Latent Parameter Estimation in Fusion Networks Using Separable Likelihoods

Murat Üney, *Member, IEEE,* Bernard Mulgrew, *Fellow, IEEE,* Daniel E. Clark, *Member, IEEE*

***Abstract*—Multi-sensor state space models underpin fusion applications in networks of sensors. Estimation of latent parameters in these models has the potential to provide highly desirable capabilities such as network self-calibration. Conventional solutions to the problem pose difficulties in scaling with the number of sensors due to the joint multi-sensor filtering involved in the evaluation of the parameter likelihood. We propose an approximation with a node-wise separable structure which can be evaluated by local filtering operations, instead. When leveraged with Markov random field models and message passing algorithms for inference, these likelihoods facilitate scalable estimation across the network and fit in both centralised and decentralised processing paradigms. We relate the approximation quality of the proposed separable likelihoods to the accuracy of state estimation using joint and local filtering schemes. We demonstrate this approach for network self-localisation using measurements from non-cooperative targets in an example.**

***Index Terms*—sensor networks, hidden Markov models, Markov random fields, pseudo-likelihood, simultaneous localisation and tracking, Monte Carlo algorithms, dynamical Markov random fields**

I. Introduction

A wide range of sensing applications including wide are surveillance are underpinned by state space models which are capable of representing a variety of dynamic phenomena such as spatio-temporal (see, e.g., [1]) and population [2] processes. In fusion (or, object tracking) networks, multi-sensor versions of stochastic state space models, also known as hidden Markov models [3], are used to estimate object trajectories in a surveillance region.

These models, however, are often specified by some latent parameters [4] some of which are unknown in practice and need to be estimated based on measurements from the state processes (or, objects). Examples of this problem setting in fusion networks include estimation of noise parameters [5], sensor biases [6] and localisation/calibration of sensors in a GPS denying environment using point detections of non-cooperative targets [7], [8]. Another example is the estimation of the orientations and positions of nodes in a camera network based on feature detections [9].

The parameter likelihood of the problem pose difficulties in the multi-sensor setting as its evaluation requires all sensor measurements across the network to be processed jointly in Bayesian filtering recursions. The poor scalability of likelihood evaluation arises as an issue in both maximum likelihood (ML) algorithms (e.g., [10]) and Bayesian estimation recursions (e.g., [11]). These difficulties are exacerbated when there are multiple objects with measurement association uncertainties: For the case, set valued state variables and multi-object distributions can be used, or, alternatively, these association variables can be treated as model parameters to be estimated jointly with the state trajectories. Evaluation of the parameter likelihood using multi-sensor filtering under either of modelling approaches is a problem with combinatorial complexity (see, e.g., [12] and [13]) motivating the development of strategies for efficient approximations.

An example of such strategies in the context of estimating parametric distributions from samples is the use of surrogate local functions in problems with likelihoods over high dimensional domains [14]–[16]. In the context of latent parameter estimation using samples from multiple measurement processes that are conditioned on underlying state space evolutions, we previously proposed separable likelihood functions which facilitate estimation using local filtering operations [17], [18]. These likelihoods consider two sensors and factorise into a pair of local terms leading to a dual-term likelihood approximation. When combined with pairwise Markov random fields (MRFs) –an undirected graph structure describing Markov (or, conditional independence) relations among the variables– and message passing algorithms for inference on them [19], efficient network wide estimation schemes are obtained. Such models and algorithms have been widely used for distributed estimation of various types of parameters in sensor networks (see, e.g., [20]–[22] and the references therein).

The main contribution of this work is an alternative separable likelihood for latent parameter estimation in multi-sensor state space models. This approximation similarly considers a pair of sensors and has a quadrature product structure each factor of which can be evaluated in terms of local filtering operations. As an indication of the approximation quality, we consider the Kullback-Leibler divergence (KLD) [23] of the proposed likelihood with respect to the actual likelihood obtained by joint filtering and relate it to the uncertainties in predicting and estimating the underlying state using individual and joint sensor histories. We show that with more accurate local filters the approximation quality improves and the proposed quad-term separable likelihood has an improved error bound compared to the aforementioned dual-term likelihood.

We equip pairwise MRFs with these likelihoods as edge potentials and nodes associated with latent parameters local to

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grants EP/J015180/1 and EP/K014277/1, and the MOD University Defence Research Collaboration in Signal Processing.

Murat Üney and Bernard Mulgrew are with the Institute for Digital Communications, School of Engineering, University of Edinburgh, EH9 3JL, Edinburgh, UK (e-mail: m.uney@ed.ac.uk, b.mulgrew@ed.ac.uk).

Daniel E. Clark is with the School of Engineering and Physical Sciences, Heriot–Watt University, EH14 4AS, Edinburgh, UK (e-mail: d.e.clark@hw.ac.uk).

sensors. The marginal distributions of this model approximates posterior marginals of the latent parameters to be estimated. We estimate these marginals iteratively using Belief Propagation (BP) [24] which consists of successive message passings among neighbouring nodes and updating of local marginals based on these messages. This computational structure lends itself for decentralised estimation, as well as scalable computation in a centralised fashion.

The realisation of the resulting algorithm is based on the nonparametric BP [25] approach and involves sampling from the updated marginals followed by separable likelihood evaluations in the message passing stage. As BP iterations converge to a fixed point, the empirical average of the samples from the marginals constitute (an approximate) minimum mean squared error (MMSE) estimate of the latent parameters. The edge potential evaluations are done using the entire measurement history within a selected time period in an offline fashion which is a strategy similar to particle Markov chain Monte Carlo (MC) algorithms [26]. As such, we differ from [18] in which windowing of measurements are used together with artificial parameter dynamics for enabling online processing.

Preliminary results of the proposed approach can be found in [27]. This article provides a complete account of our solution strategy and structured as follows: Section II provides the problem statement. Then, we introduce the quad-term node-wise separable likelihood approximation in Section III. Section IV details how these likelihoods are combined with pairwise MRFs to model the problem across the sensors and give details of a MC based implementation particularly for the estimation of sensor registration/calibration parameters. Then, we extend this framework for the case involving additional model parameters in Section V, which addresses, for example, estimation in the presence of multiple objects with measurement association uncertainties. In Section VI, we demonstrate the efficacy of our approach in network self-localisation using measurements from multiple non-cooperative targets. Finally, we conclude in Section VII.

II. Problem Definition

For simplicity in exposition but without loss of generality, we build the discussion upon a two sensors and a single Markov process case. Let us consider observation processes $\{Z_k^i\}_{k=1}^t$ and $\{Z_k^j\}_{k=1}^t$ associated with sensors i and j, respectively, together with a Markov process $\{X_k\}_{k=1}^t$ specified by an initial state distribution and a transition density, all spanning time from $k=1$ to t. The state space model with parameters θ is then specified as follows [28]: The state value x_k is a point in the state space $\mathcal{X}$ and is generated by the chain

$$\begin{aligned} X_k|(X_{1:k-1}=x_{1:k-1}) &\sim \pi(x_k|x_{k-1};\theta), \\ X_1 &\sim \pi_1(x_1;\theta), \end{aligned} \tag{1}$$

where $.|.$ denotes conditioning. Measured values z_k^i and z_k^js are points in $\mathcal{Z}^i$ and $\mathcal{Z}^j$ respectively, and, they are generated independently in accordance with the likelihood models

$$\begin{aligned} Z_k^i|(X_{1:k}=x_{1:k}, Z_{1:k}^i=z_{1:k}^i) &\sim g_i(z_k^i|x_k;\theta) \\ Z_k^j|(X_{1:k}=x_{1:k}, Z_{1:k}^j=z_{1:k}^j) &\sim g_j(z_k^i|x_k;\theta) \end{aligned} \tag{2}$$

where subscript $1:k$ indicates a vector concatenation over time.

