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 when evaluating the parameter likelihood. In this article, we propose a separable pseudo-likelihood which is a more accurate approximation compared to a previously proposed alternative under typical operating conditions. In addition, we consider using separable likelihoods in the presence of many objects and ambiguity in associating measurements with objects that originated them. To this end, we use a state space model with a hypothesis based parameterisation, and, develop an empirical Bayesian perspective in order to evaluate separable likelihoods on this model using local filtering. Bayesian inference with this likelihood is carried out using belief propagation on the associated pairwise Markov random field. We specify a particle algorithm for latent parameter estimation in a linear Gaussian state space model and demonstrate its efficacy for network self-calibration using measurements from non-cooperative targets in comparison with alternatives.

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 area surveillance is underpinned by state space models which are capable of representing a variety of dynamic phenomena such as spatio-temporal (see, e.g., [1]) processes. In fusion (or, object tracking [2]) 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], [7] and localisation/calibration of sensors in a GPS denying environment (e.g., in underwater sensing [8])

using point detections from non-cooperative targets [9], [10]. Another example is the estimation of the orientations and positions of nodes in a camera network based on feature detections [11].

Such problems fall in the domain of parameter estimation in state space models (see, e.g., [12] for a review). The parameter likelihood of the multi-sensor problem, however, does not scale well with the number of sensors which specifies the dimensionality of the unknown, or, the length of the measurement window that will be used for estimation. In the presence of multiple objects, the scalability issue is exacerbated by the measurement origin (or, data association) uncertainties that arise. Exact evaluation of the likelihood in this case has combinatorial complexity with the number of sensors [13], and, in general multiple object models, it is intractable even for a single sensor [14]. Estimation using a maximum likelihood (ML) or a Bayesian approach requires repeated evaluation of this likelihood (see, e.g., [12], [15], [16]) necessitating the use of efficient approximation strategies.

Intractable or computationally prohibitive likelihoods have motivated a number of lines of work in the statistics literature including likelihood free methods, or, approximate Bayesian computation [17], and, composite likelihood/pseudo-likelihood approaches [18]. Likelihood free methods can be used for sampling from the parameter posterior in state space models [19] including those capable of modelling multiple objects [20]. The latter approach is based on developing surrogates to replace the original likelihood, e.g., block based approximations in maximum likelihood [21]. The pseudo-likelihood perspective has been useful in networked settings in which constraints on i) the availability of parts of data, and/or, ii) scalability in processing with the number of sources arise. Examples include surrogates built upon local functions for estimation of parametric probability measures (e.g., exponential family distributions) from distributedly stored high dimensional samples [22]–[24].

It is not straightforward to find such pseudo-likelihoods for parameter estimation in state space models, however, that can resolve these two issues that arise when there are multiple data sources (or, sensors). It is worthwhile to develop and analyse surrogates that provide scalability with the number of sources, and, are suitable to local computations (e.g., local filtering). In [25], we proposed a pseudo-likelihood which is a product of "dual-term" approximations replacing their intractable exact counterparts. These approximations are separable in that they can be evaluated using single sensor filtering. This feature underpins scalability with the number of sensors. In [26], we have investigated the quality of the dual-term approximation,

and, related it to the level of uncertainty in the prediction and estimation of the underlying state process.

In this work, we propose an alternative pseudo-likelihood which is provably a more accurate approximation for parameter estimation in multi-sensor state space models, under typical operating conditions. This approximation is also separable in that it is a scaled product of quadruple terms each of which can be found using single sensor filtering. In order to exploit this quad-term likelihood when there are multiple objects, extra attention should be paid to the handling of the data association uncertainties. We propose to use a hypothesis based parameterisation for the multi-object state space model as detailed in [14]& [27] in order to facilitate the use of the quad-term surrogate in this setting. In the parameterised model, we explicitly point out the combinatorial complexity of exact likelihood evaluation with the number of sensors. Then, we introduce an empirical Bayesian [28] interpretation of local filtering that facilitates the use of separable likelihoods within this model. These modelling aspects detailing the use of separable likelihoods in hypothesis based multi-object models constitute the second contribution of this work.

Separable likelihoods fit well in distributed fusion architectures in which locally filtered distributions are transmitted in the network, as opposed to sensor measurements [29]. Moreover, they facilitate parameter estimation using a message passing computational structure which is desirable in networked problems. Specifically, the proposed likelihood surrogate together with independent parameter priors lead to a pairwise Markov random field (MRF) posterior model. 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) [30] which consists of successive message passings among neighbouring nodes and updating of local marginals based on these messages. This computational structure lends itself to decentralised estimation, as well as scalable computation at fusion centre.

As an indication of the approximation quality, we consider the Kullback-Leibler divergence (KLD) [31] of the quad-term likelihood with respect to the actual pairwise likelihood 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 approximation.

We provide a Monte Carlo algorithm for sensor self-calibration in this framework for linear Gaussian state space (LGSS) models. The algorithm is based on the nonparametric BP approach [32] and involves sampling from the updated marginals followed by quad-term 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 are evaluated 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 (MCMC) algorithms [15]. As such, we differ from [26] in which windowing of measurements are used for enabling online processing.

Preliminary results of the proposed pseudo-likelihood can be found in [33]. This article provides a complete account of our solution strategy in multiple object models and is structured as follows: Section II provides the probabilistic model and the problem statement. Then, we detail pairwise pseudo-likelihoods in parameterised multi-object models, and, relate this perspective to latent parameter estimation via inference over pairwise MRFs, in Section III. The proposed quad-term node-wise separable likelihood approximation is detailed in Section IV. Section V details the structural and computational properties of the quad-term approximation when the unknowns are respective quantities. Based on these results, we propose a distributed sensor localisation algorithm in linear Gaussian multi-object state space models in Section VI. The efficacy of this algorithm is demonstrated in comparison to the approach in [26], in Section VII. Finally, we conclude in Section VIII.

II. PROBLEM DEFINITION

A. Probabilistic model

Let us consider a set of sensors $\mathcal{V} = \{1, \ldots, N\}$ networked over communication links listed by $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is undirected (i.e., the links are bi-directional), connected, and, might contain cycles.

