RECURSIVE ESTIMATE-MAXIMIZE (EM) ALGORITHMS FOR TIME VARYING PARAMETERS WITH APPLICATIONS TO MULTIPLE TARGET TRACKING *

L. Frenkel and M. Feder
Dept. Electrical Engineering - Systems
Tel-Aviv University
Tel-Aviv, Israel

ABSTRACT

We investigate the application of EM algorithm to the classical problem of multiple target tracking (MTT) for a known number of targets. Conventional algorithms, have a computational complexity that depends exponentially on the targets' number, and usually divide the problem into a localization stage and a tracking stage. The new algorithms achieve a linear dependency, and integrate those two stages. Three major optimization criteria are proposed, using deterministic and stochastic dynamic models for the targets.

1. INTRODUCTION

The problem of tracking superimposed signals, embedded in noise, is important in sonar, radar, spectral estimation, and other fields. The observed data $y_1(t), y_1(t), \ldots, y_k(t), \ldots$, can be a non-linear noisy function of the tracks parameters $\theta_1^{(n)}, \theta_1^{(n)}, \ldots, \theta_k^{(n)}, \ldots$,

$$y_k(t) = \sum_{n=1}^{N} s_k^{(n)}(t, \theta_k^{(n)}) + n_k(t) \tag{1}$$

where each $y_k(t)$ is a vector of M signals that is composed of N superimposed signals $s_k^{(n)}(t, \theta_k^{(n)})$, and $n_k(t)$ is the observation noise. k is the snapshot index. In a sonar problem, for example, the data vector $y_k(t)$ is generated by an array of M sensors, sampled in time. This data is a function of the parameters, which can be the locations and velocities of the N targets, for each sample time.

The parameters $\theta_k^{(n)}$ themselves, can be modeled as a stochastic process, or as a deterministic vector. Any brute force attempt to write an algorithm that will produce an exact MSE, MAP or Maximum Likelihood criterion optimization for the estimation of these parameters, is unfeasible, and the produced algorithm will have computational complexity that is exponential with respect to the number of the snapshots and the targets. Consequently, the algorithms that are traditionally used for multiple target tracking (MTT), are not optimal. The problem is usually considered in two separate stages: *Localization*, in which the new parameters $\theta_k^{(n)}$ are estimated from the recent snapshot $y_k(t)$, and *tracking*, that makes use of some estimations from the localization part to produce the final track. Even the estimation in the localization part is usually not optimal (in a minimum estimation error sense), because optimality requires a computational complexity that depends exponentially on the number of the targets.

In this paper, we investigate the use of the EM algorithm to this classical problem of MTT with a known number of targets. The algorithms that are used, integrates the localization stage with the tracking stage, and achieve a linear computational complexity with respect to the targets number.

2. THE EM ALGORITHM

The EM algorithm iterates between a "complete data" estimation (E-step), and the parameters estimation (M-step). In our case, the complete data can be obtained by decomposing the observed data $y_k(t)$ into its signal components (see [3]),

$$x_k^{(n)}(t) = s_k^{(n)}(t, \theta_k^{(n)}) + n_k^{(n)}(t) \tag{2}$$

with constraints on the noise decomposition, to obtain

$$y_k(t) = \sum_{n=1}^{N} x_k^{(n)}(t) = H x_k(t), \quad H = [1 1 \ldots 1] \tag{3}$$

We utilize those previous results to develop the algorithms in the sequel. Three major approaches are proposed.

3. MAXIMUM LIKELIHOOD ESTIMATION

The first approach, uses a deterministic model for the parameters. A "forgetting" mechanism is used, to allow changes over time. Suppose $y_1(t), y_2(t), \ldots, y_k(t), \ldots$ are independent ergodic snapshots (the "incomplete" data), each with a probability density $f_Y(y_k, \theta_k)$. The parameters θ_k are unknown vectors that vary according to

$$\theta_k = F \theta_{k-1} \tag{4}$$

where F is a constant transition matrix. According to this model, estimating the last parameter θ_K, will automatically drive the estimation of all the preceding parameters. Consequently, the parameters estimation can be accomplished by a multi-parameter *maximum likelihood* search on θ_K.

Using the computationally efficient recursive EM algorithm, the search over the targets is avoided. Let $x_1, x_2, \ldots x_k, \ldots$ represent an independent "complete" data, that is related to the observations by (3). The batch EM algorithm starts

* This work has been supported by the US-Israel Binational Science Foundation (BSF).