The parameter vector $\theta \in \mathcal{B}$ can be used to represent a wide variety of parameters of the model some of which can be intrinsic to either sensor i or j alone such as parameters pertaining to their noise models. The rest can be classified as respective parameters such as sensor locations and similar "calibration" parameters. When θ is known, inference problem in this model often involves estimation of the distribution of X_k conditioned on the measurement histories, i.e., multi-sensor filtering. When θ is unknown, its likelihood measures a goodness of fit to the measurements and is evaluated via multi-sensor filtering [4, Sec.IV]:

$$l\left(z_{1:t}^i, z_{1:t}^j|\theta\right) = \prod_{k=1}^{t} p\left(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta\right), \tag{3}$$

where

$$\begin{aligned} &p\left(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta\right) = \\ &\int \left(g_i(z_k^i|x_k,\theta)g_j(z_k^j|x_k,\theta)\right) p(x_k|z_{1:k-1}^i, z_{1:k-1}^j, \theta)\mathrm{d}x_k. \end{aligned} \tag{4}$$

The equality in (3) follows from the chain rule of probabilities. The term in (4) is the contribution at time step k which updates the likelihood of the previous time step and is found using the Markov property that the sensor measurements are independent of the measurement histories, conditioned on the current state and θ. Let us denote this relation by $Z_k^j \perp\!\!\!\perp Z_{1:k-1}^j | X_k, \theta$ (see, e.g., [19]). The integrands in the expression for the likelihood update term in (4) are the multi-sensor likelihood and the prediction density for X_k based on the history of both sensors i and j. In other words, (4) is the scale factor for the posterior density of Bayesian recursions, or, the "centralised" filter:

$$\begin{aligned} &p(x_k|z_{1:k}^i, z_{1:k}^j, \theta) = \\ &\qquad \frac{\left(g_i(z_k^i|x_k,\theta)g_j(z_k^j|x_k,\theta)\right) p(x_k|z_{1:k-1}^i, z_{1:k-1}^j, \theta)}{p\left(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta\right)} \end{aligned} \tag{5}$$

$$\begin{aligned} &p(x_k|z_{1:k-1}^i, z_{1:k-1}^j, \theta) = \\ &\qquad \int \pi(x_k|x_{k-1},\theta) p(x_{k-1}|z_{1:k-1}^i, z_{1:k-1}^j, \theta)\mathrm{d}x_{k-1} \end{aligned} \tag{6}$$

We consider the mimimum mean squared error (MMSE) estimate of θ which can be found as the expected value of the posterior distribution

$$p(\theta|z_{1:k-1}^i, z_{1:k-1}^j) \propto l(z_{1:k-1}^i, z_{1:k-1}^j|\theta) p_0(\theta), \tag{7}$$

$$\hat{\theta} = \int_{\mathcal{B}} \theta p(\theta|z_{1:k-1}^i, z_{1:k-1}^j)\mathrm{d}\theta \tag{8}$$

This estimate can be computed by sampling from the posterior distribution in (7) using, for example, Markov Chain Monte Carlo methods [26] and using these samples to find the Monte Carlo estimate of the integral in (8) –which amounts to computing the empirical mean of these samples.

The main computational burden, however, is the joint filtering involved when evaluating the parameter likelihood.

This problem is exacerbated when multiple Markov processes (modelling multiple targets in fusion networks) and/or multiple sensors are involved, which are later discussed in Sections V and IV.

A similar issue arises for the maximum likelihood (ML) paradigm, as well. Consider the case in which $\mathcal{B}$ is bounded and let us select a uniform prior over $\mathcal{B}$. Under this modelling assumption –which is reasonably accurate for most practical problem settings– the MMSE estimate of θ is equivalent to its ML estimate given by

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\mathcal{B}} l(z^i_{1:k-1}, z^j_{1:k-1}|\theta).$$

The computational bottleneck for the solver of the optimisation problem above is the evaluation of the likelihood term, the complexity of which is dominated by filtering.

Therefore, we consider the problem of estimating θ in a way that scales with the number of targets and sensors. In the following sections, we introduce a solution which features scalability by building upon local filtering and message passing computational schemes.

III. Quad-term node-wise separable likelihoods

In this section, we introduce an approximation for the parameter likelihood which factorises into terms local to nodes i and j, i.e., a node-wise separable approximation. Let us consider the likelihood update term in (4). This term factorises in alternative ways as follows:

$$\begin{aligned}
&p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\\
&= p(z^i_k|z^j_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta)p(z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta) \quad (9)\\
&= p(z^j_k|z^i_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta)p(z^i_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta) \quad (10)\\
&= \Big(p(z^i_k|z^j_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta)p(z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\Big)^{1/2}\\
&\quad\times\Big(p(z^j_k|z^i_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta)p(z^i_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\Big)^{1/2} \quad (11)
\end{aligned}$$

In the first and second lines above, the chain rule is used. The third equality can be found by taking the geometric mean of the first two expressions. The conditioning of the four factors in Eq.(11) to the measurement histories of both sensors prevents decentralisation. In order to avoid this, let us leave out the history of sensor i (sensor j) in the first two (last two) terms of (11), i.e.,

$$\begin{aligned}
&q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\\
&\quad\triangleq \frac{1}{\kappa_k(\theta)}\Big(p(z^i_k|z^j_k, z^j_{1:k-1}, \theta)p(z^j_k|z^j_{1:k-1}, \theta)\Big)^{1/2}\\
&\qquad\times\Big(p(z^j_k|z^i_k, z^i_{1:k-1}, \theta)p(z^i_k|z^i_{1:k-1}, \theta)\Big)^{1/2} \quad (12)
\end{aligned}$$

$$\begin{aligned}
\kappa_k(\theta) = \int\int \Big(&p(z^i_k, z^j_k|z^j_{1:k-1}, \theta)\\
&\times p(z^i_k, z^j_k|z^i_{1:k-1}, \theta)\Big)^{1/2}\mathrm{d}z^i_k\mathrm{d}z^j_k \quad (13)
\end{aligned}$$

where $\kappa_k(\theta)$ is the normalisation constant that guarantees q to integrate to unity. Note that κ_k is a function of the parameters θ.

The appeal of this quadruple term is that the factors depend on single sensor histories. As such, they require filtering of sensor histories of i and j individually and in turn allow us to avoid centralisation. This point is discussed later in this section.

Next, we consider the difference between the update term (4) and the quad-term approximation introduced in (12). Because these terms are probability densities over the sensor measurements, their "divergence" can be quantified using the KLD [23]. For this reason, we incorporate θ in the joint probabilistic model that encompasses the HMM model as a random variable Θ associated with the prior density – equivalently, its marginal in the joint model– $p(\theta)$:

Proposition 3.1: The KLD between the centralised update and the node-wise separable approximation in (12) is bounded by the average of the mutual information (MI) [23] between the current measurement pair and a single sensor's history conditioned on the history of the other sensor, i.e.,

$$\begin{aligned}
&D(p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)||q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta))\\
&\qquad\leqslant \frac{1}{2}I(Z^i_k, Z^j_k; Z^i_{1:k-1}|Z^j_{1:k-1}, \Theta)\\
&\qquad\quad+\frac{1}{2}I(Z^i_k, Z^j_k; Z^j_{1:k-1}|Z^i_{1:k-1}, \Theta). \quad (14)
\end{aligned}$$

The proof can be found in Appendix A. The upper bound given in Proposition 3.1 measures the departure of the current measurements and one of the sensor histories from conditional independence when they are conditioned on the history of the other sensor. Note that these variables, when conditioned on X_k, are conditionally independent, i.e., $(Z^i_k, Z^j_k)\perp\!\!\!\perp Z^j_{1:k-1}|X_k, \Theta$ holds and consequently

$$I(Z^i_k, Z^j_k; Z^i_{1:k-1}|X_k, \Theta) = I(Z^i_k, Z^j_k; Z^j_{1:k-1}|X_k, \Theta) = 0.$$

Similarly, the average MI term on the right hand side of (14) is zero if $(Z^i_k, Z^j_k)\perp\!\!\!\perp Z^i_{1:k-1}|Z^j_{1:k-1}, \Theta$ and $(Z^i_k, Z^j_k)\perp\!\!\!\perp Z^j_{1:k-1}|Z^i_{1:k-1}, \Theta$ hold simultaneously. This condition is satisfied, for example, in the case that either of the measurement histories $Z^i_{1:k-1}$ and $Z^j_{1:k-1}$ are sufficient statistics for X_k (i.e., it can be predicted by both sensors with probability one). This level of accuracy should not be expected as the transition density of state space models introduce some uncertainty. Therefore, it is instructive to relate the KLD in (14) further to the uncertainty on X_k given the sensor histories:

Corollary 3.2: The KLD considered in Proposition 3.1 is upper bounded by the weighted sum of uncertainty reductions in the local target prediction and posterior distributions achieved when the other sensor's history is included jointly:

$$\begin{aligned}
&D\Big(p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)||q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\Big)\\
&\leqslant\frac{1}{2}\Bigg(\Big(H(X_k|Z^j_{1:k-1}, \Theta) - H(X_k|Z^j_{1:k-1}, Z^i_{1:k-1}, \Theta)\Big)\\
&\quad+\Big(H(X_k|Z^i_{k-1}, \Theta) - H(X_k|Z^j_{1:k-1}, Z^i_{1:k-1}, \Theta)\Big)\Bigg)\\
&\quad+\frac{1}{2}\Bigg(\Big(H(X_k|Z^j_{1:k}, \Theta) - H(X_k|Z^j_{1:k}, Z^i_{1:k-1}, \Theta)\Big)
\end{aligned}$$

$$+\left(H(X_k|Z^i_{1:k},\Theta)-H(X_k|Z^i_{1:k},Z^j_{1:k-1},\Theta)\right)\Bigg), \quad (15)$$

where H denotes the Shannon entropy [23].

