Expectation Maximisation for Sensor Data Fusion

Felix Opitz
Defence and Communications Systems
EADS
Wörthstr.85
89077 Ulm, Germany

felix.opitz@eads.com

Abstract: The expectation maximisation algorithm offers several applications in sensor data fusion. An overview of some of this applications and a short course in expectation maximisation algorithm and its properties is given.

The expectation maximisation algorithm (EM) was introduced by Dempster, Laird and Rubin in 1977 [DLR77]. The basic of expextation maximisation is maximum likelihood estimation (MLE). In modern sensor data fusion expectation maximisation becomes a substantial part in several applications, e.g. multi target tracking with probabilistic multi hypothesis tracking (PMHT), target extraction within probability hypothesis density (PHD) filter, cluster analysis within multidimensional data association, or image computing.

1 Maximum likelihood estimation

Let $\mathcal{Z}$ be a random vector with probability density function $p(\mathcal{Z} \mid \mathcal{X})$ depending on a parameter $\mathcal{X} \in \Omega$. The aim is to estimate the parameter $\mathcal{X}$ using the observation $\mathcal{Z}$. Therefore, the likelihood function $L(\mathcal{X}) = p(\mathcal{Z} \mid \mathcal{X})$ is formed. An estimation $\hat{\mathcal{X}}$ of $\mathcal{X}$ is obtained by the solution of

$$\left.\frac{\partial L(\mathcal{X})}{\partial \mathcal{X}}\right|_{\hat{\mathcal{X}}} = 0 \Longleftrightarrow \left.\frac{\partial \log L(\mathcal{X})}{\partial \mathcal{X}}\right|_{\hat{\mathcal{X}}} = 0 \tag{1}$$

The Fisher information matrix is given by

$$I(\mathcal{X} \mid \mathcal{Z}) = E_{\mathcal{X}}\left[\frac{\partial^2 \log L(\mathcal{X})}{\partial \mathcal{X} \partial \mathcal{X}^T}\right] = E_{\mathcal{X}}\left[S(\mathcal{Z} \mid \mathcal{X}) S^T(\mathcal{Z} \mid \mathcal{X})\right] \tag{2}$$

with the so called score statistic $S(\mathcal{Z} \mid \mathcal{X}) = \frac{\partial \log L(\mathcal{X})}{\partial \mathcal{X}}$ based on the observed data $\mathcal{Z}$. It is well known, that the asymptotic covariance matrix of the estimation $\hat{\mathcal{X}}$ can be approxiamted by $I^{-1}(\mathcal{X} \mid \mathcal{Z})$.

2 Expectation Maximisation for Maximum Likelihood Estimation

The expectation maximisation is well suited when the observed data vector $\mathcal{Z}$ is incomplete but the maximum likelihood estimation is straightforward for an augmented (complete) observation vector $\mathcal{W}$. Another application area is, when the complete log likelihood function cannot be maximised analytically.

Suppose now, that incomplete data $\mathcal{Z}$ is given and one wishes to maximise $L(\mathcal{X}) = p(\mathcal{Z} \mid \mathcal{X})$. Let $\mathcal{W}$ denote the complete version of $\mathcal{Z}$. i.e. $\mathcal{W}^T = (\mathcal{Z}^T, \mathcal{A}^T)$, with explicitly known probability density function $p(\mathcal{W} \mid \mathcal{X})$. The augmentation part is denoted by $\mathcal{A}$. It includes the non observable states. Then the EM procedure is as follows [MK97], [Web99]:

$$L(\mathcal{X}) = p(\mathcal{Z} \mid \mathcal{X}) = \int p(\mathcal{Z}, \mathcal{A} \mid \mathcal{X}) d\mathcal{A} \tag{3}$$

The EM procedure generates a sequence of estimates of $\mathcal{X}$, $\{\mathcal{X}^{(m)}\}$, from an initial estimate by two steps:

E-step: Evaluate

$$\begin{aligned} Q(\mathcal{X}, \mathcal{X}^{(m)}) &= E[log(p(\mathcal{W} \mid \mathcal{X})) \mid \mathcal{Z}, \mathcal{X}^{(m)}] \\ &= \int \log(p(\mathcal{Z}, \mathcal{A} \mid \mathcal{X})) p(\mathcal{A} \mid \mathcal{Z}, \mathcal{X}^{(m)}) d\mathcal{A} \end{aligned} \tag{4}$$

M-step: Find $\mathcal{X} = \mathcal{X}^{(m+1)}$ that maximes $Q(\mathcal{X}, \mathcal{X}^{(m)})$, i.e.

$$\mathcal{X}^{(m+1)} = M(\mathcal{X}^{(m)}) = \arg\max_{\mathcal{X} \in \Omega} Q(\mathcal{X}, \mathcal{X}^{(m)}) \tag{5}$$

