DISTRIBUTED ESTIMATION OF LATENT PARAMETERS IN STATE SPACE MODELS USING SEPARABLE LIKELIHOODS

Murat Üney and Bernard Mulgrew

IDCOM, School of Engineering,
The University of Edinburgh,
EH9 3JL, Edinburgh, UK

Daniel Clark

School of Eng. & Physical Sciences,
Heriot-Watt University
EH14 4AS, Edinburgh, UK

ABSTRACT

Motivated by object tracking applications with networked sensors, we consider multi sensor state space models. Estimation of latent parameters in these models requires centralisation because the parameter likelihood depend on the measurement histories of all of the sensors. Consequently, joint processing of multiple histories pose difficulties in scaling with the number of sensors. We propose an approximation with a node-wise separable structure thereby removing the need for centralisation in likelihood computations. When leveraged with Markov random field models and message passing algorithms for inference, these likelihoods facilitate decentralised estimation in tracking networks as well as scalable computation schemes in centralised settings. We establish the connection between the approximation quality of the proposed separable likelihoods and the accuracy of state estimation based on individual sensor histories. We demonstrate this approach in a sensor network self-localisation example.

Index Terms— sensor networks, hidden Markov models, Markov random fields, pseudo-likelihood, simultaneous localisation and tracking

1. INTRODUCTION

State space models and their stochastic versions known as hidden Markov models [1] are used to represent a wide range of dynamic phenomena including spatio-temporal processes (see, e.g., [2]), individual dynamic systems and their populations [3]. As such, sensing applications, in which sensors dispersed over a region collect measurements from moving objects in their field of view, benefit from these models in solving problems involving the estimation of object trajectories.

These models often involve latent parameters [4] which need to be estimated based on measurements that the objects induce on the sensors. In fusion (or, object tracking) networks, localisation/calibration of sensors in a GPS denying environment using point detections of non-cooperative targets [5–7] can be treated as an instance of this problem setting. Another example is the estimation of the orientations and positions of nodes in a camera network based on feature detections [8].

The difficulty in multi-sensor settings is that the parameter likelihood requires all the target measurements collected across the net-

work to be filtered jointly. When a large number of sensors are involved, the complexity of joint filtering overwhelms likelihood evaluation both in maximum likelihood (ML) algorithms (e.g., [5]) and Bayesian estimation recursions (e.g., [6]). In this setting, even a quadratic complexity with the number of sensors might be prohibitive as is the case in linear Gaussian state space models [9, 10]. These difficulties are exacerbated in the case of more complex state space models, for example, set valued state variables [11], or, those with measurement association uncertainties [12]. Consequently, the joint filtering approach suffers from poor scalability with the number of sensors.

Decentralised paradigms, on the other hand, have more desirable properties such as scalability, better resource utilisation and flexibility. This is often achieved by filtering sensor histories individually thereby maintaining a linear $\mathcal{O}(N)$ cost for filtering. The filtered posteriors are then communicated and combined using fusion algorithms leading to a trade-off between the estimation accuracy and scalability. This perspective has parallels with the use of message passing schemes for distributed estimation [13–15]. The structure of the problem we consider, however, differs in that the parameter likelihood does not readily factorise into local functions implying a Markov model as in this line of work.

The need for centralised operation for evaluating a parameter likelihood also arises in the context of estimation of parametric distributions from samples. In this context, the use of surrogate local functions has been proposed [16–18], for efficient estimation. In our earlier work, we used a similar approach together with the perspective of local filtering and proposed separable likelihood functions for our problem setting [19, 20]. These likelihoods factorise into a pair of local terms and lead to a dual-term parameter likelihood approximation in state space models.

In this work, we propose an alternative separable likelihood for parameter estimation in distributed state space models. The separable structure allows its evaluation to be done using local filtering operations which in turn results with a linear filtering cost across the network. We assess the approximation quality using the Kullback-Leibler divergence (KLD) [21] of the proposed likelihood with respect to the actual likelihood obtained by joint filtering. In particular, we relate this divergence 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 leverage the node-wise separable likelihoods with pairwise Markov random fields (MRFs) and obtain a message passing scheme for parameter estimation in distributed multi-sensor state space models similar to, for example, [13–15].