Next, let us consider a single object with state evolution modelled as a Markov process X_k for time index $k \geq 1$. This process is specified by an initial state distribution and a transition density. The state space model with parameters θ is then specified as follows [12]: The state value x_k is a point in the state space $\mathfrak{X}$ and is generated by the chain

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

where $.|.$ denotes conditioning. A measured value $z_k^i \in \mathfrak{Z}^i$ at sensor $i \in \mathcal{V}$ is generated independently in accordance with the likelihood model

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

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

In fusion scenarios, there are multiple such objects denoted by a multi-object state

$$\mathcal{X}_k \triangleq [X_{k,1}, \ldots, X_{k,M_k}], \tag{3}$$

that induce measurements according to the above state space model resulting with sensors collecting a multitude of measurements

$$\mathcal{Z}_k^i \triangleq [Z_{k,1}^i, \ldots, Z_{k,O_k^i}^i], \tag{4}$$

where M_k is the number of objects and O_k^i is the number of measurements collected at sensor i at time k. Here, the origin of $Z_{k,j}^i$s are unknown, i.e., the *data association*s which encode a mapping from these measurement (random) variables to the elements of $\mathcal{X}_k$ (and, equivalently to the previously collected measurements from the same objects) are not known [2].

In the general multi-object tracking model, M_k and O_k^i are random variables with laws determined by probability models regarding how these objects appear in the surveillance region and disappear (which is often referred to as their birth and death, respectively), the law for the false alarms, etc. For the sake of simplicity and ease of presentation in the limited space especially when relating computational complexity to the number of sensors and objects in the following discussion, we assume that all of the objects that exist at time step $k = 1$ remain in the scene for the time window considered and there are no missed detections and false alarms in sensor measurements which imply that $M_k = M$ and $O_k^i = M_k$, respectively, for some positive integer M with probability one.

In this simplified "closed world" model the multi-object state transition is given by

$$\pi(\mathbf{X}_k|\mathbf{X}_{k-1};\theta) = \prod_{m=1}^{M} \pi(x_{k,m}|x_{k-1,m};\theta). \qquad (5)$$

The likelihood of the measurements collected by sensor i is conditioned not only on the multi-object state $\mathbf{X}_k$, but, also on a (data association) hypothesis τ_k^i that encodes the association of measurements to the objects within $\mathbf{X}_k$:

$$l_i(\mathbf{Z}_k^i|\mathbf{X}_k,\tau_k^i;\theta) = \prod_{o=1}^{M} g_i(z_{k,o}^i|x_{k,\tau_k^i(o)};\theta), \qquad (6)$$

and, the prior on τ_k^i assigns equal probability to all $M!$ permutations of $[1,\ldots,M]$ that τ_k^i can take, i.e.,

$$p(\tau_k^i) = \frac{1}{M!}. \qquad (7)$$

B. Statement of the problem

We are interested in estimating $\theta \in \mathcal{B}$ using the measurements collected across the network by sensors $i \in \mathcal{V}$ for a time window of length t. The parameter likelihood of the problem quantifies how well these measurements fit into the state space model with the selected value of the parameter, and, is evaluated via multi-sensor filtering [4, Sec.IV]:

$$l\left(\mathbf{Z}_{1:t}^1,\ldots,\mathbf{Z}_{1:t}^N|\theta\right) =$$
$$\prod_{k=1}^{t} p\left(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N,\theta\right), \quad (8)$$

where the time updates on the right hand side are given by

$$p\left(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N,\theta\right) =$$
$$\sum_{\tau_k^1}\cdots\sum_{\tau_k^N} p(\tau_k^1) \times \ldots \times p(\tau_k^N)$$
$$\times \int_{\mathcal{X}^M} l(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{X}_k,\tau_k^1,\ldots,\tau_k^N;\theta)$$
$$\times p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N;\theta)\mathrm{d}\mathbf{X}_k, \quad (9)$$

and, the multi-sensor likelihood inside the integration factorises as

$$l(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{X}_k,\tau_k^1,\ldots,\tau_k^N;\theta) = \prod_{i\in\mathcal{V}} l_i(\mathbf{Z}_k^i|\mathbf{X}_k,\tau_k^i;\theta). \qquad (10)$$

where the terms in the product are given by (6).

Here, (8) follows from the chain rule of probabilities. The term in (9) 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 mutually independent of the measurement histories, conditioned on the current state for any value of θ. Let us denote this relation by $\mathcal{Z}_k^j \perp\!\!\!\perp \mathcal{Z}_{1:k-1}^j|\mathcal{X}_k,\theta$ for $i \in \mathcal{V}$ (see, e.g., [34], for this notation). (10) follows from that the measurements of different sensors are mutually independent, i.e., $\mathcal{Z}_k^i \perp\!\!\!\perp \mathcal{Z}_k^j|\mathcal{X}_k,\theta$ for $(i,j) \in \mathcal{V} \times \mathcal{V}$.

This likelihood can be used in a MMSE estimator of $\theta \in \mathcal{B}$, in principle, for a random variable Θ associated with a prior density $p(\theta)$. This estimate is given by the expected value of the posterior distribution

$$p(\theta|\mathbf{Z}_{1:t}^1,\ldots,\mathbf{Z}_{1:t}^N) \propto l(\mathbf{Z}_{1:t}^1,\ldots,\mathbf{Z}_{1:t}^N|\theta)\,p(\theta), \qquad (11)$$
$$\hat{\theta} = \int_{\mathcal{B}} \theta\, p(\theta|\mathbf{Z}_{1:t}^1,\ldots,\mathbf{Z}_{1:t}^N)\,\mathrm{d}\theta. \qquad (12)$$

The MMSE estimate can be computed by generating L samples from the posterior distribution in (11) using, for example, MCMC methods [15] and using these samples to find a Monte Carlo estimate of the integral in (12). In both this approach and maximum likelihood (ML) solutions aiming to maximise (8) with iterative optimisation, repeated evaluations of the likelihood are required.

The evaluation of this likelihood is intractable, however, not only because of the $(M!)^N$ summations in (9), but, also because of the complexity in finding the integrations involved. The integrands here are i) the multi-sensor likelihood in (10), and, ii) the prediction density for $\mathbf{X}_k$ based on the network's entire measurement history up to time k. In other words, (9) is the scale factor for the posterior density of Bayesian recursions, or, the "centralised" filter given by

$$p(\mathbf{X}_k,\tau_k^{1:N}|\mathbf{Z}_{1:k}^1,\ldots,\mathbf{Z}_{1:k}^N;\theta) =$$
$$\frac{l(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{X}_k,\tau_k^{1:N};\theta)}{p\left(\mathbf{Z}_k^1,\ldots,\mathbf{Z}_k^N|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N\right)} \times p(\tau_k^{1:N})$$
$$\times p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^j,\theta), \qquad (13)$$
$$p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N,\theta) =$$
$$\sum_{\tau_{k-1}^1}\cdots\sum_{\tau_{k-1}^N} \int_{\mathcal{X}^M} \pi(\mathbf{X}_k|\mathbf{X}_{k-1},\theta)$$
$$\times p(\mathbf{X}_{k-1},\tau_{k-1}^{1:N}|\mathbf{Z}_{1:k-1}^1,\ldots,\mathbf{Z}_{1:k-1}^N,\theta)\mathrm{d}\mathbf{X}_{k-1}, \qquad (14)$$

where we denote by $\tau_k^{1:N}$ the concatenation of τ_k^is and it has $(M!)^N$ different configurations. Here, both the prediction (14) and update (13) are $\mathcal{O}((M!)^N)$.

