

Ensemble Learning Online Filtering in Wireless Sensor Networks

Hichem Snoussi

ISTIT/M2S, University of Technology of Troyes,
12, rue Marie Curie, 10000, France
Email: snoussi@utt.fr

Cédric Richard

ISTIT/M2S, University of Technology of Troyes,
12, rue Marie Curie, 10000, France
Email: richard@utt.fr

Abstract—In many applications, the observed system is assumed to evolve according to a probabilistic state space model. The data likelihood function is, in general, non linear or/and non Gaussian leading to analytically intractable inference. Particle filter is a popular approximate Monte Carlo solution based on a particle representation of the filtering distribution. However, power constraints in sensor networks requires an additional approximation (compression) when communicating the particle based representation. In this contribution, we propose an alternative ensemble learning (variational) approximation suitable to the communication constraints of sensor networks. The efficiency of the variational approximation relies on the fact that the online update of the filtering distribution and its compression are simultaneously performed. In addition, the variational approach has the nice property to be parameterization-independent ensuring the robustness of the data processing. The selection of the leader node is based on a trade-off between communication constraints and information content relevance of measured data.

I. INTRODUCTION

Collaborative information processing in sensor networks is becoming a very attractive field of research [1]. In such a sensor network, the sensors role is not limited to detect and transmit the data to a central unit where they are processed. The sensors have the ability to collaborate, exchange information to ensure optimal decision. Such sensors are called “smart” sensors. Contrary to the centralized approach, the system does not depend on a unique processing unit whose damaging leads to the entire system failure. Every smart sensor is able to play a central role and provide a suboptimal decision. The system is thus very robust against a probable foreign attack or a technical failure of the central unit. In addition, as collected data are locally processed, only pertinent information is exchanged between smart nodes, limiting hence the required channel communication bandwidth. In fact, the nodes batteries have a limited lifetime and a wise management of the energy consumption has to be considered in such a network.

In this paper, we consider target tracking application where the observed system evolution is described by a non linear state space model, the non linear function depending on the sensor. The proposed approach could be however extended to deal with other applications as region monitoring, multi-cameras visual tracking,...

The filtering consists of estimating the posterior marginal probability of the continuous hidden state. To solve the inference problem, the particle filter is frequently used as an

approximate Monte Carlo inference method able to deal with the non linear aspect of the system dynamics. Recently, distributed particle filters were proposed in literature [2], [3]. To meet the energy constraints, only one leader node is selected at each time step to implement the particle filter. However, distributed particle filters need the exchange of large sample based representations of the filtering distribution between selected leaders. Message approximations should therefore be considered in order to reduce the communication energy consumption. In [3], a message approximating scheme based on a greedy KD-tree approximation is proposed. In [2], a full collaborative strategy is proposed with a GMM (Gaussian Mixture Model) message approximation implemented by an EM (Expectation-Maximization) algorithm. However, with the GMM approximation, the error propagation is not controlled and the model refinement is not straightforward as with the KD-tree approximation. Unlike the GMM modeling, the KD-tree Gaussian mixture modeling allows a more flexible control of the trade-off between the approximation precision and the communication constraints. However, in both cases, an inference error is propagated by the successive message approximations.

In this contribution, we propose a collaborative distributed ensemble learning (variational) filter, based on online updating a free form approximation of the filtering distribution. The ensemble learning approach allows a natural adaptive compression of the non Gaussian filtering distribution. The approximating message can then be communicated between leader nodes without loss. Furthermore, the ensemble learning filtering outperforms the particle filter even if the sample based representations are entirely communicated [4]. In fact, when the hidden state dimension is high or the likelihood is sharply peaked, the particle filter is likely to degenerate keeping only a few particles in regions with high likelihood. The ensemble learning filter circumvents this problem allowing a robust target tracking in a distributed environment. A node selection protocol is proposed as well. It is mainly based on updating the predictive target density within an ensemble learning procedure.