0-7803-2431-5/95 $4.00 © 1995 IEEE

with an arbitrary initial guess $\theta_K^{(0)}$, and after l iterations estimates θ_K by $\hat{\theta}_K^{(l)}$. Each iteration cycle is can be described by two steps:

E: Evaluate

$$Q(\theta_K,\hat{\theta}_K^{(l)}) = E_{\hat{\theta}_K^{(l)}}\{\log f(x_1,\dots,x_K;\theta_1,\dots,\theta_K) \mid y_1,\dots,y_K\}$$
$$= \sum_{l=1}^{K} E_{\hat{\theta}_K^{(l)}}\{\log f(x_l,\mathbf{F}^{l-K}\theta_K) \mid y_1,\dots,y_K\}$$
$$= \sum_{l=1}^{K} E_{\hat{\theta}_K^{(l)}}\{\log f(x_l,\mathbf{F}^{l-K}\theta_K) \mid y_l\}$$

M: $\mathbf{Max}_{\theta_K}\ Q(\theta_K,\hat{\theta}_K^{(l)}) \rightarrow \hat{\theta}_K^{(l+1)}$

where $E_{\hat{\theta}_K^{(l)}}\{\cdot\}$ denotes a statistical expectation with respect to the last parameter estimation. In each iteration all the data has to be processed.

In [2] Titterington has suggested a sequental algorithm for the case where the parameters are constant. Based on this approach, in our case (of time varying parameters) we obtain the following sequential algorithm:

$$\text{E: } Q(\theta_{k+1},\hat{\theta}_k) = L_{k+1}(\theta_{k+1}) = \tag{5}$$
$$= \sum_{l=1}^{k+1} \gamma^{k+1-l} E_{\mathbf{F}\hat{\theta}_l}\{\log f(x_l,\mathbf{F}^{l-k-1}\theta_k)/y_l\}$$
$$= \gamma L_k(F^{+}\theta_{k+1}) + E_{\mathbf{F}\hat{\theta}_k}\{\log f(x_{k+1},\theta_{k+1})/y_{k+1}\}$$

$$\text{M: } \mathbf{Max}_{\theta_{k+1}}\ Q(\theta_{k+1},\hat{\theta}_k) \rightarrow \hat{\theta}_{k+1} \tag{6}$$

where each of the statistical expectations of the log-likelihood is done once, and used for the next expectation. The constant γ was suggested in [1]. For varying parameters, γ is expected to be a tradeoff between good tracking ability (γ small) and noise insensitivity ($\gamma = 1$).

It can be proven, that by applying a Newton second order approximation, we obtain a new, fast stochastic algorithm, that benefits from the given model:

The Recursive EM-Newton algorithm:	
$\hat{\theta}_{k+1} = F\hat{\theta}_k + Ic_{k+1}^{-1} S(y_{k+1},\mathbf{F}\hat{\theta}_k)$	(7)
$Ic_{k+1} = \gamma F^{-1\,T} Ic_k F^{-1} + Ic(F\hat{\theta}_k)$	(8)

where $S(y_k,\theta_k)$ denotes the score vector, and $Ic(\theta_k)$ is Fisher information matrix corresponding to a single complete snapshot, that is

$$S(y_k,\theta_k) = \nabla_{\theta_k} \log f_Y(y_k;\theta_k)$$
$$Ic(\theta_k) = -E_{\theta k}\{\nabla^2_{\theta_k} \log f_X(x_k;\theta_k)\}$$

Notice, that for *F=Identity* (constant parameters), and for $\gamma = 1$ (no forgetting factor), this algorithm becomes

$$\bar{\theta}_{k+1} = \bar{\theta}_k + \left[\sum_{l=1}^{k} Ic(\hat{\theta}_l)\right]^{-1} S(y_{k+1},\hat{\theta}_k)$$

and for a small change of $Ic(\hat{\theta}_l)$ near the ML estimation θ, we obtain, approximately,

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left[k\, Ic(\hat{\theta}_l)\right]^{-1} S(y_{k+1},\hat{\theta}_k))$$

which is a stochastic approximation algorithm, suggested by Titterington [2] for constant parameters. The conventional approach to handle time varying parameters (see [9]), is to substitute the converging series ($1/k$) with a small positive constant γ_0 in the algorithm obtained for constant parameters, that is,