In order to address these challenges, multi-object filtering (or, tracking) algorithms often employ two approximations: First, they aim to find the most probable data association hypothesis in (13) denoted by $\bar{\tau}_k^{1:N}$, instead of both evaluating this expression for all possible associations and storing them. The benefits of doing so are that i) one can generate tracks (or, object trajectories) as simply marginals of $p(\mathbf{X}_k,\tau_k^{1:N} = \bar{\tau}_k^{1:N}|.)$ for $k = 1,...,t$, and, ii) evaluations of the integral in (14) in the next time step can be restricted to this value of

the hypothesis variable. Equivalently, the posterior distribution in (13) is factorised as

$$
\begin{aligned}
&p(\mathbf{X}_k, \tau_k^{1:N} | \mathbf{Z}_{1:k}^1, \ldots, \mathbf{Z}_{1:k}^N; \theta) \\
&= p(\mathbf{X}_k | \mathbf{Z}_{1:k}^1, \ldots, \mathbf{Z}_{1:k}^N, \theta, \tau_k^{1:N}) p(\tau_k^{1:N} | \mathbf{Z}_{1:k}^1, \ldots, \mathbf{Z}_{1:k}^N, \theta) \\
&= p(\mathbf{X}_k | \mathbf{Z}_{1:k}^1, \ldots, \mathbf{Z}_{1:k}^N, \theta, \tau_k^{1:N}) p(\tau_k^{1:N} | \mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N, \theta)
\end{aligned} \quad (15)
$$

where the association variables appear as model parameters in the first term and the second term is similar to a prior distribution on these models with the difference that it is conditioned on the current measurements. At time $k-1$, let us select this "empirical prior" as

$$
p(\tau_{k-1}^{1:N} | \mathbf{Z}_{k-1}^1, \ldots, \mathbf{Z}_{k-1}^N, \theta) \leftarrow \delta_{\bar{\tau}_{k-1}^{1:N}}(\tau_{k-1}^{1:N}) \quad (16)
$$

where δ is Kronecker's delta function and $\leftarrow$ denotes assignment. The second approximation follows from the first one: Evaluation of the prediction stage in (14) reduces to evaluation of the Chapman-Kolmogorov equation for only the most likely value of the association parameters. This approach is often referred to as empirical Bayes [28], and is used to facilitate approximate solutions to otherwise intractable problems.

A similar approximation can be used when evaluating the parameter posterior in (11). This leads to the following likelihood

$$
\begin{aligned}
&l\left(\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N | \theta, \tau_{1:t}^{1:N}\right) = \\
&\prod_{k=1}^t p_{\tau_k^{1:N}}\left(\mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N | \mathbf{Z}_{1:k-1}^1, \ldots, \mathbf{Z}_{1:k-1}^N, \theta\right),
\end{aligned} \quad (17)
$$

conditioned on $\tau_{1:t}^{1:N}$ where the factors are defined by

$$
\begin{aligned}
&p_{\tau_k^{1:N}}\left(\mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N | \mathbf{Z}_{1:k-1}^1, \ldots, \mathbf{Z}_{1:k-1}^N, \theta\right) \triangleq \\
&\int_{\mathfrak{X}^M} l(\mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N | \mathbf{X}_k, \tau_k^{1:N}; \theta) \\
&\qquad \times p(\mathbf{X}_k | \mathbf{Z}_{1:k-1}^1, \ldots, \mathbf{Z}_{1:k-1}^N; \theta) \mathrm{d}\mathbf{X}_k.
\end{aligned} \quad (18)
$$

This likelihood evaluated at $\tau_{1:k}^{1:N} = \bar{\tau}_{1:k}^{1:N}$ replaces the one in (8) when the empirical (model) prior is selected as in (16) (see Appendix A for details). We will refer to (17) as the empirical likelihood. Note that (18) is the integral term in (9).

The empirical likelihood update term is computationally more convenient, however, alone it is not sufficient for scalability with the number of sensors N: Finding $\bar{\tau}_k^{1:N}$ is equivalently an $N+1$-dimensional assignment problem which is NP hard even for $N=2$ sensors [35] which partly underlies the local filtering paradigm for multi-sensor processing and our interest in compatible solutions. For a moment, let us consider the problem for a single object, i.e., for $M=1$. In this case τ_k^i for $i=1,\ldots,N$ have only one possible configuration (i.e., there is no data association uncertainty). Because the dimensionality of θ is specified by N and (8) will be evaluated for roughly NL samples (when estimating (12) (see, e.g., [36])) each of which costing – in the simplest linear Gaussian measurements case (9) [1]– at the least $O(N^2 t)$,

---

[1]Specifically, for linear Gaussian measurements with no data association uncertainty, the marginal parameter likelihood involves computation of the innovation covariance for the so called group-sensor measurements in joint multi-sensor filtering.

the computational cost will be cubic in the number of sensors which can easily become prohibitive for large N.

The networked setting has additional constraints to take into account: The sensors perform local filtering of their measurements and exchange filtered (track) distributions over $\mathcal{G}$ as opposed to transmitting their measurements [29]. As a result, the network-wide measurements are not available to evaluate the likelihood of the problem. Instead, local distributions we denote by $p(\mathbf{X}_k, \tau_k^j = \hat{\tau}_k^j | \mathbf{Z}_{1:k}^j)$ are made available to neighbouring nodes where $\hat{\tau}_k^j$ is an approximation to the most probable association configuration $\bar{\tau}_k^j$ found locally, based on only the local sensor measurements at sensor j. There are computationally efficient algorithms for finding such solutions for the single sensor problem (see, e.g., [35] and the references therein). Therefore, a viable solution needs to build upon these densities and local data associations $\hat{\tau}_k^i$ as opposed to joint multi-sensor filtering in the network.

The problem we address in this work is the design of scalable approximations to (8) for estimating θ in a networked setting based on local filtering results at the nodes. The proposed approach also addresses the aforementioned computational bottleneck at fusion centres in centralised multi-sensor architectures with a designated node receiving unfiltered sensor measurements.