The paper is organized as follows: in Section II, the probabilistic sequential Bayesian filtering is introduced. The distributed particle filter method is briefly described. The section III is the main contribution of this paper. We propose

a collaborative ensemble learning filter in a distributed sensor network where the smart sensors exchange only Gaussian statistics (sufficient statistics) and select the leader nodes in an autonomous way. Finally, some simulation results corroborating the efficiency of the proposed algorithms are presented.

II. CENTRALIZED AND DISTRIBUTED PARTICLE FILTERING

In this section, we briefly recall the particle filter method for filtering in nonlinear dynamical systems. It is an approximate Monte Carlo method estimating, recursively in time, the marginal posterior distribution of the continuous hidden state of the system, given the observations. The particle filter provides a point mass approximation of these distributions. For more details and a comprehensive review of the particle filter see [5].

The observed system evolves in time according to the following nonlinear dynamics:

$$\begin{cases} \mathbf{x}_t & \sim p_x(\mathbf{x}_t | \mathbf{x}_{t-1}, u_t) \\ \mathbf{y}_t^{(m)} & \sim p_m(\mathbf{y}_t^{(m)} | \mathbf{x}_t, u_t), \quad m = 1..M, \end{cases} \quad (1)$$

where $\mathbf{y}_t^{(m)} \in \mathbb{R}^{n_y}$ denotes the observations transmitted from the sensor m at time t , $\mathbf{x}_t \in \mathbb{R}^{n_x}$ denotes the unknown continuous state, and $u_t \in \mathcal{U}$ denotes a known control signal. The probability distribution $p_x(\mathbf{x}_t | \mathbf{x}_{t-1}, u_t)$ models the stochastic transition dynamics of the hidden state. Given the continuous state, the observations $\mathbf{y}_t^{(m)}$ from each sensor m follow a stochastic model $p_m(\mathbf{y}_t | \mathbf{x}_t, u_t)$, where the stochastic aspect reflects the observation noise.

A. Centralized particle filtering

In this paper, we assume that, given the states \mathbf{x}_t , the sensor noises are stochastically independent:

$$p(\mathbf{y}_t^{(1)}, \dots, \mathbf{y}_t^{(M)} | \mathbf{x}_t, u_t) = \prod_{m=1}^M p_m(\mathbf{y}_t^{(m)} | \mathbf{x}_t, u_t).$$

Consequently, concatenating the observations gathered in the central unit, $\mathbf{y}_t = [\mathbf{y}_t^{(1)}, \dots, \mathbf{y}_t^{(M)}]$, and replacing the distribution product $\prod p_m$ by an observation distribution p_y , the stochastic model (1) is rewritten as:

$$\begin{cases} \mathbf{x}_t & \sim p_x(\mathbf{x}_t | \mathbf{x}_{t-1}, u_t) \\ \mathbf{y}_t & \sim p_y(\mathbf{y}_t | \mathbf{x}_t, u_t). \end{cases} \quad (2)$$

1) Bayesian Importance Sampling: The Bayesian filtering is based on the estimation of the posterior marginal probability $p(\mathbf{x}_t | \mathbf{y}_{1:t})$. The nonlinear and the non Gaussian aspect of the transition distributions leads to intractable integrals when evaluating the marginals. Therefore, one has to resort to Monte Carlo approximation where the joint posterior distribution $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$ is approximated by the point-mass distribution of a set of weighted samples (called particles) $\{\mathbf{x}_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$:

$$\hat{p}_N(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\mathbf{x}_{0:t}^{(i)}}(d\mathbf{x}_{0:t}),$$

where $\delta_{\mathbf{x}_{0:t}^{(i)}}(d\mathbf{x}_{0:t})$ denotes the Dirac function.