The article is structured as follows: We provide the problem statement in Section 2. Then, we introduce the quad-term node-wise separable likelihood approximation in Section 3. In Section 4, we combine these likelihoods with pairwise MRFs. In section 5, we demonstrate the efficacy of our approach through a linear Gaussian state space model. Finally, we conclude in Section 6.

2. 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 $\{Z_t^i\}_{t=1}^k$ and $\{Z_t^j\}_{t=1}^k$ associated with sensors i and j, respectively, together with a Markov process $\{X_t\}_{t=1}^k$ specified by an initial state distribution and a transition density, all spanning time from $t = 1$ to k. The state space model with parameters θ is then specified as follows [22]: The state value x_t is a point in the state space $\mathcal{X}$ and is generated by the chain

$$X_t|(X_{1:t-1} = x_{1:t-1}) \sim \pi(x_t|x_{t-1};\theta), \quad X_1 \sim \pi_1(x_1;\theta),$$

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

$$Z_t^i|(X_{1:t} = x_{1:t}, Z_{1:k}^i = z_{1:t}^i) \sim g_i(z_t^i|x_t;\theta)$$
$$Z_t^j|(X_{1:t} = x_{1:t}, Z_{1:k}^j = z_{1:t}^j) \sim g_j(z_t^i|x_t;\theta)$$

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

Here, θ can capture a wide range of parameters such as noise model parameters and sensor locations. When θ is known, inference in this model reduces to the estimation of X_t by 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]:

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

$$p\left(z_{t+1}^i, z_{t+1}^j|z_{1:t}^i, z_{1:t}^j, \theta\right) = \int \left(g_i(z_{t+1}^i|x_{t+1}, \theta)g_j(z_{t+1}^j|x_{t+1}, \theta)\right) \times p(x_{t+1}|z_{1:t}^i, z_{1:t}^j, \theta)\mathrm{d}x_{t+1}, \quad (2)$$

where the first line follows from the chain rule of probabilities. The second line above is the contribution at time step $t + 1$ 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 $Z_{t+1}^j \perp\!\!\!\perp Z_{1:t}^j|X_{t+1}, \theta$ (see, e.g., [23]). The integrands in the expression for the likelihood update term in (2) are the multi-sensor likelihood and the prediction density for X_{t+1} based on the history of both sensors i and j. In other words, (2) is the scale factor for the posterior density of Bayesian recursions, or, the "centralised" filter. Therefore, the computation of the parameter likelihood (1) requires that a centre collects the measurements of both sensors and sequentially filters them jointly for $t = 1, \ldots, k-1$ to compute (2).

We are interested in estimating θ without resorting to joint filtering and with a computational structure that is well suited for distributed/decentralised operation such as message passings.

3. 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 the RHS of (2). This term factorises in alternative ways as follows:

$$p(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta)$$
$$= p(z_k^i|z_k^j, z_{1:k-1}^i, z_{1:k-1}^j, \theta)p(z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta) \quad (3)$$
$$= p(z_k^j|z_k^i, z_{1:k-1}^i, z_{1:k-1}^j, \theta)p(z_k^i|z_{1:k-1}^i, z_{1:k-1}^j, \theta) \quad (4)$$
$$= \left(p(z_k^i|z_k^j, z_{1:k-1}^i, z_{1:k-1}^j, \theta)p(z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta)\right)^{1/2}$$
$$\times \left(p(z_k^j|z_k^i, z_{1:k-1}^i, z_{1:k-1}^j, \theta)p(z_k^i|z_{1:k-1}^i, z_{1:k-1}^j, \theta)\right)^{1/2} \quad (5)$$

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.(5) 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 (5), i.e.,

$$q(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta) \triangleq$$
$$= \frac{1}{\kappa_k(\theta)}\left(p(z_k^i|z_k^j, z_{1:k-1}^j, \theta)p(z_k^j|z_{1:k-1}^j, \theta)\right)^{1/2}$$
$$\times \left(p(z_k^j|z_k^i, z_{1:k-1}^i, \theta)p(z_k^i|z_{1:k-1}^i, \theta)\right)^{1/2} \quad (6)$$
$$\kappa_k(\theta) = \iint \mathrm{d}z_k^i \mathrm{d}z_k^j \left(p(z_k^i, z_k^j|z_{1:k-1}^j, \theta)p(z_k^i, z_k^j|z_{1:k-1}^i, \theta)\right)^{1/2} \quad (7)$$