The proof is provided in Appendix B. Corollary 3.2 relates the approximation quality of the quad-term node-wise separable updates to the uncertainties in the target state prediction and posterior distributions when individual node histories and their combinations are considered. The first term in the RHS of (15) is the sum of the uncertainty reductions in the predicted target state by taking the other sensor's history into account. The second term involves, similarly, the reductions in the target state estimation by the introduction of the other sensor's history. Therefore, a better quality of approximation should be expected for the cases that the local filtering densities involved concentrate around a single point in the state space. Equivalently, the quality of approximation is better if sensor i or j does not achieve a significant improvement in accuracy for predicting and estimating X_k when the measurement history of the other sensor is incorporated into local filtering.

It can be shown that the upper bound in (14) for the quad-term approximation is smaller than the exact KLD of the aforementioned dual term approximation [18]. In other words, the quad-term approximation is closer to the centralised update compared to the dual-term approximation, in terms of its KLD. On the other hand, the scaling factor of the dual term approximation is unity regardless of θ thereby admitting a significant amount of flexibility in the range of the distributions and likelihoods that can be accommodated in the state space model, for example, random finite set models [18]. Evaluation of the scaling factor in (13) might not be straightforward for general problem settings.

A node-wise separable likelihood built upon the quad-term update is obtained after substituting q as the update term in (3). Let us consider (12) and define

$$\tilde{r}^k_{ij}(\theta) \triangleq p(z^i_k|z^j_k, z^j_{1:k-1},\theta) \quad (16)$$

$$p(z^i_k|z^j_k, z^j_{1:k-1},\theta) = \int p(z^i_k|x_k,\theta)p(x_k|z^j_{1:k},\theta)\mathrm{d}x_k \quad (17)$$

$$\tilde{s}^k_j(\theta) \triangleq p(z^j_k|z^j_{1:k-1},\theta), \quad (18)$$

$$p(z^j_k|z^j_{1:k-1},\theta) = \int p(z^j_k|x_k,\theta)p(x_k|z^j_{1:k-1},\theta)\mathrm{d}x_k. \quad (19)$$

Let us also express the normalisation factor in (13) equivalently as

$$\kappa^{ij}_k(\theta) = \int\int \left(\tilde{r}^k_{ij}(\theta)\tilde{s}^k_j(\theta)\right)^{1/2}\left(\tilde{r}^k_{ji}(\theta)\tilde{s}^k_i(\theta)\right)^{1/2}\mathrm{d}z^i_k\mathrm{d}z^j_k \quad (20)$$

Using these terms, the corresponding approximate separable likelihood is given by

$$\begin{aligned}\tilde{l}\left(z^i_{1:t}, z^j_{1:t}|\theta\right) &= \prod_{k=1}^{t} q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1},\theta)\\ &= \prod_{k=1}^{t}\frac{1}{\kappa^{ij}_k(\theta)}\left(\tilde{r}^k_{ij}(\theta)\tilde{s}^k_j(\theta)\right)^{1/2}\left(\tilde{r}^k_{ji}(\theta)\tilde{s}^k_i(\theta)\right)^{1/2}. \quad (21)\end{aligned}$$

We refer to this likelihood as the quad-term separable likelihood as it can also be expressed as a (scaled) product of four factors each of which are the products of the four terms inside the product in (21) over k.

Let us consider the update terms (17) and (19) which are based on the measurement history of sensor j. They are computed using a Bayesian recursive filter's posterior and the prior distributions, respectively, which process only the measurements of sensor j. Computation of $\tilde{r}^k_{ji}$ and $\tilde{s}^k_i$ similarly requires only single sensor filtering of sensor i's history.

As a result, the approximate likelihood can be evaluated in a distributed fashion: Sensor node j performs local filtering with its measurements and obtain posterior distributions $p(x_k|z^j_{1:k},\theta)$ for $k = 1,\ldots,t$ that capture the local information on the trajectory of the underlying state process (or, "target track" in fusion networks). Simultaneously $\tilde{s}^k_j$ terms are computed using (18) and (19). Then, sensor j transmits its posteriors to sensor i which, then, computes $\tilde{r}^k_{ij}$s using (16) and (17), and, vice versa. Finally, all terms are exchanged to find the normalisation factors in (20) and the quad-term separable likelihood in (21).

Corollary 3.3: The KLD between the parameter likelihood in (3) and the node-wise separable approximation in (21) is a sum of the KLD terms considered Proposition 3.1, i.e.,

$$\begin{aligned}&D\left(l\left(z^i_{1:t}, z^j_{1:t}|\theta\right)||\tilde{l}\left(z^i_{1:t}, z^j_{1:t}|\theta\right)\right) =\\ &\sum_{k=1}^{t} D\left(p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1},\theta)||q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1},\theta)\right)\end{aligned} \quad (22)$$

and bounded by the terms on the right hand sides of (14) and (15) summed over $k = 1,\ldots,t$.

Proof. Eq. (22) can easily be found after expanding the KLD term explicitly and expressing the logarithm of products involved as sums over logarithms of the factors. Boundedness follows from non-negativity of KLDs and summing both sides of (14) and (15) over $k = 1,...,t$. ■

IV. Pairwise Markov random field models for scalable network wide estimation

The benefits of the quad-term node-wise separable likelihood introduced in the previous section can be extended to the case of $N > 2$ sensors by using the product of separable likelihoods for pairs of sensors. The filtering cost is linear with the number of sensors regardless of the number of quad-terms to be evaluated. Selecting all possible $\binom{N}{2}$ pairs from $\mathcal{V} = \{1,\ldots,N\}$ to contribute in this model might undermine the computational efficiency gained. We are interested in cases in which the selected pairs $(i,j) \subset \mathcal{V}$ render a connected planar graph thereby yielding the number of pairs less than $\mathcal{O}(N^2)$. For example, the Gabriel graph, also known as the relative neighbourhood graph, of the sensor nodes is such a graph which is exploited in geographic routing for sensor networks [29]. The minimum number of quad-terms one can select is the number of edges in a spanning tree, i.e., $N-1$.

A particularly interesting case is when θ is a concatenation of unknown parameters local to nodes, i.e., $\theta = [\theta_1,...,\theta_N]$. Let us denote by $\mathcal{E}$ the set of pairs for which the quad-term

Algorithm 1 Conceptual LBP iterations for estimating θ over a MRF model equipped with quad-term separable likelihoods.

for all $j \in \mathcal{V}$ **do** ▷ Local filtering
  for $k = 1, \ldots, t$ **do**
    Find $p(x_k|z^j_{1:k}, \theta_j)$
    Find $\tilde{s}^k_j(\theta_j)$ in (18) and (19).
  end for
end for
for all $(i,j) \in \mathcal{E}$ **do** ▷ Find edge potentials
  for $k = 1, \ldots, t$ **do**
    Find $\tilde{r}^k_{ij}(\theta_i, \theta_j)$ in (16) and (17)
    Find $\tilde{r}^k_{ji}(\theta_i, \theta_j)$
    Find $\kappa^{ij}_k(\theta_i, \theta_j)$ in (20)
  end for
end for
Compute $\psi^t_{i,j}(\theta_i, \theta_j)$ in (24) using (21)
for $s = 1, ..., S$ **do** ▷ S-steps of LBP
  for all $(i,j) \in \mathcal{E}$ **do** ▷ Find LBP messages
    Find $m_{ji}(\theta_i)$ in (25)
  end for
  for all $i \in \mathcal{V}$ **do** ▷ Update local marginals
    Find $\tilde{p}_i(\theta_i)$ in (26)
    $\hat{\theta}_i \leftarrow E_{\tilde{p}_i}\{\theta_i\}$
  end for
end for

likelihood will be incorporated and by $\mathcal{V}$ the set of sensor nodes. Together with *a priori* distributions selected for Θ_is, the corresponding parameter posterior is a pairwise Markov random field over $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [19]:

$$p(\theta|Z^1_{1:t}, ..., Z^N_{1:t}) \propto \prod_{i\in\mathcal{V}} \psi_i(\theta_i) \prod_{(i,j)\in\mathcal{E}} \psi^t_{ij}(\theta_i, \theta_j), \quad (23)$$

$$\psi_i(\theta_i) = p_{0,i}(\theta_i),$$
$$\psi^t_{ij}(\theta_i, \theta_j) = \tilde{l}(z^i_{1:t}, z^j_{1:t}|\theta_i, \theta_j), \quad (24)$$

where the node potential functions (i.e., ψ_is) are arbitrary priors for θ_i (e.g., uniform distributions over bounded sets θ_is take values from) and the edge potentials (i.e., ψ^t_{ij}s) are the quad-term separable likelihoods for the pairs (i,j)s based on sensor histories up to time t.