2.1 Fisher information matrix and score statistic

$I(\mathcal{X} \mid \mathcal{Z}) = -\frac{\partial^2 \log L(\mathcal{X})}{\partial \mathcal{X} \partial \mathcal{X}^T}$ is the second order partial derivative of the (incomplete-data) log likelihhod function and $I_c(\mathcal{X} \mid \mathcal{Z}) = -\frac{\partial^2 \log p(\mathcal{Z}, \mathcal{A} \mid \mathcal{X})}{\partial \mathcal{X} \partial \mathcal{X}^T}$ and $I_m(\mathcal{X} \mid \mathcal{Z}) = -\frac{\partial^2 \log \frac{p(\mathcal{Z}, \mathcal{A} \mid \mathcal{X})}{p(\mathcal{Z} \mid \mathcal{X})}}{\partial \mathcal{X} \partial \mathcal{X}^T}$ the complete one. Then

$$I(\mathcal{X} \mid \mathcal{Z}) = I_c(\mathcal{X} \mid \mathcal{Z}) - I_m(\mathcal{X} \mid \mathcal{Z}) \tag{6}$$

2.2 Monotonicity

The incomplete-data likelihood function $L(\mathcal{X})$ is not decreased after an EM step, i.e.

$$L(\mathcal{X}^{(m+1)}) \geq L(\mathcal{X}^{(m)}) \tag{7}$$

2.3 Convergence

As a consequence of the monotonicity a sequence of likelihood values $\{L(\mathcal{X}^{(k)})\}$ bounded above converges monotonically to a value $L^\star$. However, it is in general not ensured that $L^\star$ is a stationary value, i.e. the existence of a stationary point $\mathcal{X}^\star$, s.t. $L^\star = L(\mathcal{X}^\star)$ holds. One has the following result [MK97]:

Suppose that $Q(\mathcal{X};\mathcal{Y})$ is continuous in both variable. Then the limit points of any instance $\{\mathcal{X}^{(k)}\}$ of the EM algorithm are stationary points of $L(\mathcal{X})$, and $L(\mathcal{X}^{(k)})$ converges monotonically to some value $L^\star = L(\mathcal{X}^\star)$

Unfortunately the convergence of the sequence of likelihood values $\{L(\mathcal{X}^{(k)})\}$ to some value $L^\star$ does not automatically imply the convergence of the corresponding sequence of iteartes $\{\mathcal{X}^{(k)}\}$ to a point $\mathcal{X}^\star$. However, one knows the following criterion [MK97]:

Assume that $L(\mathcal{X})$ is unimodal in Ω with $\mathcal{X}^\star$ being the one and only stationary point and $\frac{\partial Q(\mathcal{X};\mathcal{Y})}{\partial \mathcal{X}}$ is continuous in $\mathcal{X}$ and $\mathcal{Y}$. Then any EM sequence $\mathcal{X}^{(k)}$ converges to the unique maximizer $\mathcal{X}^\star$ of $L(\mathcal{X})$.

2.4 Expectation Maximisation for Maximum a Posteriori Estimation

The expectation maximisation is also suitable for maximum a posteriori estimation [MK97].

$$\log p(\mathcal{X} \mid \mathcal{Z}) = \log L(\mathcal{X}) + \log p(\mathcal{X}) + C \tag{8}$$

where the last term does not depend on $\mathcal{X}$. Therefore, the expectation maximisation algorithm may be applied as follows

E-step: Evaluate

$$\begin{aligned} Q_{MAP}(\mathcal{X}, \mathcal{X}^{(m)}) &= E[log(p(\mathcal{X} \mid \mathcal{W})) \mid \mathcal{Z}, \mathcal{X}^{(m)}] \\ &= \int \log(p(\mathcal{X}, \mathcal{A} \mid \mathcal{Z}))p(\mathcal{A} \mid \mathcal{X}^n, \mathcal{Z})d\mathcal{A} \end{aligned} \tag{9}$$

M-step: Find $\mathcal{X} = \mathcal{X}^{(m+1)}$ that maximes $Q_{MAP}(\mathcal{X}, \mathcal{X}^{(m)})$, i.e.

$$\mathcal{X}^{(m+1)} = \arg\max_{\mathcal{X}\in\Omega} Q_{MAP}(\mathcal{X}, \mathcal{X}^{(m)}) \tag{10}$$

3 Applications of Expectation Maximisation

3.1 PMHT

The perhaps most popular application of expectation maximisation is with the probabilistic multi hypotheses tracking [RWS96]. Therefore, assume that s enumerates the targets, $s \in \{1, ..., M\}$. The dynamic of target s is defined by the prediction:

$$x_s(t+1) = F_s(t)x_s(t) + v_s(t) \tag{11}$$

The measurement of this target related to the measurement equiation

$$y_s(t) = H_s(t)x_s(t) + w_s(t) \tag{12}$$

for $t = 1, ..., T$. $F_s(t)$ is the prediction matrix, $H_s(t)$ the projection onto the measurements space, $v_s(t)$ respectively $w_s(t)$ the Gaussian distributed process- respectively measurement noise with covariance $Q_s(t)$ respectively $R_s(t)$. Assuming that there are n_t measurements $\{z_r(t) \mid r = 1, ..., n_t\}$ at time t the data association problem is addressed to a variables $\{a_r(t) \mid r = 1, ..., n_t\}$, s.t. the rth measurement at time t comes from the $k_r(t)$th model (target), $y_{a_r(t)} = z_r(t)$. The multi target tracking problem is solved by the application of (a posteriori) expectation maximisation [RW01] on $\mathcal{X} = \{x_s(t)\}$, $\mathcal{Z} = \{z_r(t)\}$ and $\mathcal{A} = \{a_r(t)\}$:

$$\begin{aligned} p(\mathcal{Z}, \mathcal{X}, \mathcal{A}) = & \prod_{s=1}^{M} p(x_s(1)) \prod_{t=2}^{T} \prod_{s=1}^{M} p(x_s(t) \mid x_s(t-1)) \\ & \prod_{t=1}^{T} \prod_{r=1}^{n_t} \pi_{a_r(t)} N(z_r(t); \hat{y}_{a_r(t)}(t), R_{a_r(t)}(t)) \end{aligned} \tag{13}$$

$$\begin{aligned} p(\mathcal{Z}, \mathcal{X}) = & \prod_{s=1}^{M} p(x_s(1)) \prod_{t=2}^{T} \prod_{s=1}^{M} p(x_s(t) \mid x_s(t-1)) \\ & \prod_{t=1}^{T} \prod_{r=1}^{n_t} \left[\sum_{l=1}^{M} \pi_{a_r(t)} N(z_r(t); \hat{y}_l(t), R_l(t)(t)) \right] \end{aligned} \tag{14}$$

$$p(\mathcal{A} \mid \mathcal{Z}, \mathcal{X}) = \prod_{t=1}^{T} \prod_{r=1}^{n_t} \underbrace{\frac{\pi_l N(z_r(t); \hat{y}_l^n(t), R_l(t))}{\sum_{p=1}^{M} \pi_p N(z_r(t); \hat{y}_p^n(t), R_p(t))}}_{=w_{l,r}^n(t)} \tag{15}$$

Now one uses the a posteriori version of the expectation maximisation:

$$\begin{aligned} Q(\mathcal{X}^{n+1};\mathcal{X}^n) &= \sum_{\mathcal{A}} \log(p(\mathcal{X}^{n+1},\mathcal{A} \mid \mathcal{Z})) \prod_{t=1}^{T} \prod_{r=1}^{n_t} w^n_{a_r(t)}(t) \\ &= \log(\prod_{s=1}^{M} p(x_s^{n+1}(1)) \prod_{t=1}^{T} \prod_{s=1}^{M} p(x_s^{n+1}(t) \mid x_s^{n+1}(t-1))) \\ &\quad + \sum_{\mathcal{A},t,r} \log \pi_{a_r(t)} N(z_r(t); \hat{y}^{n+1}_{a_r(t)}(t), R_{a_r(t)}(t)) w^n_{a_r(t),r}(t) \end{aligned} \tag{16}$$

3.2 Other Applications

The PHD filter is also based on expectation maximisation. Here the expectation maximisation is used to extract the targets from the PHD. In [CBdSPP05] this approach is used for 3D sonar tracking whereas [TL04] apply a similar strategy for a passive radar application. In [GPRS04] the expectation maximisation is used for clustering in the multidimensional data association context.

References

[CBdSPP05] Daniel Edward Clark, Judith Bell, Yves de Saint-Pern, and Yvan Petillot. PHD Filter Multi-target Tracking in 3D Sonar. volume 1, pages 265 – 270. Oceans - Europe, June 2005.

[DLR77] A. P. Dempster, N. M. Lard, and D. B. Rubin. Maximium likelihood from incomplete data via the EM algorithm. volume 39, pages 1–38. Journal of the Royal Statistical Society B, 1977.

[GPRS04] Sabino Gadaleta, Aubrey Poore, Sean Roberts, and Benjamin J. Slocumb. Multiple Hypothesis Clustering, Multiple Frame Assignment, Tracking. Signal and Data Processing of Smal Targets, 2004.

[MK97] Geoffrey J. McLachlan and Triyambakam Krihnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics, New York, 1997.

[RW01] Yanhua Ruan and Peter Willett. An Improved PMHT Using an Idea from Coding. IEEE, 2001.

[RWS96] C. Rago, P. Willett, and R. Streit. Direct data fusion using the PMHT. Seattle, Washington, June 1996. Proc. the American Conference.

[TL04] Martin Tobias and Aaron D. Lanterman. Probability Hypothesis Density Based Multitarget Tracking with Bistatic Range and Doppler Observations. volume 5429, pages 296–305. Proceedings of SPIE, Signal Processing, Sensor Fusion, and Target Recognition XIII, August 2004.

[Web99] Andrew Webb. *Statistical Pattern Recognition*. Arnold, London, 1999.