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. Next, we consider the difference between the update term (2) and the quad-term approximation introduced in (6). Because these terms are probability densities over the sensor measurements, their "divergence" can be quantified using the KLD [21]. 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* (6) *is bounded by the average of the Mutual Information (MI) [21] between the current measurement pair and a single sensor's history conditioned on the history of the other sensor, i.e.,*

$$D(p(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta)||q(z_k^i, z_k^j|z_{1:k-1}^i, z_{1:k-1}^j, \theta)) \leq$$
$$\frac{1}{2}I(Z_k^i, Z_k^j; Z_{1:k-1}^i|Z_{1:k-1}^j, \Theta) + \frac{1}{2}I(Z_k^i, Z_k^j; Z_{1:k-1}^j|Z_{1:k-1}^i, \Theta). \quad (8)$$

Proof. We begin the proof by substituting the distributions of concern in the definition of (conditional) KL-divergence in (8) and obtain (9) and (10) where $\kappa_k(\theta)$ is the normalisation constant for q defined in (7). Eq.(9) follows from the multiplication of both the numerator and the denominator of the quotient inside the logarithm by $p(z_{1:k-1}^i|z_{1:k-1}^j, \theta)p(z_{1:k-1}^j|z_{1:k-1}^i, \theta)$ and a rearrangement of the terms. The definition of MI results with the first two terms in Eq.(10). The last term is the expectation of the normalisation constant over the joint distribution over the sensor histories

$$D(p(z_k^i, z_k^j | z_{1:k-1}^i, z_{1:k-1}^j, \theta) || q(z_k^i, z_k^j | z_{1:k-1}^i, z_{1:k-1}^j, \theta))$$

$$= \int \mathrm{d}z_k^i \mathrm{d}z_k^j \mathrm{d}z_{1:k-1}^i \mathrm{d}z_{1:k-1}^j \mathrm{d}\theta \, p(z_k^i, z_k^j, z_{1:k-1}^i, z_{1:k-1}^j, \theta) \log \frac{p(z_k^i, z_k^j | z_{1:k-1}^i, z_{1:k-1}^j, \theta)}{q(z_k^i, z_k^j | z_{1:k-1}^i, z_{1:k-1}^j, \theta)}$$

$$= \int \mathrm{d}z_k^i \mathrm{d}z_k^j \mathrm{d}z_{1:k-1}^i \mathrm{d}z_{1:k-1}^j \mathrm{d}\theta \, p(z_k^i, z_k^j, z_{1:k-1}^i, z_{1:k-1}^j, \theta) \log \left(\frac{p(z_k^i, z_k^j | z_{1:k-1}^i, z_{1:k-1}^j, \theta) \kappa_k(\theta)}{\left(p(z_k^i | z_k^j, z_{1:k-1}^j, \theta) p(z_k^j | z_{1:k-1}^j, \theta) p(z_k^j | z_k^i, z_{1:k-1}^i, \theta) p(z_k^i | z_{1:k-1}^i, \theta) \right)^{1/2}} \right)$$

$$= \int \mathrm{d}z_k^i \mathrm{d}z_k^j \mathrm{d}z_{1:k-1}^i \mathrm{d}z_{1:k-1}^j \mathrm{d}\theta \, p(z_k^i, z_k^j, z_{1:k-1}^i, z_{1:k-1}^j, \theta) \frac{1}{2} \left(\log \frac{p(z_k^i, z_k^j, z_{1:k-1}^i | z_{1:k-1}^j, \theta)}{p(z_k^i, z_k^j | z_{1:k-1}^j, \theta) p(z_{1:k-1}^i | z_{1:k-1}^j, \theta)} \right.$$

$$\left. + \log \frac{p(z_k^i, z_k^j, z_{1:k-1}^j | z_{1:k-1}^i, \theta)}{p(z_k^i, z_k^j | z_{1:k-1}^i, \theta) p(z_{1:k-1}^j | z_{1:k-1}^i, \theta)} + 2 \log \kappa_k(\theta) \right) \tag{9}$$

$$= \frac{1}{2} \left(I(Z_k^j, Z_k^i; Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta) + I(Z_k^j, Z_k^i; Z_{1:k-1}^j | Z_{1:k-1}^i, \Theta) \right) + E\{\log \kappa_k(\theta)\}. \tag{10}$$