This model is illustrated in Fig. 1 and facilitates the estimation of θ across the network: The pairwise MRF model in (23) allows the computation of the marginal densities through iterative local message passings such as Belief Propagation (BP) [24]. In BP, the nodes maintain distributions over their local variables and update them based on messages from their neighbours which summarise the information neighbours have gained on these variables. This is described for all $i \in \mathcal{V}$ by

$$m_{ji}(\theta_i) = \int \psi^t_{ij}(\theta_i, \theta_j)\psi_j(\theta_j) \prod_{i'\in ne(j)\setminus i} m_{i'j}(\theta_j)\, \mathrm{d}\theta_j, \quad (25)$$

$$\tilde{p}_i(\theta_i) \propto \psi_i(\theta_i) \prod_{j\in ne(i)} m_{ji}(\theta_i). \quad (26)$$

In BP iterations, nodes simultaneously send messages to their neighbours using (25) (often starting with constants as the previously received messages) and update their local "belief" using (26). If $\mathcal{G}$ contains no cycles (i.e., $\mathcal{G}$ is a tree), $\tilde{p}_i$s are guaranteed to converge to the marginals of (23), in a finite number of steps [24]. For the case in which $\mathcal{G}$ contains cycles iterations of (25) and (26) are known as loopy BP (LBP). For the case, convergence is not guaranteed, in general, nevertheless LBP has been been very successful in computing approximate marginals in a distributed fashion, in fusion, self-localisation and tracking problems in sensor networks [20], [21], [30]. In our problem setting, we assume that the models over spanning trees of a loopy $\mathcal{G}$ are consistent in that they lead to "similar" marginal parameter distributions, which suggests the existence of LBP fixed points [31] that will be converged when initial beliefs are selected reasonably [32].

The resulting conceptual algorithm is given in Algorithm 1. Let us consider the pairwise MRF model in (23) equipped with quad-term node-wise separable likelihoods. First, the sensor histories are filtered individually and the filtered posteriors as a function of the (continuous) state variable as well as the unknown parameter is obtained. Simultaneously, the s^k_j terms of the node-wise separable likelihood are computed. Next, r^k_{ij} and r^k_{ji} are found for neighbouring nodes i and j, using the measurements and the filtered posteriors previously found. This stage results with the edge potentials of the MRF model. Then, starting from initial marginal distributions, S step of (loopy) BP is used to iteratively improve the posterior marginal distributions, the expected value of which provides us with the estimates of the latent parameters.

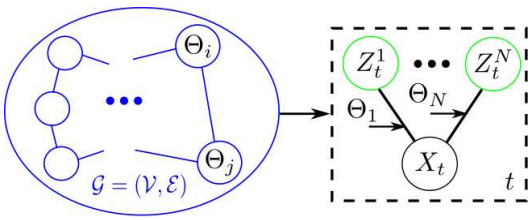

Fig. 1. A multi-sensor state space - or, hidden Markov- model (black dashed box on the right representing a chain over t) parameterised with a Markov Random field (the blue edges on the left).

A. A Monte Carlo LBP algorithm for sensor registration in fusion networks

The conceptual steps in Algorithm 1 treat the functions involved as entities that are found and stored over the entire continuous domain from which their arguments take values from. In this section, we introduce particle based representations and Monte Carlo computations [33] for the realisation of these steps in registration problems.

In this setting, the local parameters signify respective quantities such as sensor locations and orientations, and appear only in the measurement model given in (2), i.e., the state transition model in (1) is not conditioned on θ. Moreover, these parameters specify a coordinate transform among local frames of the sensors. Let x_k be a point in the reference

coordinate frame and $[x_k]_j$ denote the corresponding point in the coordinate frame of sensor j. Then, $[x_k]_j$ and $[x_k]_i$ are given in terms of a coordinate transform τ respectively as

$$[x_k]_j = \tau(x_k;\theta_j), \tag{27}$$
$$[x_k]_i = \tau(\tau^{-1}([x_k]_j;\theta_j);\theta_i), \tag{28}$$

As an example, when the state space consists of locations on a plane and θ_j is the position of sensor j, this transform is given by

$$\tau(x_k;\theta_j) = x_k - \theta_j,$$
$$\tau(\tau^{-1}([x_k]_j;\theta_j);\theta_i) = [x_k]_j + \theta_j - \theta_i.$$

Let us consider the local filtering stage. Because the sensor measurements are collected in local coordinate frames, filtering of them results with posteriors in the local coordinate frames. For the case, $\tilde{s}_j^k$ does not vary with θ_j, and, is a density over local measurements which evaluate to the scale factor (or, evidence) in the denominator of the update stage in Bayesian filtering at the current measurement, i.e.,

$$\tilde{s}_j^k = \int p(z_k^j|[x_k]_j)p([x_k]_j|z_{1:k-1}^j)\mathrm{d}[x_k]_j. \tag{29}$$

The $\tilde{r}_{ij}^k(\theta_i,\theta_j)$ terms can be found from the filtered posteriors using the transformation in (27) for a given pair of values for the parameters, i.e., (16),(17) specialises to

$$\tilde{r}_{ij}^k(\theta_i,\theta_j) = \int p\left(z_k^i|\tau(\tau^{-1}([x_k]_j;\theta_j);\theta_i)\right)p([x_k]_j|z_{1:k}^j)\mathrm{d}[x_k]_j. \tag{30}$$

As a result, local filtering of sensor histories can be performed regardless of the sample generation process for the LBP steps, in this problem setting.

Suppose that we are able evaluate (30),the scale factor (20) and, consequently, the node-wise separable likelihood (21) for any given values of the parameters. Now, let us adopt the sampling approach detailed in [18, Sec.VI] for carrying out LBP belief update and messaging in (26) and (25), respectively. Given L equally weighted samples from $\tilde{p}_i(\theta_i)$, i.e.,

$$\theta_i^{(l)} \sim \tilde{p}_i(\theta_i), \text{ for } l = 1,\ldots,L, \tag{31}$$

the edge potetials are evaluated to obtain

$$\psi_{ij}^t(\theta_i^{(l)},\theta_j^{(l)}) \text{ for } l = 1,\ldots,L. \tag{32}$$

Consider the BP message from node j to i in (25). Suppose that independent identically distributed (i.i.d.) samples from the (scaled) product of the jth local belief and the incoming messages from all neighbours except i are given, i.e.,

$$\bar{\theta}_j^{(l)} \sim \tilde{p}_j(\theta_j)\prod_{i'\in ne(j)/i} m_{i'j}(\theta_j) \text{ for } l = 1,...,L. \tag{33}$$

These samples are used with kernel approximations in order to represent the message from node j to i (scaled to one), in the NBP approach [25]. We use Gaussian kernels leading to the approximation given by

$$\hat{m}_{ji}(\theta_i) = \sum_{l=1}^{L}\omega_{ji}^{(l)}\mathcal{N}(\theta_i;\theta_{ji}^{(l)},\Lambda_{ji}), \tag{34}$$

$$\theta_{ji}^{(l)} = \tau(\tau^{-1}(\bar{\theta}_j^{(l)};\theta_j^{(l)});\theta_i^{(l)}),$$
$$\omega_{ji}^{(l)} = \frac{\psi_{i,j}^t(\theta_i^{(l)},\theta_j^{(l)})}{\sum_{l'=1}^{L}\psi_{i,j}^t(\theta_i^{(l')},\theta_j^{(l')})},$$

where the kernel weights are the normalised edge potentials. Λ_{ji} is related to a *bandwidth* parameter that can be found using Kernel Density Estimation (KDE) techniques. In particular, we use the rule-of-thumb method in [34] and find

$$\Lambda_{ji} = \left(\frac{4}{(2d+1)L}\right)^{2/(d+4)}\hat{\mathbf{C}}_{ji},$$
$$\hat{\mathbf{C}}_{ji} = \sum_{l'}\sum_{l}\omega_{ji}^{(l')}\omega_{ji}^{(l)}(\theta_{ji}^{(l')}-\hat{\mathbf{m}}_{ji})(\theta_{ji}^{(l)}-\hat{\mathbf{m}}_{ji})^T,$$
$$\hat{\mathbf{m}}_{ji} = \sum_{l=1}^{L}\omega_{ji}^{(l)}\theta_{ji}^{(l)}$$

where $\hat{\mathbf{m}}_{ji}$ and $\hat{\mathbf{C}}_{ji}$ are the empirical mean and covariance of the samples, respectively, and d is the dimensionality of θ_{ji}s.