Based on the same set of particles, the marginal posterior probability (of interest) $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ can also be approximated as follows:

$$\hat{p}_N(\mathbf{x}_t | \mathbf{y}_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(d\mathbf{x}_t),$$

In the Bayesian importance sampling (IS) method, the particles $\{\mathbf{x}_{0:t}^{(i)}\}_{i=1}^N$ are sampled according to a proposal distribution $\pi(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$ and $\{w_t^{(i)}\}$ are the corresponding normalized importance weights:

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}^{(i)}) p(\mathbf{x}_{0:t}^{(i)})}{\pi(\mathbf{x}_{0:t}^{(i)} | \mathbf{y}_{1:t})}.$$

2) Sequential Monte Carlo: Sequential Monte Carlo (SMC) consists of propagating the trajectories $\{\mathbf{x}_{0:t}^{(i)}\}_{i=1}^N$ in time without modifying the past simulated particles. This is possible for the class of proposal distributions having the following form:

$$\pi(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = \pi(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1}) \pi(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}).$$

The normalized importance weights are then recursively computed in time as:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}. \quad (3)$$

B. Distributed particle filtering

In an energy constrained sensor network, the observed data $\{\mathbf{y}_t^{(m)}\}_{m=1}^M$ are not propagated from the nodes to a central unit. Instead, at each time t , a leader node m^* is selected according to an appropriate protocol (see section III). The Bayesian sequential Monte Carlo is then based on the local data $\mathbf{y}_t^{(m^*)}$ and the past filtering distribution $\hat{p}_N(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})$ communicated by the leader node at time $t-1$. Figure 1 illustrates the distributed particle filter.

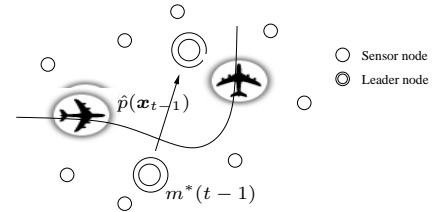


Fig. 1. The target tracking is based on the local data collected at the selected leader node and the past filtering distribution $\hat{p}(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})$ compressed and sent by the previous leader node.

Propagating all the particles $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}$ is not allowed in a wireless sensor network because of the communications constraints. Approximating the sample representation by a mixture is thus considered in [3] and [2]:

$$\hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \approx \sum_{s \in \mathcal{S}} \alpha_s \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \quad (4)$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian density.

Successive approximations by mixture of Gaussian densities necessarily lead to a propagation of errors through time. In [3], the consequences of such error propagation are explored with respect to the errors occurring in future inference. In the following section, we propose an ensemble learning approach avoiding a lossy message compression and even outperforming the particle filter based on a complete sample based representation.

III. ENSEMBLE LEARNING SEQUENTIAL FILTERING

A. State-space model

In this paper, the likelihood function in (1) is assumed to have a general form depending on the sensing model of the nodes. Concerning the transition dynamics $p_x(\mathbf{x}_t | \mathbf{x}_{t-1}, u_t)$, we adopt a mean-scale mixture model. According to this model, introduced in [4] for visual tracking, the hidden state $\mathbf{x}_t \in \mathbb{R}^{n_x}$ has a Gaussian distribution with a random mean $\boldsymbol{\mu}_t$ and a random precision matrix $\boldsymbol{\lambda}_t$. The mean follows a Gaussian random walk reflecting the time correlation of the system trajectory and the precision matrix follows a Wishart distribution:

$$\begin{cases} \boldsymbol{\mu}_t \sim \mathcal{N}(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \bar{\boldsymbol{\lambda}}) \\ \boldsymbol{\lambda}_t \sim \mathcal{W}_{\bar{n}}(\boldsymbol{\lambda}_t | \bar{\mathbf{S}}) \\ \mathbf{x}_t \sim \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t) \end{cases} \quad (5)$$

where the fixed hyperparameters $\bar{\boldsymbol{\lambda}}$, \bar{n} and $\bar{\mathbf{S}}$ are respectively the random walk precision matrix, the degrees of freedom and the precision of the Wishart distribution. Note that assuming random mean and covariance for the state \mathbf{x}_t leads to a prior probability distribution covering a wide range of tail behaviors allowing discrete jumps in the target trajectory. In fact, the marginal state distribution is obtained by integrating over the mean and precision matrix:

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \int \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t) p(\boldsymbol{\mu}_t, \boldsymbol{\lambda}_t | \mathbf{x}_{t-1}) d\boldsymbol{\mu}_t d\boldsymbol{\lambda}_t \quad (6)$$

where the integration with respect to the precision matrix leads to the known class of scale mixture distributions introduced by Barndorff-Nielsen [6]. Low values of the degrees of freedom \bar{n} reflects the heavy tails of the marginal distribution (6).

B. Updating Free form approximate distributions

According to the model (5), the augmented hidden state is now $\boldsymbol{\alpha}_t = (\mathbf{x}_t, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t)$. Instead of approximating the filtering distribution $p(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t})$ by a point-mass distribution, the ensemble learning approach consists in approximating the filtering distribution by a more tractable posterior distribution $q(\boldsymbol{\alpha}_t)$. We can minimize the Kullback-Leibler divergence between the true filtering distribution and the approximate distribution,

$$D_{\text{KL}}(q||p) = \int q(\boldsymbol{\alpha}_t) \log \frac{q(\boldsymbol{\alpha}_t)}{p(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t})} d\boldsymbol{\alpha}_t, \quad (7)$$

to obtain the optimal approximate distribution. In order to ensure that the best model is automatically chosen, we assume a free form (non parametric) approximate distribution. Choosing a separable distribution $q(\boldsymbol{\alpha}_t) = q(\mathbf{x}_t)q(\boldsymbol{\mu}_t)q(\boldsymbol{\lambda}_t)$ and minimizing the kullback-Leibler divergence (7) with variational calculus yield the following approximate distribution:

$$\begin{cases} q(\mathbf{x}_t) \propto \exp \langle \log p(\mathbf{y}_{1:t}, \boldsymbol{\alpha}_t) \rangle_{q(\boldsymbol{\mu}_t)q(\boldsymbol{\lambda}_t)} \\ q(\boldsymbol{\mu}_t) \propto \exp \langle \log p(\mathbf{y}_{1:t}, \boldsymbol{\alpha}_t) \rangle_{q(\mathbf{x}_t)q(\boldsymbol{\lambda}_t)} \\ q(\boldsymbol{\lambda}_t) \propto \exp \langle \log p(\mathbf{y}_{1:t}, \boldsymbol{\alpha}_t) \rangle_{q(\mathbf{x}_t)q(\boldsymbol{\mu}_t)} \end{cases} \quad (8)$$

The above procedure leads to an iterative algorithm to update each approximate $q(\boldsymbol{\alpha}_i)$. It is worth noting that while the distributions in the separable approximation are independent, the parameters of the distributions are in fact dependent as they are jointly updated. Furthermore, the update of the approximate distribution $q(\boldsymbol{\alpha}_t)$ can be sequentially implemented given only the approximate distribution $q(\boldsymbol{\mu}_{t-1})$. In fact, taking into account the separable approximate distribution at time $t-1$, the filtering distribution is written,

$$\begin{aligned} p(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t}) &\propto p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t, \boldsymbol{\lambda}_t | \boldsymbol{\mu}_t) \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}) q(\boldsymbol{\alpha}_{t-1}) d\boldsymbol{\alpha}_{t-1} \\ &\propto p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t, \boldsymbol{\lambda}_t | \boldsymbol{\mu}_t) \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}) q(\boldsymbol{\mu}_{t-1}) d\boldsymbol{\mu}_{t-1} \end{aligned} \quad (9)$$

where only integration with respect to $\boldsymbol{\mu}_{t-1}$ remains due to the separable form of the approximate distribution $q(\boldsymbol{\alpha}_{t-1})$. The temporal dependence on the past is hence limited to only one component approximate distribution. Communication between two successive leader nodes m_{t-1}^* and m_t^* is then based on sending $q(\boldsymbol{\mu}_{t-1})$ which is the sufficient statistic for the leader node to update the filtering distribution. As it will be shown in the following, it turns out that $q(\boldsymbol{\mu}_{t-1})$ is a Gaussian distribution and thus it can be communicated by sending only the mean and the covariance. The ensemble learning is then implemented in a collaborative sensor network without lossy compression.