$Z_{1:k-1}^i$ and $Z_{1:k-1}^j$, and, Θ. Consider the normalisation constant:

$$\kappa_k(\theta) = \int \mathrm{d}z_k^i \mathrm{d}z_k^j \left(p(z_k^i, z_k^j | z_{1:k-1}^j) p(z_k^i, z_k^j | z_{1:k-1}^i) \right)^{1/2}$$

$$\leq \left(\int \mathrm{d}z_k^i \mathrm{d}z_k^j p(z_k^i, z_k^j | z_{1:k-1}^j) \right)^{1/2}$$

$$\times \left(\int \mathrm{d}z_k^i \mathrm{d}z_k^j p(z_k^i, z_k^j | z_{1:k-1}^i) \right)^{1/2} \tag{11}$$

$$= 1.$$

The inequality (11) follows from Hölder's Inequality. Consequently, the last term in (10) is non-positive, and we obtain Eq.(8). ∎

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., $(Z_k^i, Z_k^j) \perp\!\!\!\perp Z_{1:k-1}^j | X_k, \Theta$ holds and consequently

$$I(Z_k^i, Z_k^j; Z_{1:k-1}^i | X_k, \Theta) = I(Z_k^i, Z_k^j; Z_{1:k-1}^j | X_k, \Theta) = 0.$$

Similarly, the average MI term on the right hand side of (8) is zero if $(Z_k^i, Z_k^j) \perp\!\!\!\perp Z_{1:k-1}^i | Z_{1:k-1}^j, \Theta$ and $(Z_k^i, Z_k^j) \perp\!\!\!\perp Z_{1:k-1}^j | Z_{1:k-1}^i, \Theta$ hold simultaneously. This condition is satisfied, for example, in the case that either of the measurement histories $Z_{1:k-1}^i$ and $Z_{1:k-1}^j$ 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 (8) 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:*

$$D(p||q) \leq \frac{1}{2} \left(\left(H(X_k | Z_{1:k-1}^j, \Theta) - H(X_k | Z_{1:k-1}^j, Z_{1:k-1}^i, \Theta) \right) \right.$$

$$+ \left(H(X_k | Z_{k-1}^i, \Theta) - H(X_k | Z_{1:k-1}^j, Z_{1:k-1}^i, \Theta) \right) \bigg)$$

$$+ \frac{1}{2} \left(\left(H(X_k | Z_{1:k}^j, \Theta) - H(X_k | Z_{1:k}^j, Z_{1:k-1}^i, \Theta) \right) \right.$$

$$+ \left(H(X_k | Z_{1:k}^i, \Theta) - H(X_k | Z_{1:k}^i, Z_{1:k-1}^j, \Theta) \right) \bigg), \tag{12}$$

where H denotes the Shannon Entropy [21].

Proof. We apply the chain rule of information to the MI terms on the RHS of (8) leading to to

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

$$+ I(Z_k^i; Z_{1:k-1}^i | Z_k^j, Z_{1:k-1}^j, \Theta) \tag{13}$$

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

$$+ I(Z_k^j; Z_{1:k-1}^j | Z_k^i, Z_{1:k-1}^i, \Theta) \tag{14}$$

The MI terms on the RHSs of the equations above are for random variables which form Markov chains with the current target state 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 $Z_k^i \leftrightarrow X_k \leftrightarrow Z_{1:k-1}^i | Z_k^j, Z_{1:k-1}^j, \Theta$ for the RHS of Eq.(13). The Data Processing Inequality [21] applied to these terms lead to

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

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

$$= H(X_k | Z_{1:k-1}^j, \Theta) - H(X_k | Z_{1:k-1}^j, Z_{1:k-1}^i, \Theta)$$

$$+ H(X_k | Z_k^j, Z_{1:k-1}^j, \Theta) - H(X_k | Z_k^j, Z_{1:k-1}^j, Z_{1:k-1}^i, \Theta) \tag{15}$$

A similar break down of Eq.(14) results with

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

$$\leq H(X_k | Z_{1:k-1}^i, \Theta) - H(X_k | Z_{1:k-1}^i, Z_{1:k-1}^j, \Theta)$$

$$+ H(X_k | Z_k^i, Z_{1:k-1}^i, \Theta) - H(X_k | Z_k^i, Z_{1:k-1}^i, Z_{1:k-1}^j, \Theta) \tag{16}$$

Substituting from (15) and (16) into (8) results with (12). ∎

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 (12) 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 (8) for the quad-term approximation is smaller than the exact KLD of the aforementioned dual term approximation [20]. 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 [20]. Evaluation of the scaling factor in (7) might not be straightforward for general problem settings.