Given these messages, let us consider sampling from the updated marginal in (26). We use the weighted bootstrap (also known as sampling/importance resampling) [35] with samples generated from the (scaled) product of Gaussian densities with mean and covariance found as the empirical mean and covariance of the particle sets, respectively. In other words, given $\hat{\mathbf{m}}_{ji}$ and $\hat{\mathbf{C}}_{ji}$ as above, we generate

$$\theta_i^{(l)} \sim f(\theta_i), \quad l = 1,\cdots,L,$$
$$f(\theta_i) \propto \mathcal{N}(\theta_i;\hat{\mathbf{m}}_i,\hat{\mathbf{C}}_i)\prod_{j\in ne(i)}\mathcal{N}(\theta_i;\hat{\mathbf{m}}_{ji},\hat{\mathbf{C}}_{ji}).$$

The particle weights for these samples to represent the updated marginal is given by

$$\omega_i^{(l)} = \hat{\omega}_i^{(l)}/\sum_{l'=1}^{L}\hat{\omega}_i^{(l')}$$
$$\hat{\omega}_i^{(l)} = \left(p_{0,i}(\theta_i^{(l)})\prod_{j\in ne(i)}\hat{m}_{ji}(\theta_i^{(l)})\right)/f(\theta_i^{(l)})$$

where $p_{0,i}$ is the prior density selected for θ_i (and, the node potential in (23)). Thus, the local calibration marginal is estimated by

$$\hat{P}_i(\mathrm{d}\theta_i) = \sum_{l=1}^{L}\omega_i^{(l)}\delta_{\theta_i^{(l)}}(\mathrm{d}\theta_i). \tag{35}$$

As the final step of the bootstrap, $\{\theta_i^{(l)},\omega_i^{(l)}\}_{l=1}^M$ is resampled (with replacement) leading to equally weighted particles from $\tilde{p}_i(\theta_i)$, i.e., $\{\theta_i^{(l)}\}_{l=1}^L$. We follow similar bootstrap steps in order to generate the samples in (33).

After nodes iterate the BP computations described above for S times, each node estimates its location by finding the empirical mean of $\{\theta_i^{(l)}\}_{l=1}^L$. These steps are summarised in Algorithm 2.

V. Joint model selection and parameter estimation for multiple Markov chains

The complexity of joint multi-sensor Bayesian recursions (or, centralised filtering) can be severly restrictive when the

Algorithm 2 Pseudo-code for estimation of θ using separable likelihoods within Belief Propagation.

for all $j \in \mathcal{V}$ **do** ▷ Local filtering
  for $k = 1, \ldots, t$ **do**
    Find $p(x_k|z^j_{1:k})$
    Find $\bar{s}^k_j$ in (29).
  end for
end for
for all $j \in \mathcal{V}$ **do** ▷ Sample from priors
  Sample $\theta^{(l)}_i \sim p_{0,i}(\theta_i)$ for $l = 1, \ldots, L$
end for
for $s = 1, ..., S$ **do** ▷ S-steps of LBP
  for all $(i, j) \in \mathcal{E}$ **do** ▷ Evaluate edge potentials
    for $l = 1, \ldots, L$ **do**
      Find $\tilde{r}^k_{ij}(\theta^{(l)}_i, \theta^{(l)}_j)$ in (30) for $k = 1, \ldots t$
      Find $\tilde{r}^k_{ji}(\theta^{(l)}_i, \theta^{(l)}_j)$ for $k = 1, \ldots t$
      Find $\kappa^{ij}_k(\theta^{(l)}_i, \theta^{(l)}_j)$ in (20) for $k = 1, \ldots t$
      Find $\psi^t_{i,j}(\theta^{(l)}_i, \theta^{(l)}_j)$ in (24) using (21)
    end for
  end for
  for all $(i, j) \in \mathcal{E}$ **do** ▷ Find LBP message
    Find the kernel representation $\hat{m}_{ji}(\theta_i)$ in (34)
  end for
  for all $i \in \mathcal{V}$ **do** ▷ Update local marginals
    Find the updated $\hat{P}_i$ in (35) and sample $\theta^{(l)}_i \sim \tilde{p}_i(\theta_i)$
    $\hat{\theta}_i \leftarrow \frac{1}{L}\sum_{l=1}^{L} \theta^{(l)}_i$
  end for
end for

parameter space of concern is, for example, the (Cartesian) product of an Euclidean space and a discrete space that grows combinatorially with the problem size. Such problems can arise, for examples, when there are unknown model parameters along with the latent parameters.

In fusion networks, a typical problem in this setting is inference in the case of multiple Markov chains possibly modelling multiple objects. The difficulty stems form the measurement-object association uncertainties in this case [36]. These associations can be parameterised in terms of model selection parameters which have a combinatorially increasing number of possible configurations [13]. The exact solution requires the likelihood for any given selection of calibration parameters and association configurations to be evaluated using joint filtering. In this section, we show how the use of separable likelihood approximation leads to a decomposition of the network wide association problem into local problems, removing the need for centralised filtering in likelihood evaluations.

In order to elaborate on this point, let us consider the update term in (4) for M targets constituting a concatenated meta-state variable $\bar{x}_k \triangleq [x^1_k, \ldots, x^M_k]$. For the sake of simplicity, we assume that each object induces only one measurement and does so with probability one (i.e., the probability of detection is one) and there are no false alarms (or, non-target detections). Let us denote the concatenation of such measurements at sensors i collected at time k by $\bar{z}^i_k$. This term is given by $\bar{z}^i_k \triangleq [z^{i,1}_k, \ldots, z^{i,M}_k]$.

We also assume that all targets that exists at $k = 1$ continue to exist up until $k = t$ and no new targets appear in this period of time. After substituting $\bar{x}_k$, $\bar{z}^i_k$ and $\bar{z}^j_k$ in (4), the instantenous update term is obtained as

$$p\left(\bar{z}^i_k, \bar{z}^j_k | \bar{z}^i_{1:k-1}, \bar{z}^j_{1:k-1}, \theta\right) = \prod_{m=1}^{M} \int \left(g_i(z^{i,\tau^i_k(m)}_k | x^m_k, \theta) g_j(z^{j,\tau^j_k(m)}_k | x^m_k, \theta) \right) \times p(x^m_k | \bar{z}^i_{1:k-1}, \bar{z}^j_{1:k-1}, \theta) \mathrm{d}x^m_k \quad (36)$$

where τ^i_k is a permutation of $(1, \ldots, M)$ at time k such that its mth element $\tau^i_k(m)$ is the index of sensor i's measurement that is induced by target m. Let us denote the association history of this target by $\tau^i_{1:k-1}(m)$ and the measurements it selects from the history $\bar{z}^i_{1:k-1}$ by $z^{i,m}_{1:k-1}$. The prediction density in (36) is given by

$$p(x^m_k | \bar{z}^i_{1:k-1}, \bar{z}^j_{1:k-1}, \theta) = p(x^m_k | z^{i,m}_{1:k-1}, z^{j,m}_{1:k-1}, \theta) \quad (37)$$

where the right hand side is simply (6) with relavant terms substituted.

It is worth noting that all the terms on the right hand sides of (36) and (37) are parameterised by $\tau^i_{1:k}$ and $\tau^j_{1:k}$. In other words, the product of (36) over $k = 1, \ldots, t$ is the likelihood for the joint parameter defined by $\beta \triangleq [(\tau^i_{1:t}, \tau^j_{1:t}), \theta]$ for which the results in Section III hold.

One possible way to proceed with estimating θ is to use its marginal likelihood in which the discrete (model parameter) terms are marginalised (using a non-informative prior[1]), i.e.,

$$l(\bar{z}^i_{1:t}, \bar{z}^i_{1:t} | \theta) = \sum_{(\tau^i_{1:t}, \tau^j_{1:t})} l(\bar{z}^i_{1:t}, \bar{z}^i_{1:t} | \beta = [\tau^i_{1:t}, \tau^j_{1:t}, \theta]) p(\tau^i_{1:t}, \tau^j_{1:t}) \quad (38)$$

within an estimation scheme.