It is also worth noting that the parameter vector $\theta \in \mathcal{B}$ can be used to represent a wide variety of parameters of the global model some of which can be intrinsic to sensors $i \in \mathcal{V}$ individually such as parameters pertaining to local noise models. We are particularly interested in a second class of parameters which have dependencies among sensors such as respective parameters such as sensor locations and similar "calibration" parameters. In the former setting, the estimation of local parameters decouple into independent estimation problems which can be solved using a suitable approach (see, e.g., [27], [37], [38]).

In our setting, $\theta \triangleq [\theta_1, \ldots, \theta_N]$ where θ_i is associated with $i \in \mathcal{V}$ and its estimation does not decouple and depends on all measurements across the network due to the dependencies of parameters, which, in turn, brings forward the multi-sensor aspects of the problem this work aims to address. Because local filtering is performed, on the other hand, local estimation of data association $\hat{\tau}_k^i$ is available independent of θ, which we discuss in detail later in Section V.

III. A PSEUDO-LIKELIHOOD AND A PAIRWISE MRF POSTERIOR FOR DECENTRALISED ESTIMATION

Pseudo-likelihoods are constructed from likelihood like functions which are computationally convenient and defined typically over smaller subsets of the data to overcome difficulties posed by intractable likelihoods over the entire set of data (see, e.g., [18] and the references therein). Let us denote the network-wide data set by

$$
\mathbf{Z} \triangleq [\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N].
$$

A fairly general form for a pseudo-likelihood is given by [18]

$$
\tilde{l}(\mathbf{Z}|\theta) = \prod_{s \in S} \tilde{l}(\mathbf{Z}_{d_s} | \mathbf{Z}_{c_s}, \theta)^{\omega_s} \quad (19)
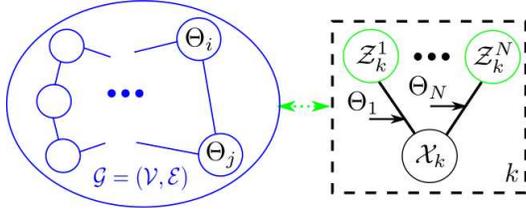$$

Fig. 1. A multi-sensor state space - or, hidden Markov- model (black dashed box on the right representing a chain over k) and a Markov Random field model of the parameter posterior (the blue edges on the left).

where S is an index set, ω_s is a positive real number, and, $\mathbf{Z}_{d_s}$ and $\mathbf{Z}_{c_s}$ are mutually exlusive subsets of $\mathbf{Z}$ (for example, $\mathbf{Z}_{d_1} = \mathbf{Z}_2^i$ and $\mathbf{Z}_{c_1} = \mathbf{Z}_1^j$, etc.). These sets can be selected in various ways ensuring that the factors are computationally convenient functions, for example, marginals and/or conditional densities, and, estimates based on $\tilde{l}(\mathbf{Z}|\theta)$ are sensible. Note that (8) is also in this form, however, with difficult to evaluate factors.

Let us consider $\theta = [\theta_1, ..., \theta_N]$ and a pseudo-likelihood surrogate for the empirical likelihood in (17):

$$\tilde{l}_{\tau_{1:t}^{1:N}}(\mathbf{Z}|\theta) = \prod_{(i,j)\in\mathcal{E}} l_{\tau_{1:t}^{i,j}}(\mathbf{Z}^i, \mathbf{Z}^j|\theta_{i,j}), \quad (20)$$

$$= \prod_{(i,j)\in\mathcal{E}} \prod_{k=1}^{t} p_{\tau_k^{i,j}}(\mathbf{Z}_k^i, \mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{k-1}^j, \theta_{i,j}). (21)$$

Here, $\mathcal{E}$ is essentially the set of sensor pairs whose likelihoods we would like to incorporate into the pseudo-likelihood, and, it is convenient to choose them as those that share a communication link, in a networked setting (Section II-A).

The pairwise structure above is beneficial to use with the MMSE estimator in (12). Note that the MMSE estimate is the concatenation of the expected values of posterior marginals, i.e., $p(\theta_i|\mathbf{Z})$ for $i = 1, \dots, N$. These distributions can be found using message passing algorithms over $\mathcal{G}$ when the surrogate (20) is used in (11) together with independent but arbitrary *a priori* distributions selected for Θ_is. Specifically, the parameter posterior corresponding to such a selection of priors and the pseudo-likelihood (20) is a pairwise Markov random field over $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [34]:

$$p(\theta|\mathbf{Z}) \propto \prod_{i\in\mathcal{V}} \psi_i(\theta_i) \prod_{(i,j)\in\mathcal{E}} \psi_{ij}(\theta_i, \theta_j), \quad (22)$$

$$\psi_i(\theta_i) = p_{0,i}(\theta_i),$$
$$\psi_{ij}(\theta_i, \theta_j) = l_{\tau_{1:t}^{i,j}}(\mathbf{Z}^i, \mathbf{Z}^j|\theta_i, \theta_j), \quad (23)$$

where the node potential functions (i.e., ψ_is) are the selected priors (e.g., uniform distributions over bounded sets θ_is take values from) and the edge potentials (i.e., ψ_{ij}s) are the pairwise likelihoods for the pairs (i, j)s. This model is illustrated in Fig. 1.

The pairwise MRF model in (22) allows the computation of posterior marginal $p(\theta_i|\mathbf{Z})$ through iterative local message passings such as Belief Propagation (BP) [30]. 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

$$m_{ji}(\theta_i) = \int \psi_{ij}(\theta_i, \theta_j) \psi_j(\theta_j) \prod_{i'\in ne(j)\backslash i} m_{i'j}(\theta_j) \, \mathrm{d}\theta_j, (24)$$

$$\tilde{p}_i(\theta_i) \propto \psi_i(\theta_i) \prod_{j\in ne(i)} m_{ji}(\theta_i). \quad (25)$$

In BP iterations, nodes simultaneously send messages to their neighbours using (24) (often using constants as the previously received messages during the first step) and update their local "belief" using (25). 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 [30]. For the case in which $\mathcal{G}$ contains cycles, iterations of (24) and (25) are known as loopy BP (LBP). For the case, convergence does not have general guarantees, 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 [39]–[41]. 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 [42] that will be converged when initial beliefs are selected reasonably [43].

IV. QUAD-TERM NODE-WISE SEPARABLE LIKELIHOODS

The pseudo-likelihood introduced in Section III leads to a parameter posterior that admits a pairwise MRF model. This is advantageous in providing a means for decentralised estimation through message passing algorithms in a network. The edge potentials (23) of this model, however, are i) conditioned jointly on two sensors' measurements simultaneous access to which is infeasible in a networked setting, and, ii) conditioned on association variables for two sensors and ideally should be evaluated at its most probable configuration $\bar{\tau}_1^{i,j}, \dots, \bar{\tau}_t^{i,j}$ each of which is NP-hard to find, as explained in Section II.