A node-wise separable likelihood built upon the quad-term update is obtained after substituting q in (6) as the update term in (1):

$$\tilde{l}\left(z_{1:k}^i, z_{1:k}^j|\theta\right) = \prod_{t=0}^{k-1} \frac{1}{\kappa_t(\theta)} \left(p(z_t^i|z_t^j, z_{1:t-1}^j, \theta)p(z_t^j|z_{1:t-1}^j, \theta)\right)^{1/2}$$
$$\times \left(p(z_t^j|z_t^i, z_{1:t-1}^i, \theta)p(z_t^i|z_{1:t-1}^i, \theta)\right)^{1/2}$$
$$= \frac{1}{\mathcal{K}^k(\theta)} \left(r_{ij}^k(\theta)s_{ij}^k(\theta)\right)^{1/2} \left(r_{ji}^k(\theta)s_{ji}^k(\theta)\right)^{1/2} \quad (17)$$

where the four terms $r_{ij}^k, s_{ij}^k, r_{ji}^k$, and s_{ji}^k have recursive forms as

$$r_{ij}^k(\theta) \triangleq r_{ij}^{k-1}(\theta)p(z_k^i|z_k^j, z_{1:k-1}^j, \theta), \quad (18)$$

$$s_{ij}^k(\theta) \triangleq s_{ij}^{k-1}(\theta)p(z_k^i|z_{1:k-1}^i, \theta), \quad (19)$$

$$r_{ji}^k(\theta) \triangleq r_{ji}^{k-1}(\theta)p(z_k^j|z_k^i, z_{1:k-1}^i, \theta), \quad (20)$$

$$s_{ji}^k(\theta) \triangleq s_{ji}^{k-1}(\theta)p(z_k^j|z_{1:k-1}^j, \theta). \quad (21)$$

The normalisation constant $\mathcal{K}^k$ also has a similar recursive form with update terms found using (7).

This approximate likelihood can be evaluated in a distributed fashion in which sensor nodes filter their own measurement histories and exchange these posteriors with the other node at every time step t. Node i, based on its local measurements and the posterior from j, computes r_{ij} and s_{ij} above for each k. Node j performs a similar operation, and, in this respect, subscript ij indicates that the associated term is computed at sensor i and will be transmitted to sensor j and vice versa. The product of the local terms and the received terms yield the quad-term node-wise separable likelihood after normalisation with (7).

4. PAIRWISE MARKOV RANDOM FIELD MODELS FOR DISTRIBUTED ESTIMATION IN STATE SPACE MODELS

The benefits of the quad-term node-wide 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 $\mathcal{O}(N)$ regardless of the number of quad-terms to be evaluated. Of course, selecting all possible $\binom{N}{2}$ terms to be included might undermine the benefits obtained in terms of filtering. We are interested in cases in which the selected pairs render a connected planar graph thereby yielding the number of pairs less than $\mathcal{O}(N^2)$. Then, 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 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
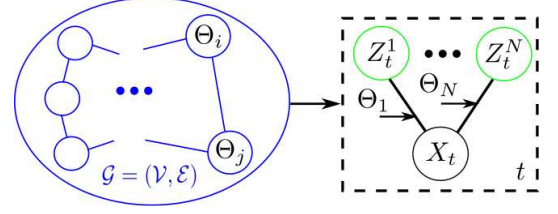

Fig. 1. A multi-sensor state space - or, hidden Markov- model (black dashed box on the right representing a chain over t) parameterised with a Markov Random field (the blue edges on the left).

posterior is a pairwise Markov random field over $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [23]:

$$p(\theta|Z_{1:k}^1, ..., Z_{1:k}^N) \propto \prod_{i\in\mathcal{V}} \psi_i(\theta_i) \prod_{(i,j)\in\mathcal{E}} \psi_{ij}^k(\theta_i, \theta_j), \quad (22)$$
$$\psi_i(\theta_i) = p_{0,i}(\theta_i),$$
$$\psi_{ij}^k(\theta_i, \theta_j) = \tilde{l}(z_{1:k}^i, z_{1:k}^j|\theta_i, \theta_j),$$

where the node potential functions (i.e., ψ_is) are arbitrary priors for θ_i (e.g., uniform distributions over $\mathcal{B}$) and the edge potentials (i.e., ψ_{ij}^ks) are the quad-term separable likelihoods for the pairs (i, j)s based on sensor histories up to time k.