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_0\left[Ic(\hat{\theta}_l)\right]^{-1} S(y_{k+1},\hat{\theta}_k) \tag{9}$$

This ad hoc procedure can now be replaced by the new EM-Newton algorithm, in order to utilize a dynamic model for the parameters

4. MSE ESTIMATION

The second approach, uses a Bayesian model, with a *MSE estimation* criterion. A natural approach to the problem of tracking superimposed signals is to model the track parameters as a stochastic process. A reasonable choice would be

$$\theta_k^{(n)} = F\theta_{k-1}^{(n)} + q_k \tag{10}$$

where q_k is the driving noise of the parameters process, and $\theta_k^{(n)}$ is the parameters describing the n-th signal the k-th snapshot.

The estimation that minimizes the MSE is

$$\hat{\theta} = E\{\theta \mid y_1 \dots y_k\} \tag{11}$$

This estimation can be accomplish by the Kalman filter, if $s_k^{(n)}(t,\theta_k^{(n)})$ in (1) is linear. Unfortunately, this is usually not the case, and a complete integration of the localization and the tracking is not possible. Using the EM algorithm for localization and a Kalman filter for tracking, with feedback to the localization, produces a fast high resolution algorithm:

The EM-Kalman algorithm

1. Initial states:

for $n = 1,2, \dots, N$ (signals):
Guess state #0, $\hat{\theta}_{0|0}^{(n)}$

Set initial error variance, $P_{1|0}$

For k=1,2,...

2. Localization:

for p = 1,2, ..., Np (iterations number):
for n = 1,2, ..., N (signals):

E-step:

$$\hat{x}_k^{(n)} = s_k^{(n)}(t,\hat{\theta}_{k|k-1}^{(n)}) + \beta_n\left[y_k(t) - \sum_{l=1}^{N} s_k^{(n)}(t,\hat{\theta}_{k|k-1}^{(n)}) \right] \quad (12)$$

where $\sum \beta_n = 1$

M-step:

$$\tilde{\theta}_k^{(n)} = H \cdot \mathbf{argmin}_{\theta_k^{(n)}} \int_T [\hat{x}_k^{(n)}(t) - s_k^{(n)}(t,\theta_k^{(n)})]^* R^{-1} [\hat{x}_k^{(n)}(t) - s_k^{(n)}(t,\theta_k^{(n)})] \quad (13)$$

3. Kalman tracking:

for n = 1,2, ..., N (signals):

$$K_k^{(n)} = P_{k/k-1}^{(n)} H^T \left[H P_{k/k-1}^{(n)} H^T + R_{EM_k} \right]$$

$$\hat{\theta}_k^{(n)} = \hat{\theta}_{k/k-1}^{(n)} - K_k^{(n)} [\tilde{\theta}_k^{(n)} - H\hat{\theta}_{k/k-1}^{(n)}]$$

$$P_k^{(n)} = [I - K_k H] P_{k/k-1}^{(n)}$$

$$\hat{\theta}_{k+1/k}^{(n)} = F\hat{\theta}_k^{(n)}$$

$$P_{k+1|k}^{(n)} = F P_k^{(n)} F^T + Q$$

At the *localization* stage, some of the parameters are estimated from the last snapshot by the EM algorithm. Denote these estimations as $\tilde{\theta}_k^{(n)}$, which estimates the vector $H\theta_k^{(n)}$. The initial state for the EM iterations was predicted by the tracking stage. This prediction is based on the previous estimations, and therefore, is close to the actual parameter. Thus, a few EM iterations may be sufficient

Since the complete data $x_k(t)$ and the incomplete data $y_k(t)$ are jointly Gaussian, related by a linear transformation (5), the estimation (12), and the log likelihood (13) can be calculated in a straight forward manner (see [3]).

At the *tracking* stage, the EM estimations are considered as a noisy observations of the real parameters

$$\tilde{\theta}_k^{(n)} = H\theta_k^{(n)} + n_{EM_k} \quad (14)$$

The driving noise and the EM-estimation noise are modeled as Gaussian, with covariance matrices $Q_k = E\left\{q_k q_k^*\right\}$ and $R_{EM_k} = E\left\{n_{EM_k} n_{EM_k}^*\right\}$. Kalman filter is used to track $\tilde{\theta}_k^{(n)}$.