Substituting the filtering distribution (9) in (8) and taking into account the prior mean-scale mixture transition model (5), the updated separable distribution $q(\boldsymbol{\alpha}_t)$ has the following form:

$$\begin{aligned} q(\mathbf{x}_t) &\propto p(\mathbf{y}_t | \mathbf{x}_t) \mathcal{N}(\mathbf{x}_t | \langle \boldsymbol{\mu}_t \rangle, \langle \boldsymbol{\lambda}_t \rangle) \\ q(\boldsymbol{\mu}_t) &\propto \mathcal{N}(\boldsymbol{\mu}_t | \boldsymbol{\mu}_t^*, \boldsymbol{\lambda}_t^*) \\ q(\boldsymbol{\lambda}_t) &\propto \mathcal{W}_{n^*}(\boldsymbol{\lambda}_t | \mathbf{S}_t^*) \end{aligned}$$

where the parameters are iteratively updated according to the following scheme:

$$\begin{aligned} \boldsymbol{\mu}_t^* &= \boldsymbol{\lambda}_t^{*-1} (\langle \boldsymbol{\lambda}_t \rangle \langle \mathbf{x}_t \rangle + \boldsymbol{\lambda}_t^p \boldsymbol{\mu}_t^p) \\ \boldsymbol{\lambda}_t^* &= \langle \boldsymbol{\lambda}_t \rangle + \boldsymbol{\lambda}_t^p \\ n^* &= \bar{n} + 1 \\ \mathbf{S}_t^* &= (\langle \mathbf{x}_t \mathbf{x}_t^T \rangle - \langle \mathbf{x}_t \rangle \langle \boldsymbol{\mu}_t \rangle^T - \langle \boldsymbol{\mu}_t \rangle \langle \mathbf{x}_t \rangle^T + \langle \boldsymbol{\mu}_t \boldsymbol{\mu}_t^T \rangle + \bar{\mathbf{S}}^{-1})^{-1} \\ \boldsymbol{\mu}_t^p &= \boldsymbol{\mu}_{t-1}^* \\ \boldsymbol{\lambda}_t^p &= (\boldsymbol{\lambda}_{t-1}^{*-1} + \bar{\boldsymbol{\lambda}}^{-1})^{-1} \end{aligned} \quad (10)$$

Note that the mean state and the precision matrix have closed forms such that their means are easily derived:

$$\langle \boldsymbol{\mu}_t \rangle = \boldsymbol{\mu}_t^*, \quad \langle \boldsymbol{\lambda}_t \rangle = n^* \mathbf{S}_t^*, \quad \langle \boldsymbol{\mu}_t \boldsymbol{\mu}_t^T \rangle = \boldsymbol{\lambda}_t^{*-1} + \boldsymbol{\mu}_t^* \boldsymbol{\mu}_t^{*T}.$$

However, the state \mathbf{x}_t does not have a closed form approximate distribution. In order to compute its mean and covariance, one can resort to an Importance sampling scheme where samples are drawn from the Gaussian $\mathcal{N}(\mathbf{x}_t | \langle \boldsymbol{\mu}_t \rangle, \langle \boldsymbol{\lambda}_t \rangle)$ and weighted according to their likelihoods:

$$\mathbf{x}_t^{(i)} \sim \mathcal{N}(\mathbf{x}_t | \langle \boldsymbol{\mu}_t \rangle, \langle \boldsymbol{\lambda}_t \rangle), w_t^{(i)} \propto p(\mathbf{y}_t | \mathbf{x}_t^{(i)}). \quad (11)$$

The mean and covariance are then obtained by their empirical approximations:

$$\langle \mathbf{x}_t \rangle = \sum_{i=1}^N w_t^{(i)} \mathbf{x}_t^{(i)}, \quad \langle \mathbf{x}_t \mathbf{x}_t^T \rangle = \sum_{i=1}^N w_t^{(i)} \mathbf{x}_t^{(i)} \mathbf{x}_t^{(i)T}$$

Note that, contrary to the distributed particle filter, the above Monte Carlo approximation remains local at the leader node and it is not communicated through the sensor network.