This model is illustrated in Fig. 1 and enables the estimation of θ in a distributed fashion: The pairwise MRF model in (22) 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

$$m_{ji}(\theta_i) = \int \psi_{ij}^k(\theta_i, \theta_j)\psi_j(\theta_j) \prod_{i'\in ne(j)\setminus i} m_{i'j}(\theta_j)\,\mathrm{d}\theta_j, \quad (23)$$

$$\tilde{p}_i(\theta_i) \propto \psi_i(\theta_i) \prod_{j\in ne(i)} m_{ji}(\theta_i). \quad (24)$$

In BP iterations, nodes simultaneously send messages to their neighbours using (23) (often starting with constants as the previously received messages) and update their local "belief" using (24). 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 (22), in a finite number of steps [24]. For the case in which $\mathcal{G}$ contains cycles, iterations of (23) and (24) are known as loopy BP (LBP) and has been very successful in computing approximate marginals in a distributed fashion, in fusion, self-localisation and tracking problems in sensor networks [13, 14, 25].

In the resulting algorithm, we consider the pairwise MRF model in (22) equipped with quad-term node-wise separable likelihoods. The sensor histories are filtered individually and the filtered posteriors are exchanged with the neighbouring sensors which are then used to compute the update terms in (18)–(21). This yields a cost of filtering linear in the number of sensors. At the end of k steps, the edge potentials in (22) are computed using (17). Finally, the marginal parameter posteriors are found using LBP message passing operations specified by (23) and (24).
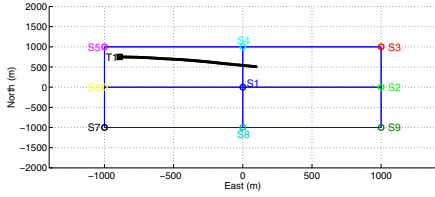
Fig. 2. Example scenario: Sensors (S1-S9) collect measurements from an object (T1) with initial position denoted by the black square. The blue lines depict the MRF model for the estimation task.

5. EXAMPLE: SELF-LOCALISATION IN LINEAR GAUSSIAN STATE SPACE MODELS

In this example, we consider a linear Gaussian state space model with unknown sensor locations θ_is given by

$$\pi(x_t|x_{t-1}) = \mathcal{N}(x_t; \mathbf{F}x_{t-1}, \mathbf{Q}) \tag{25}$$

$$g_i(z_t^i|x_t; \theta_i) = \mathcal{N}(z_t^i; \mathbf{H}_i(x_t - \theta_i), \mathbf{R}_i) \tag{26}$$

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. $\mathbf{H}_i$ is the observation matrix and $\mathbf{R}_i$ is the observation noise, both associated with sensor i.

The linear/Gaussian nature of the model renders Kalman filter as the optimal Bayesian filter. For each sensor j, we use a Kalman filter with the history $z_{1:k}^j$ which yields the following Gaussian prediction and posterior distributions at time k

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

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

After multiplying these densities with the measurement likelihood of sensor i and marginalising out the state variable, we obtain the update terms in (18) and (19) as follows:

$$p(z_k^i|z_k^j, z_{1:k-1}^j, \theta_{i,j}) = \mathcal{N}(z_k^i; \mathbf{H}_i(\hat{\mathbf{x}}_{k|k}^j + \theta_j - \theta_i), \mathbf{S}_k) \tag{27}$$

$$\mathbf{S}_k = \mathbf{R}_i + \mathbf{H}_i \mathbf{P}_{k|k}^j \mathbf{H}_i^T$$

$$p(z_k^i|z_{1:k-1}^j, \theta_{i,j}) = \mathcal{N}(z_k^i; \mathbf{H}_i(\hat{\mathbf{x}}_{k|k-1}^j + \theta_j - \theta_i), \mathbf{S}_{k|k-1}) \tag{28}$$

$$\mathbf{S}_{k|k-1} = \mathbf{R}_i + \mathbf{H}_i \mathbf{P}_{k|k-1}^j \mathbf{H}_i^T$$

The four terms in (18)–(21) as well as the scaling factor $\kappa_k(\theta_{i,j})$ in (7) can be found in closed form, the latter using integration rules of quadratic exponentials.

