IMPROVED VARIATIONAL INFERENCE FOR TRACKING IN CLUTTER

Jason L. Pacheco and Erik B. Sudderth

Department of Computer Science, Brown University, Providence, RI, USA

ABSTRACT

We apply the *expectation propagation* (EP) algorithm to temporally track targets using sensors that produce spurious clutter detections, and may sometimes fail to detect the true target. The variational inference framework underlying EP allows the tracker to be easily adapted to varying measurement models. We develop variants of EP based on single Gaussian and Gaussian mixture approximations of posterior target location distributions, which offer a tradeoff between accuracy and computational complexity. Experiments show improved tracking accuracy and uncertainty estimation relative to widely used baseline tracking algorithms.

Index Terms— target tracking, Bayesian inference, variational methods, expectation propagation

1. INTRODUCTION

Probabilistic target tracking in the presence of missed and false (*clutter*) detections poses a challenging problem, for which exact Bayesian inference is intractable [1]. There is thus a rich literature on approximate tracking algorithms.

We focus on deterministic approximate inference algorithms. Particle filters are also used for tracking [2], but lead to less compact state representations and can be unstable for high-dimensional problems. The *probabilistic data association filter* (PDAF) [3] incorporates observations sequentially via a single forward pass, approximating the state's marginal distribution as Gaussian with matched mean and covariance. The *probabilistic multi-hypothesis tracker* (PMHT) [4, 3] instead adapts the *expectation maximization* (EM) algorithm to iteratively estimate smoothed state estimates from a fixed batch of data. These algorithms are derived from different measurement models: the PDAF assumes the target produces at most one true detection per time step, while the PMHT assumes the number of true detections is binomially distributed.

In this paper, we propose a family of alternative tracking algorithms based on *expectation propagation* (EP) [5], a sophisticated variational approach to approximate inference [6]. This approach is similar in spirit to the PDAF, in that we incorporate local evidence and project to a family of tractable approximate marginal distributions. Unlike PDAF, however, our EP algorithms can produce accurate smoothed state estimates; be easily adapted to various measurement models; and employ marginal approximations of varying complexity.

After reviewing exponential families and the EP algorithm (Sec. 2), we propose three tracking algorithms (Sec. 3) which consider various clutter measurement models, and approximate marginal distributions by either single Gaussians or Gaussian mixtures. Sec. 4 validates these algorithms via Monte Carlo trials on synthetic data.

2. EXPECTATION PROPAGATION

Consider a joint distribution which factorizes according to

$$p(x \mid \mathcal{D}) \propto p_0(x) \prod_i \psi_i(x) \tag{1}$$

with latent variables x, prior distribution $p_0(x)$, and observed data $\mathcal{D}$ encoded via non-negative factors $\psi_i(x)$. We choose an approximating distribution $q(x)$ that is in a tractable *exponential family* [6] of distributions, with matched factorization

$$q(x) = p_0(x) \prod_i \widetilde{\psi}_i(x) \approx p(x \mid \mathcal{D}). \tag{2}$$

We refer to $\widetilde{\psi}_i(x)$ as *messages*, which can be thought of as local approximations. EP provides a means for iteratively refining each $\widetilde{\psi}_i(x)$ such that $q(x)$ approximates the true posterior $p(x \mid \mathcal{D})$. At each iteration, EP updates the posterior and factor approximations according to the following procedure:

$$\begin{aligned}
q^{\setminus i}(x) &= q(x)/\widetilde{\psi}_i(x) && \text{(Cavity Dist.)}\\
\hat{p}(x) &\propto q^{\setminus i}(x)\psi_i(x) && \text{(Augmented Dist.)}\\
q^{\text{new}}(x) &= \arg\min_q \ \mathrm{D}(\hat{p}(x) \parallel q(x)) && \text{(KL Projection)}\\
\widetilde{\psi}_i(x) &\propto q^{\text{new}}(x)/q^{\setminus i}(x) && \text{(New Message)}
\end{aligned}$$

The KL projection can be computed in closed form via moment-matching [6]. The messages $\widetilde{\psi}_i(x)$, as well as the cavity distribution $q_{\setminus i}(x)$, are members of an *unnormalized* exponential family; EP does nothing to explicitly enforce their normalizablity. For the KL projection to be well-posed, the augmented distribution $\hat{p}(x)$ must be normalizable. If it is not then we "halt" the update, leaving the message unchanged. For more details on EP in general, see [5, 6].