There are two main challenges, however. The first is that the number of possible configurations that can be selected jointly for (τ^i_k, τ^j_k)s is $(M!)^2$ [13] and this quantity to the power of k gives the number of elements in the summation in (38). Therefore, the number of likelihood evaluations for marginalisation renders this operation infeasible. The second is that, when estimating θ with this likelihood, this marginalisation has to be performed separately for each value of θ hypothesised within the estimation scheme, adding to the computational complexity of its use.

A reasonable approximation to this marginal likelihood is to use the joint likelihood with the substitution of an estimate of the discrete variables. This approach is similar to the empirical Bayes approach [37] and amounts to using

$$l(\bar{z}^i_{1:t}, \bar{z}^i_{1:t} | [\hat{\tau}^i_{1:t}, \hat{\tau}^j_{1:t}, \theta]) \approx l(\bar{z}^i_{1:t}, \bar{z}^i_{1:t} | \theta)$$

with the maximum a posterirori (MAP) (and, equivalently, the

[1]Because the permutations constituting $(\tau^i_{1:t}, \tau^j_{1:t})$ are *a priori* equally likely, a uniform prior assigning equal probability to every possible configuration is non-informative, in this setting [13].

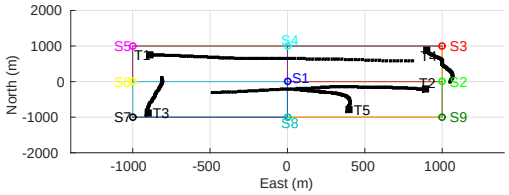

Fig. 2. Example scenario: Sensors (S1-S9) collect measurements from an object (T1) with initial position denoted by the black square. The blue lines depict the MRF model for the estimation task.

ML) estimate of the discrete model parameters given by

$$(\hat{\tau}^i_{1:t}, \hat{\tau}^j_{1:t}) = \arg \max_{\tau^i_{1:t}, \tau^j_{1:t}} p(\tau^i_{1:t}, \tau^j_{1:t} | \bar{z}^i_{1:t}, \bar{z}^i_{1:t}, \theta)$$
$$\equiv \arg \max_{\tau^i_{1:t}, \tau^j_{1:t}} \log l(\bar{z}^i_{1:t}, \bar{z}^i_{1:t} | \tau^i_{1:t}, \tau^j_{1:t}, \theta) \quad (39)$$

The optimisation problem above can be treated as a multi-dimensional assignment [36], and, has combinatorial complexity, in general. One approximate solution approach is to use relaxations of (39) over time, which lead to coupled 2 dimensional assignment problems each of which can be solved using efficient algorithms such as the Auction algorithm [38]. This approach is also equivalent to 1-scan multiple hypothesis tracking (MHT) algorithm with the aforementioned assumptions. The computational load, however, is dominated by filtering which takes place during the computation of the assignment costs [36]. The fact that there are two sensor histories for joint filtering means that the sizes of these matrices are $\mathcal{O}(M^3)$ and are found for every selected value of θ and for $k = 1, \ldots, k$.

When the proposed separable likelihoods are used, on the other hand, a relaxation of (39) is implied in which we solve for $\tau^i_{1:k}$ and $\tau^j_{1:k}$ separately, using only local sensor histories $z^i_{1:k}$ and $z^j_{1:k}$, respectively. The 2 dimensional assignment relaxations of these local problems lead to cost matrices of size $\mathcal{O}(M^2)$ for $k = 1, \ldots, k$. Finally, because the model variables now signify a local labelling of the chains, their correspondance needs to found which amounts to finding the ML setting of (τ^i_1, τ^j_1). This can be viewed as track assignment [36] the size of which remains $\mathcal{O}(M^2)$, as well.

VI. Example: Self-localisation in Linear Gaussian State space models

In this example, we consider a state space model which is linear with additive Gaussian uncertainties given (unknown) sensor locations θ_is, i.e.,

$$\pi(x_k | x_{k-1}) = \mathcal{N}(x_k; \mathbf{F} x_{k-1}, \mathbf{Q}) \quad (40)$$
$$g_i(z^i_k | x_k; \theta_i) = \mathcal{N}(z^i_k; \mathbf{H}_i(x_k - \theta_i), \mathbf{R}_i) \quad (41)$$

where $\mathcal{N}(.; \mathbf{m}, \boldsymbol{P})$ is a multi-dimensional Gaussian density with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{P}$. The linear transformation $\mathbf{F}$ specifies the state transition dynamics whereas $\mathbf{Q}$ is the process noise covariance matrix, i.e.,

$$\pi(x_n | x_{n-1}) = \mathcal{N}(x_n; F x_{n-1}, Q)$$
$$F = \begin{bmatrix} \mathbf{I}, & \mathbf{I} \\ \mathbf{0}, & \mathbf{I} \end{bmatrix}, \; Q = 0.25 \begin{bmatrix} \frac{1}{4}\mathbf{I}, & \frac{1}{2}\mathbf{I} \\ \frac{1}{2}\mathbf{I}, & \mathbf{I} \end{bmatrix}$$

where $\mathbf{I}$ and $\mathbf{0}$ are the 2×2 identity and zero matrices, respectively.

The measurement model is parameterised by the observation matrix $\mathbf{H}_i$ and the observation noise covariance $\mathbf{R}_i$, both associated with sensor i. For the sake of simplicity we use $\mathbf{H}_i = \mathbf{I}$ and $\mathbf{R}_i = \sigma^2 \mathbf{I}$ with $\sigma = 10$.

We consider the scenario depicted in Fig. 2. Nine sensors observe five moving objects. We select $\mathbf{F}$ and $\mathbf{Q}$ for each object to model near constant velocity motion with random acceleration and $\mathbf{H}$ and $\mathbf{R}$ to model noisy position measurements from the object. Location of sensor 1 is selected as the origin and the respective locations of the remaining sensors are to be estimated.

We specify the steps of Algorithm 2 for this example. In the local filtering stage, the association problem to find $\tau^j_{1:k}$s are solved using the Auction algorithm [38] as explained in Section V. Filtering for finding the association costs is carried out using Kalman filtering (KF) recursions –owing to the conditionally linear/Gaussian nature of the model. Specifically, for each sensor j and object $m \in \tau^j_1$, we use a KF with the history $z^{j,m}_{1:k}$ which yields the following Gaussian prediction and posterior distributions at time k:

$$p(x^{(m)}_k | z^{j,(m)}_{1:k-1}; \theta_{i,j}) = \mathcal{N}(x^{(m)}_k; \hat{\mathbf{x}}^{j,m}_{k|k-1} + \theta_j, \mathbf{P}^j_{k|k-1}),$$
$$p(x^{(m)}_k | z^{j,(m)}_k, z^{j,(m)}_{1:k-1}; \theta_{i,j}) = \mathcal{N}(x^{(m)}_k; \hat{\mathbf{x}}^j_{k|k} + \theta_j, \mathbf{P}^j_{k|k}).$$

The update term in (29) is a product of update terms for M objects, i.e.,

$$p(\bar{z}^j_k | \bar{z}^j_{1:k-1}) = \prod_{m=1}^{M} p(z^{j,\hat{\tau}^k_j(m)}_k | z^{j,m}_{1:k-1})$$
$$p(z^{j,\hat{\tau}^k_j(m)}_k | z^{j,m}_{1:k-1}) = \mathcal{N}(z^{j,\hat{\tau}^k_j(m)}_k; \mathbf{H}_i(\hat{\mathbf{x}}^{j,m}_{k|k-1} + \theta_j - \theta_i), \mathbf{S}_{k|k-1})$$
$$\mathbf{S}_{k|k-1} = \mathbf{R}_i + \mathbf{H}_i \mathbf{P}^{j,m}_{k|k-1} \mathbf{H}^T_i$$

Node 1 is selected as the origin of the network coordinate system. For the other nodes, the localisation prior, i.e., $p_{0,i}(\theta_i)$, is selected to be a uniform distribution over the sensing region and the initial set of samples (31) are selected as random permutations of an $L = 100$ point uniform grid over the sensing region. We consider the pairwise graph $\mathcal{G}$ in Fig. 2 depicted with blue edges.