In order to overcome these difficulties, we introduce an approximation which factorises into terms local to nodes, i.e., a node-wise separable approximation. Let us consider the "centralised" pairwise likelihood update term in (21) given some configuration $\tau_k^{i,j}$ for $k = 1, \dots, t$, and, drop them from the subscript for the sake of simplicity in notation, as well as the i, j subscript in θ, in the following discussion. This term factorises in alternative ways as follows:

$$p(\mathbf{Z}_k^i, \mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta)$$
$$= p(\mathbf{Z}_k^i|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k}^j, \theta)p(\mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \quad (26)$$
$$= p(\mathbf{Z}_k^j|\mathbf{Z}_{1:k}^i, \mathbf{Z}_{1:k-1}^j, \theta)p(\mathbf{Z}_k^i|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \quad (27)$$
$$= \left(p(\mathbf{Z}_k^i|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k}^j, \theta)p(\mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \right)^{1/2}$$
$$\times \left(p(\mathbf{Z}_k^j|\mathbf{Z}_{1:k}^i, \mathbf{Z}_{1:k-1}^j, \theta)p(\mathbf{Z}_k^i|\mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \right)^{1/2} (28)$$

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. All four factors in Eq.(28) are conditioned on the measurement histories of both sensors to which one cannot have simulatenous access in a networked

setting. We would like to aviod this by leaving out the history of sensor i (sensor j) in the first two (last two) terms of (28), i.e.,

$$
\begin{aligned}
& q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \\
& \triangleq \frac{1}{\kappa_k(\theta)} \left(p(\mathbf{Z}_k^i | \mathbf{Z}_{1:k}^j, \theta) p(\mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^j, \theta) \right)^{1/2} \\
& \qquad \times \left(p(\mathbf{Z}_k^j | \mathbf{Z}_{1:k}^i, \theta) p(\mathbf{Z}_k^i | \mathbf{Z}_{1:k-1}^i, \theta) \right)^{1/2} \quad (29)
\end{aligned}
$$

$$
\begin{aligned}
\kappa_k(\theta) = \int \int & \left(p(\mathbf{Z}_k^{i\,\prime}, \mathbf{Z}_k^{j\,\prime} | \mathbf{Z}_{1:k-1}^j, \theta) \right. \\
& \left. \times p(\mathbf{Z}_k^{i\,\prime}, \mathbf{Z}_k^{j\,\prime} | \mathbf{Z}_{1:k-1}^i, \theta) \right)^{1/2} \mathrm{d}\mathbf{Z}_k^{i\,\prime} \mathrm{d}\mathbf{Z}_k^{j\,\prime} \quad (30)
\end{aligned}
$$

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 its factors depend on single sensor histories. As such, they require filtering of sensor histories of i and j individually enabling the evaluation of their product in a network. This point is discussed later in this section.

A. Approximation quality

We consider the difference between the original centralised update term in (28) and the quad-term approximation introduced in (29). Because these terms are probability densities over sensor measurements, their "divergence" can be quantified using the KLD [31]:

Proposition 4.1: The KLD between the centralised update and the node-wise separable approximation in (29) is bounded by the average of the mutual information (MI) [31] between the current measurement pair and a single sensor's history conditioned on the history of the other sensor, i.e.,

$$
\begin{aligned}
& D(p(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \| q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta)) \\
& \qquad \leqslant \frac{1}{2} I(\mathcal{Z}_k^i, \mathcal{Z}_k^j; \mathcal{Z}_{1:k-1}^i | \mathcal{Z}_{1:k-1}^j, \Theta) \\
& \qquad\qquad + \frac{1}{2} I(\mathcal{Z}_k^i, \mathcal{Z}_k^j; \mathcal{Z}_{1:k-1}^j | \mathcal{Z}_{1:k-1}^i, \Theta). \quad (31)
\end{aligned}
$$

The proof can be found in Appendix B [2]. The upper bound in (31) measures the departure of the current pair of measurements, and, one of the sensor histories from a state of conditional independence when they are conditioned on the history of the other sensor. Note that these variables, when conditioned on $\mathcal{X}_k$, are conditionally independent, i.e., $(\mathcal{Z}_k^i, \mathcal{Z}_k^j) \perp\!\!\!\perp \mathcal{Z}_{1:k-1}^j | \mathcal{X}_k, \Theta$ holds and consequently

$$
I(\mathcal{Z}_k^i, \mathcal{Z}_k^j; \mathcal{Z}_{1:k-1}^i | \mathcal{X}_k, \Theta) = I(\mathcal{Z}_k^i, \mathcal{Z}_k^j; \mathcal{Z}_{1:k-1}^j | \mathcal{X}_k, \Theta) = 0.
$$

Similarly, the average MI term on the right hand side of (31) is zero if $(\mathcal{Z}_k^i, \mathcal{Z}_k^j) \perp\!\!\!\perp \mathcal{Z}_{1:k-1}^i | \mathcal{Z}_{1:k-1}^j, \Theta$ and $(\mathcal{Z}_k^i, \mathcal{Z}_k^j) \perp\!\!\!\perp \mathcal{Z}_{1:k-1}^j | \mathcal{Z}_{1:k-1}^i, \Theta$ hold simultaneously. This condition is satisfied, for example, in the case that either of

---

the measurement histories $\mathcal{Z}_{1:k-1}^i$ and $\mathcal{Z}_{1:k-1}^j$ are sufficient statistics for $\mathcal{X}_k$ (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 (31) further to the uncertainty on $\mathcal{X}_k$ given the sensor histories:

Corollary 4.2: The KLD considered in Proposition 4.1 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:

$$
\begin{aligned}
& D(p(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \| q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta)) \\
& \leqslant \frac{1}{2} \Bigg(\Big(H(\mathcal{X}_k | \mathcal{Z}_{1:k-1}^j, \Theta) - H(\mathcal{X}_k | \mathcal{Z}_{1:k-1}^j, \mathcal{Z}_{1:k-1}^i, \Theta) \Big) \\
& \qquad + \Big(H(\mathcal{X}_k | \mathcal{Z}_{k-1}^i, \Theta) - H(\mathcal{X}_k | \mathcal{Z}_{1:k-1}^j, \mathcal{Z}_{1:k-1}^i, \Theta) \Big) \Bigg) \\
& \quad + \frac{1}{2} \Bigg(\Big(H(\mathcal{X}_k | \mathcal{Z}_{1:k}^j, \Theta) - H(\mathcal{X}_k | \mathcal{Z}_{1:k}^j, \mathcal{Z}_{1:k-1}^i, \Theta) \Big) \\
& \qquad + \Big(H(\mathcal{X}_k | \mathcal{Z}_{1:k}^i, \Theta) - H(\mathcal{X}_k | \mathcal{Z}_{1:k}^i, \mathcal{Z}_{1:k-1}^j, \Theta) \Big) \Bigg), \quad (32)
\end{aligned}
$$

where H denotes the Shannon differential entropy [31]. The proof is provided in Appendix C. Corollary 4.2 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 difference terms on the RHS of (32) quantify the difference in uncertainty between estimating the target state $\mathcal{X}_k$ using only the local measurements, and, also taking into account the other sensor's measurements. Overall, a better quality of approximation should be expected when the local filtering densities involved concentrate around a single point in the state space.

