probabilistic methods for neural network design. His current research interests include distributed and sequential detection problems, multi-target tracking using advanced Bayesian methods, and multi-source acoustic data fusion using statistically validated belief networks.*

Michael J. Walsh *received the BA degree in physics from Bard College and the BS degree in mechanical engineering from Washington University, both in 1996 as part of a dual-degree program. In the summer of 1994, he was a student fellow at the Woods Hole Oceanographic Institution (WHOI) in the Ocean Systems and Moorings Laboratory, where he worked on hydrodynamic modeling of surface moorings. Mr. Walsh is currently employed by the Naval Undersea Warfare Center, where he works on target tracking applications for the Combat Systems Department. He is also working towards the MS degree in electrical engineering at the University of Massachusetts (Dartmouth). His current research interests include applications of estimation theory, probabilistic and statistical modeling, and numerical analysis to target tracking.*