For the evaluation of the proposed likelihoods in Algorithm 2, we first use the posterior target densities generated in the local filtering phase to find the label association, i.e., $(\hat{\tau}^j_1, \hat{\tau}^i_1)$, again using the Auction algorithm for 2 the corresponding two dimensional assignment problem. The update term in (30) is similarly a product of terms for different objects given by

$$p(\bar{z}^i_k | \bar{z}^j, \bar{z}^j_{1:k-1}, \theta_{i,j}) = \prod_{m=1}^{M} p(z^{i,\hat{\tau}^i_k(m)}_k | z^{j,m}_{1:k-1}, \theta_{i,j})$$
$$p(z^{i,\hat{\tau}^i_k(m)}_k | z^{j,m}_{1:k-1}, \theta_{i,j}) = \mathcal{N}(z^{i,\hat{\tau}^i_k(m)}_k; \mathbf{H}_i(\hat{\mathbf{x}}^{j,m}_{k|k} + \theta_j - \theta_i), \mathbf{S}_k)$$
$$\mathbf{S}_k = \mathbf{R}_i + \mathbf{H}_i \mathbf{P}^{j,m}_{k|k} \mathbf{H}^T_i$$

After multiplying the update terms, the scaling factor $\kappa^{ij}_k(\theta_{i,j})$ in (20) can be found in closed form using integration rules of quadratic exponentials.

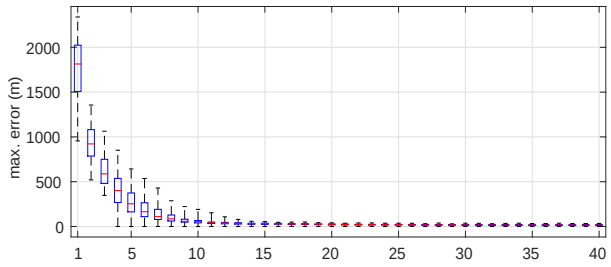

Fig. 3. Maximum localisation error in the network versus the iteration number of the proposed algorithm for 100 Monte Carlo runs. The boxes are centred at the median (red) with edges (blue) at the 25th and 75th percentiles.

We perform these computations for $t = 20$ time steps and evaluate the quad-term separable edge potential, i.e., $\psi^t_{i,j}(\theta^{(l)}_i, \theta^{(l)}_j)$, in (24) using (21). We use $S = 40$ iterations of the Monte Carlo realisation of the LBP algorithm detailed in Section IV-A.

First, we consider the accuracy of estimation in terms of the maximum localisation error in the network averaged over 100 Monte Carlo simulations. In Fig. 3, we present the box-plot of these errors for all runs with respect to the time window index n. The average error at the final iteration $n = 40$ is 15.8349m which is well within $3\sigma = 30$m bound of sensor measurement uncertainties.

In Fig. 4, we provide semi-log plot of the error margins over iterations after normalisation with $1000m$. Note that the decay speed is close to a log-linear regime and reaches a minimum corresponding to less than 2% of the normalising distance. These results demonstrate that the proposed scheme is capable of providing self-localisation with good accuracy and small error margins.

VII. Conclusions

In this work, we propose a node-wise separable likelihood for parameter estimation problems in state space models with multiple measurement sources. The separable structure enables us to evaluate this likelihood using local filtering operations thereby scaling with the number of sensors whereas the actual likelihood requires joint multi-sensor filtering which often has a prohibitive cost. Therefore, we use the proposed likelihood as an efficient approximation and relate the approximation quality to the reduction of uncertainty in predicting and estimating the underlying state when individual and pairs of sensor histories are taken into account. We show that the approximation quality improves as the local filtering accuracy gets better.

We use the proposed likelihood to build a pairwise Markov random field and exploit message passing algorithms for inference over this graphical structure. Doing that, we achieve an efficient computational structure for parameter estimation in state space models. We provide a conceptual description of the resulting algorithmic framework, as well as details of an implementation based on Monte Carlo methods for estimating calibration parameters in a fusion network. We demonstrate our approach in a sensor network self-localisation example.

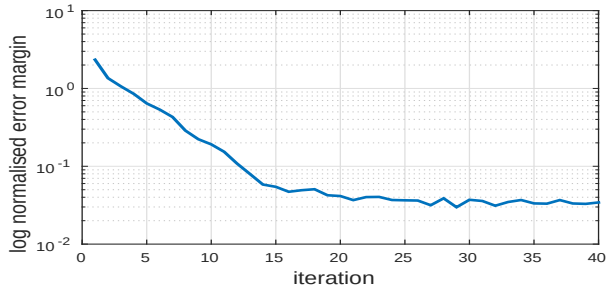

Fig. 4. Log-normalised error margin versus the iteration number n.

Appendix

A. Proof of Proposition 3.1

Proof. Let us expand the KLD term in (14) by substituting its arguments given in (4), (12) and (13):

$$
\begin{aligned}
&D\left(p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)||q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)\right) \\
&= \int \mathrm{d}z^i_k \mathrm{d}z^j_k \mathrm{d}z^i_{1:k-1} \mathrm{d}z^j_{1:k-1} \mathrm{d}\theta \, p(z^i_k, z^j_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta) \\
&\qquad \times \log \frac{p(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)}{q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)} \\
&= \int \mathrm{d}z^i_k \mathrm{d}z^j_k \mathrm{d}z^i_{1:k-1} \mathrm{d}z^j_{1:k-1} \mathrm{d}\theta \, p(z^i_k, z^j_k, z^i_{1:k-1}, z^j_{1:k-1}, \theta) \\
&\qquad \times \frac{1}{2} \left(\log \frac{p(z^i_k, z^j_k, z^i_{1:k-1}|z^j_{1:k-1}, \theta)}{p(z^i_k, z^j_k|z^j_{1:k-1}, \theta) p(z^i_{1:k-1}|z^j_{1:k-1}, \theta)} \right. \\
&\qquad \quad + \log \frac{p(z^i_k, z^j_k, z^j_{1:k-1}|z^i_{1:k-1}, \theta)}{p(z^i_k, z^j_k|z^i_{1:k-1}, \theta) p(z^j_{1:k-1}|z^i_{1:k-1}, \theta)} \\
&\qquad \quad \left. + 2 \log \kappa_k(\theta) \right) \qquad (42) \\
&= \frac{1}{2} \Big(I(Z^j_k, Z^i_k; Z^i_{1:k-1}|Z^j_{1:k-1}, \Theta) \\
&\qquad + I(Z^j_k, Z^i_k; Z^j_{1:k-1}|Z^i_{1:k-1}, \Theta) \Big) + E\{\log \kappa_k(\theta)\}. \qquad (43)
\end{aligned}
$$

In the equations above, $\kappa_k(\theta)$ is a normalisation constant given by (13). Eq.(42) is obtained after multiplying both the numerator and the denominator of the quotient inside the logarithm by $p(z^i_{1:k-1}|z^j_{1:k-1}, \theta) p(z^j_{1:k-1}|z^i_{1:k-1}, \theta)$ and a rearrangement of the terms. The definition of MI [23] results with the first two terms in Eq.(43). The last term is the expectation of the normalisation constant over the joint distribution of the sensor histories $Z^i_{1:k-1}$ and $Z^j_{1:k-1}$, and, Θ .

Let us now consider the normalisation constant:

$$
\begin{aligned}
\kappa_k(\theta) &= \int \mathrm{d}z^i_k \mathrm{d}z^j_k \left(p(z^i_k, z^j_k|z^j_{1:k-1}) p(z^i_k, z^j_k|z^i_{1:k-1}) \right)^{1/2} \\
&\leqslant \left(\int \mathrm{d}z^i_k \mathrm{d}z^j_k p(z^i_k, z^j_k|z^j_{1:k-1}) \right)^{1/2} \\
&\qquad \times \left(\int \mathrm{d}z^i_k \mathrm{d}z^j_k p(z^i_k, z^j_k|z^i_{1:k-1}) \right)^{1/2} \qquad (44) \\
&= 1.
\end{aligned}
$$

The inequality (44) follows from Hölder's Inequality. Consequently, the last term in (43) is non-positive, and, (14) is obtained. ■

B. Proof of Corollary 3.2

Proof. We apply the chain rule of information to the MI terms on the RHS of (14) leading to

$$I(Z_k^i, Z_k^j; Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta) = I(Z_k^j; Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta) + I(Z_k^i; Z_{1:k-1}^i | Z_k^j, Z_{1:k-1}^j, \Theta) \quad (45)$$

and,

$$I(Z_k^i, Z_k^j; Z_{1:k-1}^j | Z_{1:k-1}^i, \Theta) = I(Z_k^i; Z_{1:k-1}^j | Z_{1:k-1}^i, \Theta) + I(Z_k^j; Z_{1:k-1}^j | Z_k^i, Z_{1:k-1}^i, \Theta). \quad (46)$$