B. The quad-term pairwise likelihood

The quad-term update in (29) leads to a separable approximate likelihood given by

$$
\tilde{l}\left(\mathbf{Z}_{1:t}^i, \mathbf{Z}_{1:t}^j | \theta \right) = \prod_{k=1}^{t} q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) \quad (33)
$$

We refer to this term 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 factors of (29) over k. Let us define

$$
\begin{aligned}
r_{ij}^k(\mathbf{Z}_k^i, \theta) & \triangleq p(\mathbf{Z}_k^i | \mathbf{Z}_{1:k}^j, \theta), \\
s_j^k(\mathbf{Z}_k^j, \theta) & \triangleq p(\mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^j, \theta).
\end{aligned}
$$

Then, the quad-term update in (33) is given by

$$
\begin{aligned}
& q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta) = \\
& \frac{1}{\kappa_k(\theta)} \left(r_{ij}^k(\mathbf{Z}_k^i, \theta) s_j^k(\mathbf{Z}_k^j, \theta) \right)^{1/2} \left(r_{ji}^k(\mathbf{Z}_k^j, \theta) s_i^k(\mathbf{Z}_k^i, \theta) \right)^{1/2}, \\
& \hspace{10cm} (34)
\end{aligned}
$$

where the normalisation factor is given in (30), and, equivalently in terms of the four factors above as

$$\kappa_k(\theta) = \int \int \left(r_{ij}^k(\mathbf{Z}_k^{i\,'}, \theta) s_j^k(\mathbf{Z}_k^{j\,'}, \theta) \right)^{1/2}$$
$$\times \left(r_{ji}^k(\mathbf{Z}_k^{j\,'}, \theta) s_i^k(\mathbf{Z}_k^{i\,'}, \theta) \right)^{1/2} \mathrm{d}\mathbf{Z}_k^{i\,'} \mathrm{d}\mathbf{Z}_k^{j\,'}.$$

Corollary 4.3: The KLD between the parameter likelihood in (8) and the node-wise separable approximation in (33) is bounded by the terms on the right hand sides of (31) and (32) summed over $k = 1, \ldots, t$ as

$$D\left(l\left(\mathbf{Z}_{1:t}^i, \mathbf{Z}_{1:t}^j | \theta\right) || \tilde{l}\left(\mathbf{Z}_{1:t}^i, \mathbf{Z}_{1:t}^j | \theta\right) \right) = \sum_{k=1}^{t} D\left(p||q\right). \quad (35)$$

Proof. Eq. (35) 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 (31) and (32) over $k = 1, ..., t$. ∎

As a conclusion, when t is not large – e.g., on the order of tens which is typical in fusion applications– the proposed approximation can be used for parameter estimation via local filtering. Sensors that are more accurate in inferring the underlying state process result with a smaller KLD in (35), which in turn leads to a more favourable estimation performance.

One other approximation based on local filtering distributions was studied in [26] which has the following dual-term product form

$$u(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta)$$
$$\triangleq p(\mathbf{Z}_k^i | \mathbf{Z}_{1:k-1}^i, \theta) p(\mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \theta). \quad (36)$$

In Appendix D, we shown that $D(p||q) < D(p||u)$, when sensors are equivalent. The entropy bound given in (32) is also smaller than that for the dual-term approximation. In other words, the quad-term approximation is more accurate compared to the dual-term approximation, under typical operating conditions.

The scaling factor of the dual term approximation is unity regardless of θ, on the other hand, 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, the dual-term pseudo-likelihood is used with random finite set variables (RFS) in [26], [44], which, in a sense, have the association variables marginalised out making it possible to avoid multi-dimensional assignment problems in the general multi-object tracking model. For RFS distributions, however, it is not straightforward to compute the scaling factor in (30) for the quad-term. In this article, we consider a parametric model instead, which is effectively configured through association variables.

V. QUAD-TERM LIKELIHOOD FOR SENSOR CALIBRATION PARAMETERS

The results presented so far are fairly general and do not depend on the nature of Θ. When Θ represents respective parameters such as calibration parameters, there are certain simplifications of the expressions involved which provide computational benefits. In particular, parameters such as respective location and bearing angles relate the local coordinate frames of the sensors which collect measurements in their local frame. The local filtering distributions are hence over the space of state vectors in the local frame. A point $x_k \in \mathfrak{X}$ (Section II-A) is implicitly in the Earth coordinate frame (ECF), and, associated with its representation in the jth local frame $[x_k]_j$ through a coordinate transform T with the following properties

$$[x_k]_j = T(x_k; \theta_j), \quad (37)$$
$$[x_k]_i = T(T^{-1}([x_k]_j; \theta_j); \theta_i).$$

As an example, when x_k is a location on the Cartesian plane, and, θ_j is the position of sensor j, T is given by

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

For simplicity in notation, we will denote $T(T^{-1}(.; \theta_j); \theta_i)$ by $\mathcal{T}_\theta(.)$ when semantics is clear from the context[3].

In this section, it is revealed how local filtering distributions are used in the quad-term update. When θ are respective quantities, these distributions become independent of θ because both the state and the measurement variables are in the same local coordinate frame. In other words, at sensor j

$$p(\mathbf{X}_k, \tau_k^j | \mathbf{Z}_{1:k}^j, \theta_j) \equiv p([\mathbf{X}_k]_j, \tau_k^j | \mathbf{Z}_{1:k}^j), \quad (38)$$

holds for the filtering posterior, for any configuration of τ_k^j.