3. TRACKING IN CLUTTER VIA EP

In this section, we derive EP inference algorithms for the *dependent* observation assignment model underlying the PDAF,

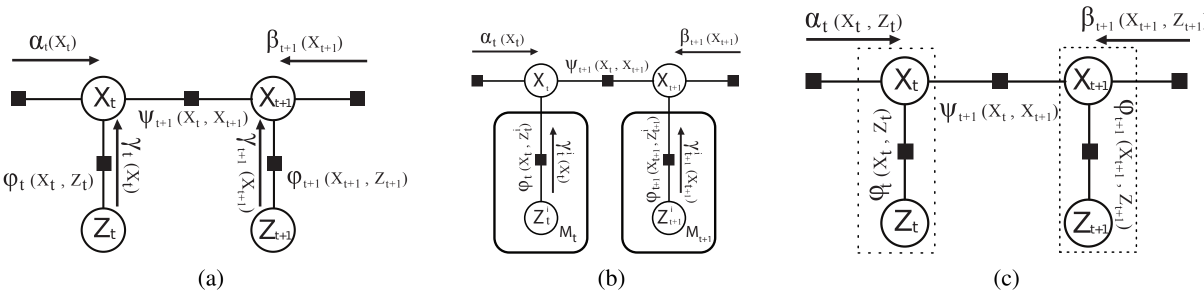


Fig. 1. Factor graphs illustrating the joint factorization and messages underlying three EP tracking algorithms. (a) EPD: Dependent observation assignments, single Gaussian state distributions. (b) EPI: Independent observation assignments, single Gaussian state distributions. (c) EPD+: Dependent observation assignments, Gaussian mixture state distributions.

and the *independent* observation assignment model underlying the PMHT. We consider approximations of state distributions by two exponential families, single Gaussian and Gaussian mixture. We do not apply the mixture approximation to the independent assignment model, where the mixture size is exponential in the number of observations per timestep.

For all of the models, the joint distribution factorizes as

$$p(X, Z) = \frac{1}{Z} p_0(x_0) \prod_{t=1}^{T} \psi_t(x_{t-1}, x_t) \varphi_t(x_t, z_t) \quad (3)$$

where the target state at scan t is $x_t \in \mathbb{R}^n$ with prior $p_0(x_0) = N(x \mid \mu_0, V_0)$ and linear Gaussian target dynamics $\psi_t(x_{t-1}, x_t) = N(x_t \mid F x_{t-1}, Q)$ where $F, Q \in \mathbb{R}^{n \times n}$. The observation likelihoods $\varphi_t(x_t, z_t)$ encode the assignment model, and depend implicitly on observed data $y_t = \{y_t^i\}_{i=1}^{M_t}$.

3.1. Observation Assignment Models

Under the dependent assignment model, at most one detection per timestep is related to the target state. Assignments are encoded as $z_t \in \{0, \ldots, M_t\}$, where $z_t = 0$ indicates that all observations are clutter. Otherwise, $y_t^{z_t}$ is target-generated:

$$\varphi_t^D(x_t, z_t) = \delta_{z_t,0} \lambda_0 + \sum_{i=1}^{M_t} \delta_{z_t,i} \lambda_i N(y_t^i \mid H x_t, R) \quad (4)$$

Here, $\delta_{\cdot,\cdot}$ is the Kronecker delta, $H \in \mathbb{R}^{m \times n}$ and $R \in \mathbb{R}^{m \times m}$. The overall potential is a mixture of M_t Gaussians and a constant. Typically, $\lambda_0 = (1 - P_d) N(y_t^i \mid 0, \Sigma_0)$ and $\lambda_i = \frac{P_d}{M_t} p(z_t = i)$ for some probability of detection P_d.