C. Node scheduling protocol

In order to prolong the sensor network lifetime, only one active node (leader) is processing data. The remaining nodes are wether turned off if a minimum region coverage is guaranteed or just sensing and locally broadcasting measurements (see the illustrative figure 2). The selection of the leader node should be based on the relevance of its measured data. Information theoretic criteria such that proposed in [7] could be considered for this purpose. In this paper, however, we propose a simpler scheme adapted to the target tracking application and efficiently implemented by the ensemble learning approach.

In some sensor network target tracking applications, the node measurement is inversely proportional to the distance between the node position and the target position, in a limited sensing range. Before collecting data \mathbf{y}_t , the best node at time t is thus the closest node to the predicted position of the target. In order to take into account all the statistical behavior of the predicted target position, the relevance of the node's data $\mathcal{I}(m)$ can be measured by the predictive distribution $p(\mathbf{x}_t | \mathbf{y}_{1:t-1})$ applied to the sensor position:

$$\mathcal{I}(m) = p(\mathbf{x}_t | \mathbf{y}_{1:t-1})|_{\mathbf{x}_t = \mathbf{s}_m}$$

where \mathbf{s}_m is the position of the sensor m .

The predictive distribution $p(\mathbf{x}_t | \mathbf{y}_{1:t-1})$ can also be updated by the ensemble learning approach. In fact, taking into account the separable approximate distribution at time $t-1$, the predictive distribution is written,

$$p(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t-1}) \propto p(\mathbf{x}_t, \boldsymbol{\lambda}_t | \boldsymbol{\mu}_t) \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}) q(\boldsymbol{\mu}_{t-1}) d\boldsymbol{\mu}_{t-1} \quad (12)$$

The exponential form solution, which minimizes the Kullback-Leibler divergence between the predictive distribution $p(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t-1})$ and the separable approximate distribution $q_{t|t-1}(\boldsymbol{\alpha}_t)$, yields Gaussian distributions for the state and its mean and Wishart distribution for the precision matrix:

$$\begin{aligned} q_{t|t-1}(\mathbf{x}_t) &\propto \mathcal{N}(\mathbf{x}_t | \langle \boldsymbol{\mu}_t \rangle, \langle \boldsymbol{\lambda}_t \rangle) \\ q_{t|t-1}(\boldsymbol{\mu}_t) &\propto \mathcal{N}(\boldsymbol{\mu}_t | \boldsymbol{\mu}_t^*, \boldsymbol{\lambda}_t^*) \\ q_{t|t-1}(\boldsymbol{\lambda}_t) &\propto \mathcal{W}_{n^*}(\boldsymbol{\lambda}_t | \mathbf{S}_t^*) \end{aligned}$$

where the parameters are updated according to the same iterative scheme (10), except that the mean and covariance states are now exactly evaluated without resorting to the importance sampling method:

$$\langle \mathbf{x}_t \rangle = \langle \boldsymbol{\mu}_t \rangle, \quad \langle \mathbf{x}_t \mathbf{x}_t^T \rangle = \langle \boldsymbol{\mu}_t \rangle + \langle \boldsymbol{\mu}_t \rangle \langle \boldsymbol{\mu}_t \rangle^T.$$