5. MAP ESTIMATION

The third approach, uses the a Bayesian model, with a *MAP estimation* criterion. The parameters are modeled as the discrete stochastic process (10). An exact solution can be achieved using the Hidden Markov Model (HMM), (see [6], [7] for frequency line tracking). However, the HMM algorithm does not reduce the exponential dependence of the computational complexity on the targets' number, and in a typical sonar problem, for example, tracking even two targets may be unfeasible. Integrating the EM algorithm with the HMM, eliminates this problem, and a fast, powerful algorithm, that converges to the optimal solution, is produced.

The batch EM-HMM algorithm:

1. Initial track guess

for n = 1,2, ..., N (signals):
guess track #0, $\hat{\theta}_{k,0}^{(n)}$ $\quad (1 \le k \le K)$

for p = 1,2, ..., Np :

2. Recursion:

for k =1,2, ...,K (snapshots):
for n = 1,2, ..., N (signals):

E-step:

$$\hat{x}_{k,p}^{(n)} = s_k^{(n)}(t,\hat{\theta}_{k,p-1}^{(n)}) + \beta_n\left[y_k(t) - \sum_{l=1}^{N} s_k^{(n)}(t,\hat{\theta}_{k,p-1}^{(n)}) \right]$$

where $\sum \beta_n = 1$ $\quad (1 \le k \le K)$

M-step:

for j = 1,2, ..., Ns (new state)

$$\delta_k^{(n)}(j) = \begin{cases} \pi_j b_j(x_{1,p}^{(n)}(t)) \\ \mathbf{max}_{S_i,1\le i\le N_s}\left[\delta_{k-1}^{(n)}(S_i)a_{ij}\right] b_j(x_{k,p}^{(n)}(t)) \end{cases}$$

$$\psi_k^{(n)}(S_j) = \mathbf{argmax}_{S_i,1\le i\le N_s}\left[\delta_{k-1}^{(n)}(S_i)a_{ij}\right] \quad (2 \le k \le K)$$

3. Termination:

for n = 1,2, ..., N (signals):

$$i_{K,p}^{(n)} = \mathbf{argmax}_{1\le i\le N_s}\left[\delta_K^{(n)}(i)\right], \quad \hat{\theta}_{K,p}^{(n)} = S_{i_{K,p}^{(n)}} \quad (2 \le k \le K)$$

4. Backtracking

for k = K-1,K-2, ...,1 (snapshots)
for n = 1,2, ..., N (signals):

$$i_{k,p}^{(n)} = \psi_{k+1}^{(n)}(i_{k+1,p}^{(n)}), \quad \hat{\theta}_{k,p}^{(n)} = S_{i_{k,p}^{(n)}}$$

The heart of this HMM algorithm is the Viterby algorithm, in which the new state probability distribution for states j=1...Ns, at time k ($\delta_k^{(n)}(j)$), is calculated from the one saved for time k-1, and from the new complete data $x_k^{(n)}(t)$ [5].

$A=\{a_{ij}\}$ is the state transition matrix for one signal from state $\theta_k^{(n)}=S_i$ to state $\theta_{k+1}^{(n)}=S_j$, where

$$a_{ij}=P\left\{\theta_{k+1}^{(n)}=S_j|\theta_k^{(n)}=S_i\right\}$$
$$=c\cdot\int_{s_j\in \mathcal{S}_j}\exp\{-\tfrac{1}{2}[s_j-FS_i]^*Q^{-1}[s_j-FS_i]\}\,ds_j$$

The initial state distribution vector is $\Pi=\{\pi_i\}$, where $\pi_i=P\left\{\theta_0^{(n)}=S_i\right\}$. $B=\{b_j(x_k^{(n)}(t))\}$, where $b_j(x_k^{(n)}(t))$ is the continuous probability distribution of $x_k^{(n)}(t)$ given state $\theta_k^{(n)}=S_j$, that is

$$b_j(x_k^{(n)}(t))= \tag{15}$$
$$=f_{X_k(t)}(x_k^{(n)}(t)\mid\theta_k^{(n)}=S_j)$$
$$=c\cdot\exp\left\{-\int_T[x_k^{(n)}(t)-s_k^{(n)}(t,S_j)]^*R^{-1}[x_k^{(n)}(t)-s_k^{(n)}(t,S_j)]\right\}$$

The probability measures $\boldsymbol{A}$, $\boldsymbol{B}$, Π and the states number Ns, define an HMM that produces the MAP estimation for the n-th signal, given the complete data as observations,

$$\hat{\theta}^{(n)}=\max_{\theta^{(n)}}\left\{f(\theta^{(n)}\mid x_1^{(n)}\ldots x_k^{(n)})\right\} \tag{16}$$

and thus constitutes an M-step in the EM algorithm, that takes all the snapshots into account.