The independent assignment model assumes the M_t detections are marginally independent, where $z_t^i = 0$ if detection i is clutter, and $z_t^i = 1$ if it is related to the target:

$$\varphi_t^I(x_t, z_t^i) = \delta_{z_t^i,0} \lambda_0 + \delta_{z_t^i,1} \lambda_1 N(y_t^i \mid H x_t, R) \quad (5)$$

The overall observation likelihood at time t is then the product $\prod_{i=1}^{M_t} \varphi_t^I(x_t, z_t^i)$, a mixture of $\mathcal{O}(2^{M_t})$ Gaussians plus a constant term.

3.2. EPD: Dependent Assignment, Single Gaussian

We begin with a Gaussian marginal posterior approximation $q_t(x_t) = N(x_t \mid m_t, V_t)$ defined as the product of a *forward* message $\alpha_t(x_t)$, a *backward* message $\beta_t(x_t)$, and a *measurement* message $\gamma_t(x_t)$:

$$q_t(x_t) \propto \alpha_t(x_t) \gamma_t(x_t) \beta_t(x_t) \approx p(x_t \mid Y_1^T) \quad (6)$$

The messages are parameterized as unnormalized Gaussians in information form,

$$\alpha_t(x_t) = s_t^\alpha \exp(-\frac{1}{2} x_t^T \Lambda_t^\alpha x_t + x_t^T \eta_t^\alpha), \quad (7)$$

with similar definitions for $\beta_t(x_t)$ and $\gamma_t(x_t)$. Figure 1(a) shows a factor graph [6] for this model along with overlays denoting the direction and type of messages.

The forward pass augmented distribution at scan t yields a Gaussian density. Dropping explicit dependence on x_t,

$$\hat{p}_t(\cdot) \propto \gamma_t(\cdot) \beta_t(\cdot) \int_{\mathcal{X}} \alpha_{t-1}(x) \, \gamma_{t-1}(x) \, \psi_t(x, \cdot) \, dx \quad (8)$$

The EP projection step introduces no approximation, so $q_t^{\text{new}} = \hat{p}_t$. The forward messages $\alpha_t(x_t)$ are as in a conventional Kalman filter, and the reverse messages $\beta_t(x_t)$ as in a two-pass Kalman smoother. In contrast, the measurement messages $\gamma_t(x_t)$ involve a projection step since the augmented distribution is non-Gaussian:

$$\hat{p}_t(x_t) \propto \sum_{z_t=0}^{M_t} \alpha_t(x_t) \varphi_t^D(x_t, z_t) \beta_t(x_t) \quad (9)$$

The projection $q_t^{\text{new}} \propto \arg\min_q \mathrm{D}(\hat{p}_t \| q)$ matches the mean and variance of the Gaussian mixture $\hat{p}_t(x_t)$. The measurement message update is $\gamma_t^{\text{new}}(x_t) = \frac{q_t^{\text{new}}(x_t)}{\alpha_t(x_t)\beta_t(x_t)}$.

A single forward pass of this algorithm, iteratively updating α_t and γ_t, is equivalent to the PDAF [3]. To see this, note the correspondence between the PDAF prediction step and the calculation of the forward messages α_t. Similarly, the PDAF association probabilities correspond to the mixture weights of the augmented distribution of Eq. (9). The projection step yields a Gaussian posterior $q_t(x_t)$, the mean of which corresponds to the minimum mean square error (MMSE) state prediction of PDAF.

Further iterations of EPD provide a novel way of generalizing PDAF to produce smoothed state estiamtes. Each iteration has linear complexity $\mathcal{O}(N)$, where $N = \sum_{t=1}^{T} M_t$ is the total number of detections.