The data node relevance has consequently a quadratic logarithm:

$$\log \mathcal{I}(m) = -\frac{1}{2}(\mathbf{s}_m - \langle \boldsymbol{\mu}_t \rangle)^T \langle \boldsymbol{\lambda}_t \rangle (\mathbf{s}_m - \langle \boldsymbol{\mu}_t \rangle) + \text{const.}$$

Each node has a limited energy reserve and also a limited communication range. In addition, when communicating with other nodes, the leader node has to increase the transmission power to reach far message destinations. Given that the communication energy consumption is higher than the data processing energy consumption, the relevance criteria should involve a term penalizing long distances when selecting the next leader node. Therefore, the following selection criteria is suitable under sensor network energy constraints:

$$\begin{aligned} m_{t-1}^* &= \arg \min_m \mathcal{J}(m) \\ \mathcal{J}(m) &= \beta (\mathbf{s}_m - \langle \boldsymbol{\mu}_t \rangle)^T \langle \boldsymbol{\lambda}_t \rangle (\mathbf{s}_m - \langle \boldsymbol{\mu}_t \rangle) + \\ &\quad (1 - \beta) (\mathbf{s}_m - \mathbf{s}_{m_{t-1}^*})^T (\mathbf{s}_m - \mathbf{s}_{m_{t-1}^*}) \end{aligned} \quad (13)$$

where m_{t-1}^* is the leader node at time $t-1$ and the coefficient $\beta \in [0, 1]$ reflects the trade off between the data relevance and the communication cost.

Additional considerations have to be taken into account before the hand-off decision between two successive leaders such as conflicting situations when tracking multiple objects, forcing the hand-off when battery reserve is entirely consumed.... All these considerations are not reported here due to space limitation.

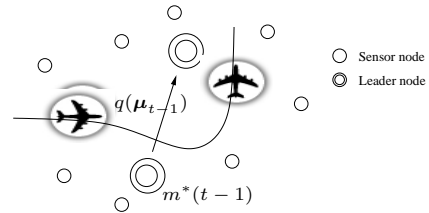


Fig. 2. The filtering distribution is updated by the leader node after receiving the mean and covariance of $q(\boldsymbol{\mu}_{t-1})$ by the previous leader node

IV. NUMERICAL RESULTS

In order to illustrate the effectiveness of the proposed ensemble learning target tracking in a wireless sensor network, we have considered the tracking of simulated sinusoidal trajectory in a 2-dimensional field (figure 3). A set of 200 nodes are randomly deployed in $100m \times 100m$ square area. Each node has a sensing range set to 20m. At each time t , within this range, the leader node obtains an observation of the target position through a range-bearing model:

$$y_t^r = \frac{p}{\|\mathbf{s}_m - \mathbf{x}_t\| + 0.5} + v_t^r$$

$$y_t^\theta = \arctan \frac{s_2 - x_2}{s_1 - x_1} + v_t^\theta$$

where $\mathbf{s}_m = (s_1, s_2)$ and $\mathbf{x}_t = (x_1, x_2)$ are the node and the target positions at time t , p (set to 10) is the energy emitted by the target (measured at a reference distance of 1 meter), v_t^r and v_t^θ are the corrupting noises due to modeling error, instrumental noise and background additive interfering signals. The noises v_t^r and v_t^θ are assumed white Gaussian with variances σ_ϵ^2 and σ_θ^2 respectively ($\sigma_\epsilon^2 = \sigma_\theta^2 = 0.05$).

The hyperparameters of the transition dynamical state model are set to the following values:

$$\bar{\mathbf{\lambda}} = 10^{-2} \mathbf{I}, \bar{n} = 100, \bar{\mathbf{S}} = 10^2 \mathbf{I},$$

where the hyperparameters values allow a general non informative prior. It is worth noting that in target tracking applications, an informative prior, involving the target velocity and acceleration, is usually assumed. Here, the transition prior has a more general form which can be used in other sensor network applications.