In the prediction stage of filtering, the multi-object transition kernel in (5) also becomes independent of Θ, so, the Chapman-Kolmogorov equation for finding the prediction density at sensor j (together with the empirical Bayes selection of association priors during iterations as explained in Section II) becomes

$$p([\mathbf{X}_k]_j | \mathbf{Z}_{1:k-1}^j) = \int \pi([\mathbf{X}_k]_j | [\mathbf{X}_{k-1}]_j)$$
$$\times p([\mathbf{X}_{k-1}]_j, \tau_{k-1}^j = \bar{\tau}_{k-1}^j | \mathbf{Z}_{1:k-1}^j) \mathrm{d}[\mathbf{X}_k]_j. \quad (39)$$

Note also that the entropy terms in (32) that are conditioned on a single sensor's measurements measure the uncertainty of the above densities. Consequently, they also become independent of Θ, i.e., $H(\mathcal{X}_k | \mathcal{Z}_{1:k-1}^j, \Theta)$ equals to $H(\mathcal{X}_k | \mathcal{Z}_{1:k-1}^j)$ for example, which highlights its relevance to the local prediction accuracy.

A. The quad-term update for calibration

Now, let us expand the quad-term time updates in (34), and, explicitly show the aforementioned simplifications. We start with s_j^k which is the scale factor of the local Bayesian filter at sensor j:

$$s_j^k(\mathbf{Z}_k^j, \theta) = p_{\tau_k^j}(\mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^j, \theta),$$
$$= \int l_j(\mathbf{Z}_k^j | \mathbf{X}_k, \tau_k^j, \theta) p(\mathbf{X}_k | \mathbf{Z}_{1:k-1}^j, \theta) \mathrm{d}\mathbf{X}_k,$$
$$= \int l_j(\mathbf{Z}_k^j | [\mathbf{X}_k]_j, \tau_k^j) p([\mathbf{X}_k]_j | \mathbf{Z}_{1:k-1}^j) \mathrm{d}[\mathbf{X}_k]_j, (40)$$

---

[3]Note that, when the calibration parameters also include orientation angles, $\mathcal{T}_\theta(.)$ involves rotation matrices accordingly.

where in the last line independence from θ is asserted. Because s_j^k does not depend on θ, we denote $s_j^k(\mathbf{Z}_k^j,\theta)$ by $s_j^k(\mathbf{Z}_k^j)$ in the rest of the article.

Next, let us consider r_{ij}^k which has terms in different coordinate frames:

$$r_{ij}^k(\mathbf{Z}_k^i,\theta) = p_{\tau_k^{i,j}}(\mathbf{Z}_k^i|\mathbf{Z}_{1:k}^j,\theta)$$

$$= \int l_i(\mathbf{Z}_k^i|\mathbf{X}_k,\tau_k^i,\theta)p(\mathbf{X}_k,\tau_k^j|\mathbf{Z}_{1:k}^j,\theta)\mathrm{d}\mathbf{X}_k$$

$$= \int l_i\left(\mathbf{Z}_k^i|T(\mathbf{X}_k;\theta_i),\tau_k^i\right)p\left(T^{-1}(\mathbf{X}_k;\theta_j),\tau_k^j|\mathbf{Z}_{1:k}^j\right)\mathrm{d}\mathbf{X}_k$$

$$= \int l_i\left(\mathbf{Z}_k^i|\mathcal{T}_\theta([\mathbf{X}_k]_j),\tau_k^i\right)p\left([\mathbf{X}_k]_j,\tau_k^j|\mathbf{Z}_{1:k}^j\right)\mathrm{d}[\mathbf{X}_k]_j. \quad (41)$$

In the third line above, the coordinate transformations are substituted explicitly. The last line follows from that the filtering distribution (38) is in the jth local frame.

B. Evaluation of the quad-term update based on single sensor filtering distributions

Here, we discuss the evaluation of the quad-term likelihood given local filtering distributions and single sensor association configurations which we denote for sensor j by $p([\mathbf{X}_k]_j,\tau_k^j=\hat{\tau}_k^j|\mathbf{Z}_{1:k}^j)$ and $\hat{\tau}_k^j$, respectively, as explained in Section II-B.

Instead of considering evaluation for the most probable association hypothesis $\bar{\tau}_k^{i,j}$ which is infeasible to find, we propose to use the local results $\hat{\tau}_k^{i,j} \triangleq (\hat{\tau}_k^i,\hat{\tau}_k^j)$ as a reasonable approximation to this configuration and substitute them in (40)–(41). These approximations can be found regardless of θ as discussed earlier in this section by using one of the well studied algorithms in the literature [2] such as solving a $2-D$ association problem at each time step [35] to find $\hat{\tau}_k^j$. We detail this approach for a linear Gaussian state space model in Section VI.

Given these local results and their exchange over the network, one can consider an in-network computation scheme for evaluating (40)–(41). Specifically, these terms (and, the other factors of the quad-term update which are obtained by replacing i and j in these expressions) can be found at the sensor platform where the measurements to be substituted for evaluation are stored, i.e., sensors j and i, respectively, for s_j^k and r_{ij}^k. Substitution of the measurement histories on the conditioning side will have been carried out by local filtering.

More explicitly, s_j^k in (40) (or, s_i^k) becomes a product of similar terms when $\hat{\tau}_j^k$ is substituted in (40), and, its computation is carried out during the local filtering of sensor j's (or, sensor i's) measurements using

$$s_j^k(\mathbf{Z}_k^j) = \prod_{o=1}^M s_{j,o}^k(z_{k,o}^j) \quad (42)$$

$$s_{j,o}^k(z_{k,o}^j) \triangleq \int g_j(z_{k,o}^j|x_k')p_{m'}(x_k'|\mathbf{Z}_{1:k-1}^j)\mathrm{d}x_k'$$

where $m' = \hat{\tau}_k^j(o)$ and the density in the integral is the m'th marginal of the local prediction density, the mth of which is

given by

$$p_m(x'|\mathbf{Z}_{1:k-1}^j) \triangleq \int p(\mathbf{X}_k=[x_{k,1},\ldots,x_{k,m-1},x',x_{k,m+1},$$

$$x_{k,m+2},\ldots,x_{k,M}],\tau_k^j=\hat{\tau}_k^j|\mathbf{Z}_{1:k-1}^j)$$

$$\mathrm{d}x_{k,1}\ldots\mathrm{d}x_{k,m-1}\mathrm{d}x_{k,m+1}\ldots\mathrm{d}x_{k,M}.$$