6. SIMULATION

Simulations were performed for the case of a sonar with linear array of omnidirectional sensors. The algorithms achieve a high resolution in range and azimuth, in low SNR, and random phase. The EM-Kalman algorithm was implemented for the FTV 3-dimensional sonar [4], and tested with real data.

Fig. 1 shows the first snapshot of two Hanning signals embedded in noise. The signals' time delay represents the targets' ranges.

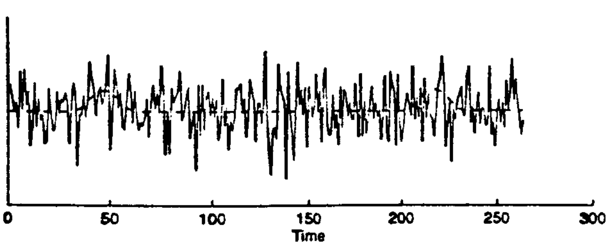

Figure 1: The first snapshot. the original signal (--), and the noisy signal (-)

In the following snapshots the ranges change, and targets maneuver in sinus-like tracks. Fig. 2 shows the time delay estimations of the conventional stochastic approximation algorithm (9). Fig. 3 shows the EM-Newton algorithm (7)-(8) estimates, using a linear model for the track. Clearly, for this case there is an advantage of using the model-based EM-Newton algorithm.

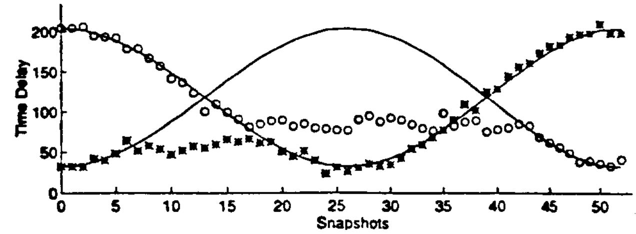

Figure 2: The actual track (-) and a conventional stochastic approximation algorithm estimations (o,*)

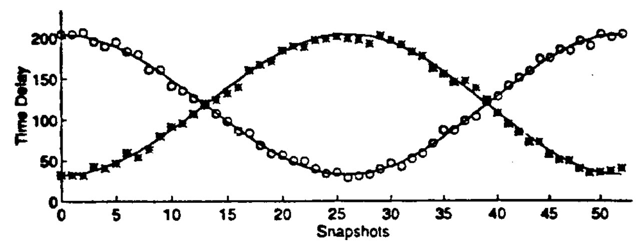

Figure 3: The actual track (-) and the EM-Newton algorithm estimates (o,*)

7. REFERENCES

[1] E. Weinstein M. Feder "Sequential algorithms for parameter estimation based on the Kullbak-Leibler information matrix," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.38, pp.1652-1654, Sep. 1990

[2] D. M. Titterington "Recursive parameter estimation using incomplete data ," *J. Roy. Stat. SOC.*, vol. 46(B), pp. 256-267, 1984.

[3] M. Feder E. Weinstein and A. V. Oppenheim, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Signal Processing*, vol.36, pp. 477-489, Apr. 1988

[4] J. S. Jaffe, E. Reuss, D. McGehee, G. Chandran, "FTV, a sonar for tracking zooplankton in 3-dimensions," Marine Physical Lab, Scripps Institution Oceanography, 1993

[5] L. R. Rabiner, "A tutorial on hidden Markov model and selected applications in speech recognition," *Proc. IEEE*, vol. 77, No. 2, pp. 257-286, Feb. 1989

[6] R. L. Streit and R. F. Barrett, "Frequency line tracking using hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, No. 4, pp. 586-598, Apr. 1990

[7] X. Xie and R. J. Evans , *"Multiple Target Tracking and Multiple Frequency Line Tracking Using Hidden Markov Models,"* *IEEE Trans. Signal Processing* , vol. 39, No. 12, pp. 2659-2676, Dec. 1991

[8] S. S. Blackman , *Multiple-Target Tracking with Radar Applications*. Artech House, 1986

[9] L. Ljung and Torsten Soderstrom, *Theory and Practice of Recursive Identification*. The MIT Press, 1987