The ensemble learning tracking is applied on simulated data. Figure 3 depicts the estimated trajectory superimposed with the true sinusoidal trajectory. Note the accuracy of the target tracking. The position is estimated by the empirical mean of the point-mass approximate distribution $q(\mathbf{x}_t)$:

$$\hat{\mathbf{x}}_t = \sum_{i=1}^N w_t^{(i)} \mathbf{x}_t^{(i)}$$

On the same figure 3, the selected leaders are plotted in circles. Note the ability of the algorithm to select the relevant node while respecting the communication constraints. In fact, during the hand-off, the leader node sends the mean and the covariance of $q(\boldsymbol{\mu}_t)$ which requires 80 bits if all the parameters are transmitted up to 16 bits of resolution.

In figure 4, the leader selection criteria (13) used in the scheduling protocol is plotted within the set of randomly deployed nodes. The criteria is quadratic and therefore easy to implement. The same criteria was heuristically used in [8] exploiting only the second order statistics of the predictive distribution. The ensemble learning yields however a Gaussian predictive distribution approximating the predictive distribution in an optimal in the Kullback-Leibler metric.

V. CONCLUSION

An ensemble learning approach is proposed for online filtering in wireless sensor networks. At each time step, a leader node is selected according to a criteria representing a trade-off between data relevance and communication constraints. Approximation the filtering distribution by a separable free form distribution is implemented by a simple variational iterative algorithm. The approximate distribution yields a natural and adaptive compression of the filtering distribution which is propagated in the sensor network without lossy compression. In addition, the ensemble learning tracking outperforms the particle filter as it is implicitly based on a robust importance proposal.

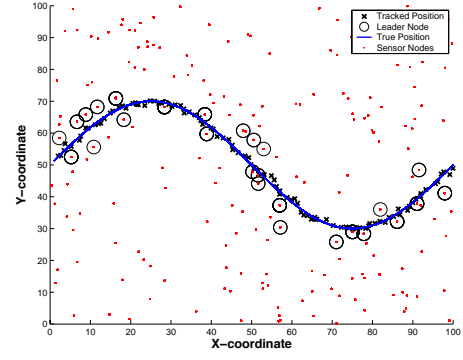


Fig. 3. Ensemble learning tracking in collaborative sensor network

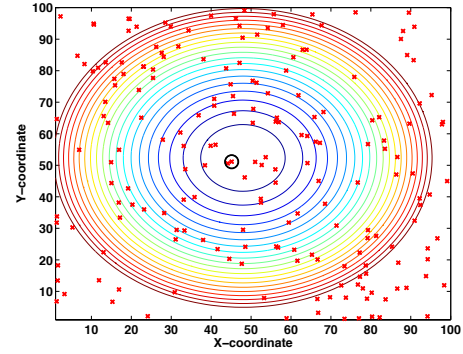


Fig. 4. Contour plot of the leader selection criteria. The selected node is plotted inside a circle.

REFERENCES

- [1] "Special issue on self-organizing distributed collaborative sensor networks", *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, 2005.
- [2] X. Sheng, Y. Hu, and P. Ramanathan, "Distributed particle filter with GMM approximation for multiple targets localization and tracking in wireless sensor network", in *Information Processing in Sensor Networks, 2005. IPSN 2005*, April 2005, pp. 181–188.
- [3] A. Ihler, J. Fisher III, and A. Willsky, "Particle filtering under communications constraints", in *Proc. Statistical Signal Processing (SSP) 2005*, 2005.
- [4] J. Vermaak, N. Lawrence, and P. Perez, "Variational inference for visual tracking", in *Conf. Computer Vision and Pattern Recog. CVPR'03*, June 2003.
- [5] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [6] O. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size", in *Proc. Roy. Soc., London*, 1977, vol. 353, pp. 401–419.
- [7] A. Doucet, B. Vo, C. Andrieu, and M. Davy, "Particle filtering for multitarget tracking and sensor management", in *Proc. Int. Conf. Inform. Fusion*, March 2002, vol. 1, pp. 474–481.
- [8] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration for tracking applications", *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, 2002.