We consider the scenario depicted in Fig. 2. Nine sensors observe a moving object. We select $\mathbf{F}$ and $\mathbf{Q}$ 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 perform the edge potential computations using (27)-(28) in (18)–(21) and (17). These evaluations take place within LBP message passing iterations given by (23)-(24). In particular, we consider a particle based implementation strategy known as non-parametric BP [26]. The iterations are initiated with a set of samples $\{\theta_i^{(l)}\}_{l=1}^L$ for each sensor i generated from the local position prior $p_{0,i}(\theta_i)$. Non-parametric BP approach represents the message from node j to i by a sample set $\{\bar{\theta}_i^{(l)}\}_{k=1}^L$ associated with the weights given by

$$\bar{\omega}_i^{(l)} \propto \psi_{ij}^k(\bar{\theta}_i^{(l)}, \theta_j^{(l)}), \quad \theta_j^{(l)} \sim \left(\psi_j(\theta_j) \prod_{i' \in ne(j)\setminus i} m_{i'j}(\theta_j) \right)$$
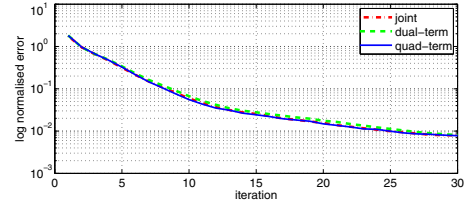

Fig. 3. Log normalised average error versus LBP iterations for the joint, dual-term and the quad-term separable likelihoods.

where the node and edge potentials have been selected as the prior densities and quad-term separable likelihoods in (22).

We use 30 iterations of LBP on the pairwise graph $\mathcal{G}$ in Fig. 2. We compare the performance of the proposed quad-term likelihoods with the actual edge potentials found by joint filtering as well as the dual-term approximation in [20] using 200 Monte Carlo simulations. First, we consider the accuracy of estimation in terms of the average maximum localisation error in the network. In Fig. 3, we provide log-plots of this term over iterations after normalisation with $1000m$. All three edge potentials result with a maximum of less than $10m$ error on the average by iteration 30. Note that the quad-term approximation performs on par with the baseline joint filtering term, both of which perform slightly better than the dual-term approximation. The computation time for both of the separable edge potentials are similar and average to $12.2s$ per edge per iteration whereas the joint filtering results with an average of $28.3s$. Therefore, the proposed approximation provides significant savings in computational time for a negligible loss in estimation accuracy, in this example.

6. CONCLUSIONS

We propose a node-wise separable likelihood for local parameter estimation in distributed state space models. 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. We provide closed form expressions and bounds for the KLD of the proposed approximation with respect to the actual likelihood and show that the approximation quality improves as the local filtering accuracy gets better. We leverage the node-wise separable likelihoods with pairwise Markov random fields and obtain a message passing scheme for parameter estimation in distributed state space models. We demonstrate the resulting distributed estimation scheme on a sensor self-localisation example.

7. REFERENCES

[1] R.J. Elliott, L. Aggoun, and J.B. Moore, *Hidden Markov Models: Estimation and Control*, Stochastic Modelling and Applied Probability. Springer New York, 2008.

[2] 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.

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

[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] 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.

[6] 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.

[7] 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.

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

[9] B.S. Rao and H.F. Durrant-Whyte, "Fully decentralised algorithm for multisensor kalman filtering," *Control Theory and Applications, IEE Proceedings D*, vol. 138, no. 5, pp. 413–420, Sep 1991.

[10] Hugh Durrant-Whyte, "Multi-sensor data fusion," ACFR, Univ. of Sydney, January 2001, course notes.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] M. Uney, B. Mulgrew, and D. Clark, "A cooperative approach to sensor localisation in distributed fusion networks," *IEEE Transactions on Signal Processing*, 2015, accepted for publication.

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

[22] 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.

[23] 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.

[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] 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.

[26] 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.