3.3. EPI: Independent Assignment, Single Gaussian

As in the EPD algorithm, we approximate the state posterior via a single Gaussian distribution:

$$q_t(x_t) \propto \alpha_t(x_t) \left(\prod_{i=1}^{M_t} \gamma_t^i(x_t) \right) \beta_t(x_t) \approx p(x_t \mid Y_1^T) \quad (10)$$

Note that we have a separate measurement message $\gamma_t^i(x_t)$ for each observation, and the posterior depends on the product of all of these messages. Figure 1(b) shows a factor graph for this model with overlays indicating the forward, backward, and measurement messages.

The forward pass augmented distribution at scan t yields a Gaussian density. Dropping explicit dependence on x_t,

$$\hat{p}_t(\cdot) \propto \beta_t(\cdot) \prod_{i=1}^{M_t} \gamma_t^i(\cdot) \int_{\mathcal{X}} \alpha_{t-1}(x) \prod_{i=1}^{M_{t-1}} \gamma_{t-1}^i(x) \psi_t(x, \cdot)\, dx$$

This is Gaussian, so as in EPD the forward and backward messages correspond to conventional Kalman filters and smoothers. The measurement message update at each scan is equivalent to an instance of EP for the *clutter problem* [5].

One full iteration of EPI has linear complexity $\mathcal{O}(N)$, where N is again the total number of detections. This algorithm does not appear to be equivalent to classical tracking algorithms, for any message schedule. EPI assumes the same assignment model as the PMHT, but the algorithm is distinct.

3.4. EPD+: Dependent Assignment, Gaussian Mixture

Returning to the dependent assignment model of EPD, we extend EP to employ a more flexible, Gaussian mixture marginal approximation. A closely related algorithm has been used for inference in switching state-space models [7]. Tractability of the posterior is maintained by limiting the marginal at scan t to a mixture approximation with M_t modes, one for each possible hypothesis z_t as to which detection is of the target:

$$q_t(x_t, z_t) = \sum_{j=0}^{M_t} \delta_{z_t,j} p_{t,j} N(x_t \mid m_{t,j}, V_{t,j}) \approx p(x_t, z_t \mid Y_1^T)$$

Note that unlike the simpler EPD approximation, $q_t(x_t, z_t)$ is defined over the target state x_t and assignments z_t jointly. Measurement messages are not necessary, because the measurement potential lies in the approximating family. We define the marginal as the product of forward and backward messages $q_t(x_t, z_t) \propto \alpha_t(x_t, z_t)\beta_t(x_t, z_t)$. The messages are parameterized as unnormalized Gaussian mixtures,

$$\alpha_t(x_t, z_t) = \sum_{j=0}^{M_t} \delta_{z_t,j} p_{t,j}^{\alpha} \exp(-\frac{1}{2} x_t^T \Lambda_{t,j}^{\alpha} x_t + x_t^T \eta_{t,j}^{\alpha}) \quad (11)$$

with a similar definition for $\beta_t(x_t, z_t)$. Figure 1(c) depicts a factor graph representation of this model with overlays for the forward and backward messages.

The augmented distribution in the forward pass at scan t, again dropping explicit dependence on x_t, is now

$$\hat{p}_t(\cdot, k) \propto \beta_t(\cdot, k) \varphi_t^D(\cdot, k) \sum_{j=0}^{M_{t-1}} \int_{\mathcal{X}} \alpha_{t-1}(x, j) \psi(x, \cdot)\, dx$$

For each candidate $z_t = k$, the augmented distribution $\hat{p}_t(x_t, k)$ is a Gaussian mixture with M_{t-1} components. We project each of these mixtures to a single Gaussian $q_t^{\text{new}}(x_t, z_t = k)$ with matched mean and covariance. The posterior approximation q_t^{new} is then a mixture of $M_t + 1$ Gaussians, indexed by z_t.

The updates of backward messages $\beta_t(x_t, z_t)$ proceed similarly to the forward pass. A single forward pass of EPD+ is similar to the *Gaussian Pseudo-Bayesian estimator of second order* (GPB2) [3], which is a forward filter for estimation in a *switching linear dynamical system* (SLDS). One or more forward and backward passes of EPD+ correspond to smoothed generalizations of GPB2, and thus novel algorithms for tracking in clutter. If there are M detections per time step, one iteration of EPD+ has computational complexity $\mathcal{O}(TM^2) = \mathcal{O}(NM)$. This is greater than EPD but still linear in T, and often tractable.