The MI terms on the RHSs of the equations above are for random variables which form Markov chains with the current the state variable X_k. Consider the (conditional) chains $Z_k^j \leftrightarrow X_k \leftrightarrow Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta$ and $Z_k^i \leftrightarrow X_k \leftrightarrow Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta$ for the RHS of Eq.(45). The Data Processing Inequality [23] applied to these terms lead to

$$\begin{aligned} &I(Z_k^i, Z_k^j; Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta) \\ &\leqslant I(X_k; Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta) + I(X_k^i; Z_{1:k-1}^i | Z_k^j, Z_{1:k-1}^j, \Theta) \\ &= H(X_k | Z_{1:k-1}^j, \Theta) - H(X_k | Z_{1:k-1}^i, Z_{1:k-1}^j, \Theta) \\ &+ H(X_k | Z_k^j, Z_{1:k-1}^j, \Theta) - H(X_k | Z_k^j, Z_{1:k-1}^j, Z_{1:k-1}^i, \Theta) \end{aligned} \quad (47)$$

A similar break down of Eq.(46) results with

$$\begin{aligned} &I(Z_k^i, Z_k^j; Z_{1:k-1}^j | Z_{1:k-1}^i, \Theta) \\ &\leqslant H(X_k | Z_{1:k-1}^i, \Theta) - H(X_k | Z_{1:k-1}^i, Z_{1:k-1}^j, \Theta) \\ &+ H(X_k | Z_k^i, Z_{1:k-1}^i, \Theta) - H(X_k | Z_k^i, Z_{1:k-1}^i, Z_{1:k-1}^j, \Theta). \end{aligned} \quad (48)$$

Substituting from (47) and (48) into (14) results with (15). ■

REFERENCES

[1] A. Zammit-Mangion, G. Sanguinetti, and V. Kadirkamanathan, "Variational estimation in spatiotemporal systems from continuous and point-process observations," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3449–3459, July 2012.

[2] Jeremie Houssineau, *Representation and estimation of stochastic populations*, Ph.D. thesis, Heriot Watt University, 2015.

[3] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics. Springer New York, 2006.

[4] Olivier Cappé, Simon J. Godsill, and Eric Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, pp. 899–924, 2007.

[5] Sumeetpal S. Singh, Nick Whiteley, and Simon J. Godsill, "Approximate likelihood estimation of static parameters in multi-target models," in *Bayesian Time Series Models*, David Barber, A. Taylan Cemgil, and Silvia Chiappa, Eds., pp. 225–244. Cambridge University Press, 2011, Cambridge Books Online.

[6] X. Lin, Y. Bar-Shalom, and T. Kirubarajan, "Multisensor multitarget bias estimation for general asynchronous sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 899–921, July 2005.

[7] N.K. Kantas, S.S. Singh, and A. Doucet, "Distributed maximum likelihood for simultaneous self-localization and tracking in sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5038–5047, 2012.

[8] M. Uney, B. Mulgrew, and D. Clark, "Target aided online sensor localisation in bearing only clusters," in *Sensor Signal Processing for Defence (SSPD) 2014*, Sept 2014, pp. 1–5.

[9] Dhanya Devarajan, Zhaolin Cheng, and R.J. Radke, "Calibrating distributed camera networks," *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1625–1639, 2008.

[10] Zhenhua Li, Siyue Chen, H. Leung, and E. Bosse, "Joint data association, registration, and fusion using em-kf," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 2, pp. 496–507, April 2010.

[11] B. Ristic, D.E. Clark, and N. Gordon, "Calibration of multi-target tracking algorithms using non-cooperative targets," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 390–398, 2013.

[12] Ronald Mahler, "The multisensor PHD filter: I. general solution via multitarget calculus," in *Proceedings of the SPIE*, 2009, vol. 7336, pp. 73360E–73360E–12.

[13] Lei Chen, Martin J. Wainwright, MüJdat ÇEtin, and Alan S. Willsky, "Data association based on optimization in graphical models with application to sensor networks," *Math. Comput. Model.*, vol. 43, no. 9-10, pp. 1114–1135, May 2006.

[14] A. Wiesel and A.O. Hero, "Distributed covariance estimation in gaussian graphical models," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 211–220, Jan 2012.

[15] Qiang Liu and Alexander Ihler, "Distributed parameter estimation via pseudo-likelihood," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, John Langford and Joelle Pineau, Eds., New York, NY, USA, July 2012, ICML '12, pp. 1487–1494, Omnipress.

[16] Yariv D. Mizrahi, Misha Denil, and Nando D. Freitas, "Distributed parameter estimation in probabilistic graphical models," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, Eds., pp. 1700–1708. Curran Associates, Inc., 2014.

[17] M. Uney, B. Mulgrew, and D. Clark, "Cooperative sensor localisation in distributed fusion networks by exploiting non-cooperative targets," in *IEEE Workshop on Statistical Signal Processing (SSP) 2014*, June 2014, pp. 516–519.

[18] M. Uney, B. Mulgrew, and D. Clark, "A cooperative approach to sensor localisation in distributed fusion networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1187–1199, March 2016.

[19] Martin J. Wainwright and Michael I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1-2, pp. 1–305, Jan. 2008.

[20] A.T. Ihler, J.W. Fisher, R.L. Moses, and A.S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 809–819, 2005.

[21] M. Cetin, Lei Chen, J.W. Fisher, A.T. Ihler, R.L. Moses, M.J. Wainwright, and A.S. Willsky, "Distributed fusion in sensor networks: A graphical models perspective," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, 2006.

[22] Mei Leng, Wee Peng Tay, T.Q.S. Quek, and Hyundong Shin, "Distributed local linear parameter estimation using gaussian spawn," *Signal Processing, IEEE Transactions on*, vol. 63, no. 1, pp. 244–257, Jan 2015.

[23] Thomas M Cover and Joy A Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[24] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss, "Understanding belief propagation and its generalizations," Tech. Rep. TR-2001-22, Mitsubishi Electric Research Laboratories, San Francisco, CA, USA, Jan. 2002.

[25] Erik B. Sudderth, Alexander T. Ihler, Michael Isard, William T. Freeman, and Alan S. Willsky, "Nonparametric belief propagation," *Commun. ACM*, vol. 53, no. 10, pp. 95–103, Oct. 2010.

[26] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein, "Particle markov chain monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, 2010.

[27] M. Uney, B. Mulgrew, and D. Clark, "Distributed estimation of latent parameters in state space models using separable likelihoods," in *Proceedings of the ICASSP 2016*, 2016.

[28] Nikolas Kantas, Arnaud Doucet, Sumeetpal S. Singh, Jan Maciejowski, and Nicolas Chopin, "On particle methods for parameter estimation in state-space models," *Statist. Sci.*, vol. 30, no. 3, pp. 328–351, 08 2015.

[29] Jie Gao and Leonidas Guibas, "Geometric algorithms for sensor networks," *Philosophical Transactions of the Royal Society A*, vol. 370, no. 1958, pp. 27–51, January 2012.

[30] M. Uney and M. Cetin, "Graphical model-based approaches to target tracking in sensor networks: An overview of some recent work and challenges," in *Image and Signal Processing and Analysis, 2007. ISPA 2007. 5th International Symposium on*, 2007, pp. 492–497.

[31] M.J. Wainwright, T.S. Jaakkola, and A.S. Willsky, "Tree-based reparameterization framework for analysis of sum-product and related algorithms," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1120–1146, May 2003.
[32] Jonathan S. Yedidia, W.T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2282–2312, July 2005.
[33] George Casella and Christian P. Robert, *Monte Carlo Statistical Methods*, Springer, second edition, 2005.
[34] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
[35] A. F. M. Smith and A. E. Gelfand, "Bayes statistics without tears: A sampling-resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, May 1992.
[36] Aubrey B. Poore and Sabino Gadaleta, "Some assignment problems arising from multiple target tracking," *Mathematical and Computer Modelling*, vol. 43, no. 9 - 10, pp. 1074 – 1091, 2006, Optimization and Control for Military ApplicationsOptimization and Control for Military Applications.
[37] Bradley P. Carlin and Thomas A. Louis, *Bayes and empirical Bayes methods for data analysis*, Chapman and Hall/CRC, 2nd edition, 2000.
[38] D. P. Bertsekas, "The auction algorithm: A distributed relaxation method for the assignment problem," *Annals of Operations Research*, vol. 14, no. 1, pp. 105–123, 1988.