3.5. KNN: Nearest Neighbor Association Baseline

The Kalman filter with nearest neighbor association (KNN) provides a baseline comparison [1]. Given approximate filtered marginals $\hat{p}_t(x_t) = N(x_t \mid m_t, P_t)$, we predict the state evolution as follows:

$$\hat{p}(x_{t+1} \mid Y_1^t) = N(x_{t+1} \mid Fm_t, Q + FP_tF^T)$$

We refer to $\hat{x}_{t+1|t} = Fm_t$ as the predicted target state and $\hat{P}_{t+1|t} = Q + FP_{t-1}F^T$ as the predicted covariance. The predicted measurement is given by $\hat{y}_{t+1} = H\hat{x}_{t+1|t}$. Assuming Gaussian noise, the most likely associated measurement can be selected based on the detection nearest to $\hat{y}_{t+1}$:

$$\hat{z}_{t+1} = \underset{z \in \{1,\ldots,M_t\}}{\arg\min} \; \|\hat{y}_{t+1} - y_{t+1}^z\|_2^2$$

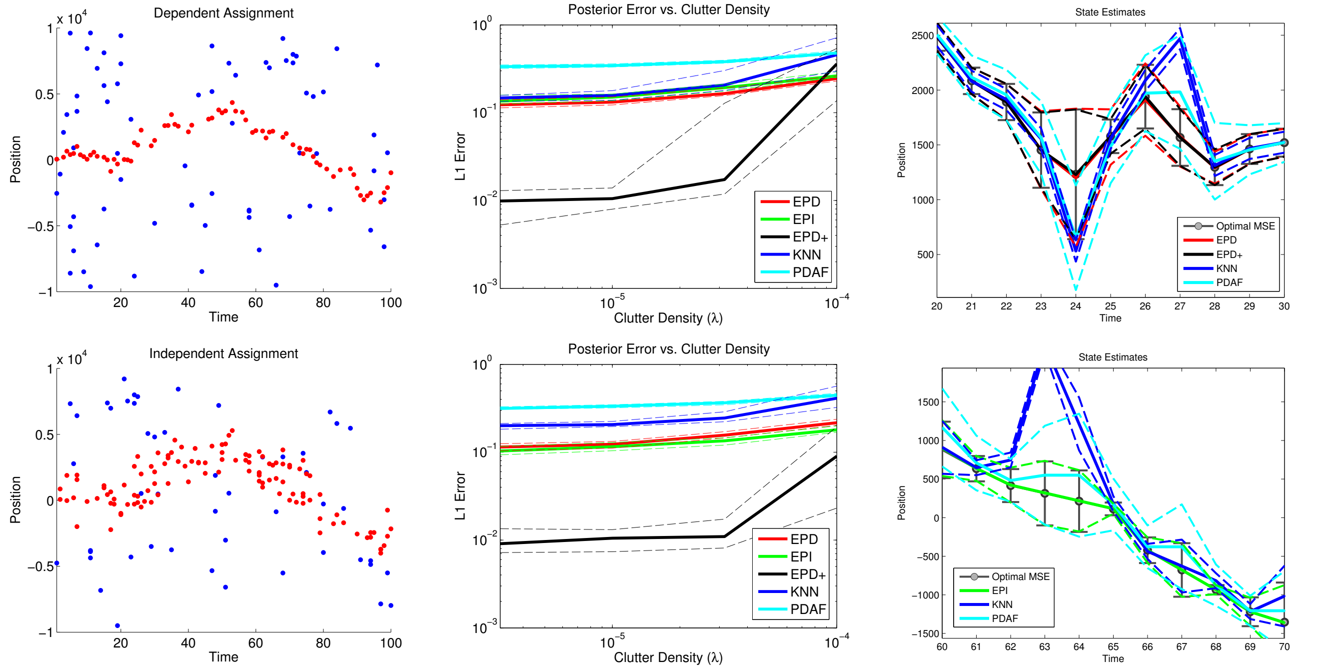
The measurement residual is calculated based on the nearest-neighbor association as $\nu_{t+1} = y_{t+1}^{\hat{z}_{t+1}} - \hat{y}_{t+1}$. Incorporating the measurement we update the marginal as,

$$\hat{p}_{t+1}(\cdot) = N(\cdot \mid \hat{x}_{t+1|t} + W\nu_{t+1}, \hat{P}_{t+1|t} - W_{t+1}S_{t+1}W_{t+1}^T)$$

where W_{t+1} and S_{t+1} are the typical Kalman gain and innovation covariance, respectively. The smoothed posterior marginal $\hat{p}(x_t \mid Y_1^T)$ is computed as the product of forward and reverse-time filters, using associations as above.

4. EXPERIMENTAL RESULTS

We conduct a Monte Carlo simulation for a one-dimensional latent state x_t with random walk dynamics $x_t \sim N(x_{t-1}, \sigma_p^2)$, initialized uniformly in the observation region. Under either assignment model, target detections $y_t^i \sim N(x_t, \sigma_m^2)$ and clutter detections $y_t^i \sim N(0, \sigma_0^2)$. The *clutter density* λ is proportional to the number of false detections.



(a) (b) (c)

Fig. 2. Data sampled from the dependent assignment model (top) and independent assignment model (bottom). (a) Example scenario with $P_d = 0.7$ and $\lambda = 10^{-4.5}$. True target detections are red, clutter detections blue. (b) Across 100 instances, we plot the median (solid) and $(0.25, 0.75)$ quantiles (dashed) of L_1 error versus clutter density λ. (c) Close-up track estimates (solid), and one standard deviation error estimates (dashed), for multiple methods applied to a single instance of each dataset.

We evaluate algorithm performance by the L_1 distance from the true posterior marginals, accurately approximated by finely discretizing the state space and running the forward-backward algorithm for hidden Markov models (HMMs) [6]. This numerical baseline is possible with one-dimensional states, but intractable for higher-dimensional problems where our EP algorithms remain feasible.

We vary $\lambda \in \{10^{-5.5}, 10^{-5.0}, 10^{-4.5}, 10^{-4.0}\}$, fixing the probability of detection as $P_d = 0.7$. For every setting of parameters we sample 100 random instances, each with $T = 100$ time points. While convergence is not guaranteed in these models, we achieve adequate convergence for our results by damping the conventional EP message updates [5, 7], with damping parameter $\alpha = 0.5$.

Figure 2 shows results for data sampled from the both the dependent and independent assignment models. As measured by L_1 error, EP consistently outperforms baseline methods, and the Gaussian mixture approximation of EPD+ is superior in almost all cases. In general EP seems robust to model mismatch, as EPD+ is effective even for data from the independent assignment model. EPD clearly improves over PDAF.

Figure 2(c) shows close-up track estimates for particular instances sampled from each assignment model. KNN consistently underestimates posterior variance, while PDAF overestimates it. State estimates among the EP algorithms are generally comparable or superior to baselines. EPD+ more accurately estimates the posterior variance than other methods.

5. CONCLUSION

We have used EP to develop a family of target tracking algorithms which are significantly more accurate than baseline methods. Our approach allows significant flexibility in the choice of observation model, and in choosing the marginal approximation to tradeoff accuracy and computation.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, New Jersey, 1979.

[2] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*, Springer Verlag, 2001.

[3] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, John Wiley and Sons, Inc., 2002.

[4] R. L. Streit and T. E. Luginbuhl, "Probabilistic multi-hypothesis tracking," Tech. Rep. 10,428, NUWC-NPT, Feb. 1995.

[5] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *UAI*, vol. 17, pp. 362–369, 2001.

[6] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, pp. 1–305, 2008.

[7] T. Heskes and O. Zoeter., "Expectation propagation for approximate inference in dynamic bayesian networks," *Uncertainty in Artificial Intelligence*, vol. 18, pp. 216–223, 2002.