The term r_{ij}^k in (41) (or, r_{ji}^k) is also computed based on these local filtering distributions. The integration in the RHS of (41) implicitly assumes that the ordering of individual objects in the local multi-object vectors are the same. In a networked setting, however, this is not necessarily the case and the identities of the fields in the state vector may differ [45]. In order to tackle with this unknown correspondance, we introduce an additional permutation random variable γ_k (see, e.g., [46]) for relating the fields of a multi-object vector $\mathbf{X}_k$ as ordered locally at sensor i and j, such that the mth field of the state vector at sensor i refers to the same object in the $\gamma_k(m)$th field of the state vector at sensor j. For example,

$$[x_{k,m}]_i = \mathcal{T}_\theta([x_{k,\gamma_k(m)}]_j).$$

Suppose that an estimate $\hat{\gamma}_k$ of this quantity is provided. After substituting in (41) together with $\hat{\tau}_k^{i,j}$ one obtains

$$r_{ij}^k(\mathbf{Z}_k^i,\theta) = \prod_{o=1}^M r_{ij,o}^k(z_{k,o}^i,\theta), \quad (43)$$

$$r_{ij,o}^k(z_{k,o}^i,\theta) \triangleq \int g_i\left(z_{k,o}^i|\mathcal{T}_\theta(x_k')\right)p_{m'}(x_k'|\mathbf{Z}_{1:k}^j)\mathrm{d}x_k'$$

where $m' = \hat{\gamma}_k(\hat{\tau}_k^i(o))$ and the density inside the integral is the m'th marginal of the filtering distribution local to sensor j. In Appendix E, we show that (43) replaces the likelihood for γ_k when $m' = \gamma_k(\hat{\tau}_k^i(o))$ and an ML estimate $\hat{\gamma}_k$ can be found in a way similar to solving the data association problem in local filtering, which is detailed later in Section VI-A.

Finally, the scale factor (30) is computed. This involves finding the measurement distributions in (30) using the prediction distribution in (39) for both sensors i and j together with their likelihoods. This leads to the following two decomposition: The first term in (30) is found as

$$p(\mathbf{Z}_k^{i\,'},\mathbf{Z}_k^{j\,'}|\mathbf{Z}_{1:k-1}^i,\theta) = \prod_{o=1}^M p_o(z_{k,o}^{i\,'},z_{k,\xi(o)}^{j\,'}|\mathbf{Z}_{1:k-1}^i,\theta)$$

$$p_o(z^i,z^j|\mathbf{Z}_{k-1}^i,\theta) \triangleq \int g_i(z^i|x_k')p_{m'}(x_k'|\mathbf{Z}_{1:k-1}^i)\mathrm{d}x_k'$$

$$\times \int g_j(z^j|\mathcal{T}_\theta^{-1}(x_k'))p_{m'}(x_k'|\mathbf{Z}_{1:k}^i)\mathrm{d}x_k' \quad (44)$$

where $\xi(o) = \hat{\tau}_k^{j-1} \circ \hat{\tau}_k^i(o)$ maps the oth measurement at sensor i to the corresponding one in sensor j, and, $m' = \hat{\tau}_k^i(o)$ in the second line.

The second term in (30) is found as

$$p(\mathbf{Z}_k^{i\,'},\mathbf{Z}_k^{j\,'}|\mathbf{Z}_{1:k-1}^j,\theta) = \prod_{o=1}^M p_o(z_{k,o}^{i\,'},z_{k,\xi(o)}^{j\,'}|\mathbf{Z}_{k-1}^j,\theta)$$

$$p_o(z^i,z^j|\mathbf{Z}_{k-1}^j,\theta) \triangleq \int g_j(z^j|x_k')p_{m'}(x_k'|\mathbf{Z}_{1:k-1}^j)\mathrm{d}x_k'$$

$$\times \int g_i(z^i|\mathcal{T}_\theta(x'_k))p_{m'}(x'_k|\mathbf{Z}^j_{1:k}) \, dx'_k \tag{45}$$

where $m' = \hat{\gamma}_k \circ \hat{\tau}^i_k(o)$ in the last line is the object that corresponds to the oth measurement at sensor i.

Consequently, the scale factor is found as

$$\kappa_k(\theta) = \prod_{o=1}^{M} \kappa_{k,o}(\theta) \tag{46}$$

$$\kappa_{k,o}(\theta) \triangleq \int \left(p_o(z^i, z^j|\mathbf{Z}^i_{1:k-1}, \theta) \right.$$
$$\left. \times p_o(z^i, z^j|\mathbf{Z}^j_{1:k-1}, \theta) \right)^{1/2} dz^i dz^j,$$

using the densities found in (44) and (45).

The expressions above describe the evaluation of the quad-term likelihood in terms of single sensor filtering distributions that can be obtained using any filtering algorithm with individual measurement histories. The use of the local filtering distributions provides scalability with the number of sensors for parameter estimation in the state space model in Section II.

These computations can be distributed in the pair (i, j) as follows: Both sensors i and j perform local filtering and exchange the resulting posterior densities at every step, as well as s_i and s_j, respectively, found using (42). Based on the received densities, sensors i and j evaluate r_{ij} and r_{ji}, respectively, using (43). As part of the filtering process, they realise the Chapman-Kolmogorov equation in (39) with their local posterior, as well as the remote posterior recently received. These densities are then used in (44), (45), and, (46) to compute the scale factor for the next step. The scale factor hence found in the previous step, therefore, is substituted in (34) together with the four terms computed. This is repeated for $k = 1, \ldots, t$, and, the quad-term likelihood (33) is computed, as a result.

VI. A MONTE CARLO LBP ALGORITHM FOR SENSOR CALIBRATION IN LINEAR GAUSSIAN STATE SPACE MODELS

In this section, we consider a linear Gaussian state space (LGSS) model within the probabilistic graphical model in Fig. 1, and, specify an algorithm for estimation of θs that combines the quad-term calibration likelihood evaluation detailed in Section V with BP message passing on the resulting pairwise MRF model (Section III). This algorithm uses Monte Carlo methods for realising BP [32] and facilitates scalability by building upon single sensor filtering as required by the quad-term approximation. As a result, an efficient inference scheme over the model in Fig. 1 is achieved.

The state space model we consider is specified by a linear state transition with process noise that is additive and Gaussian, and, linear measurements with independent Gaussian measurement noise, i.e.,

$$\pi(x_k|x_{k-1}) = \mathcal{N}(x_k; \mathbf{F}x_{k-1}, \mathbf{Q}) \tag{47}$$
$$g_j(z^j_k|x_k; \theta_j) = \mathcal{N}(z^j_k; \mathbf{H}_j[x_k]_j, \mathbf{R}_j) \tag{48}$$

for $j = 1, \ldots, N$, where $\mathcal{N}(.; \mu, \mathbf{P})$ is a multi-dimensional Gaussian density with mean vector μ and covariance matrix $\mathbf{P}$.

Here, x_k is the concatenation of position and velocity (on a $2 - D$ Euclidean plane, without loss of generality). The

matrices $\mathbf{F}$ and $\mathbf{Q}$ model motion with unknown acceleration (equivalently, manoeuvres), and, are selected as

$$\mathbf{F} = \begin{bmatrix} \mathbf{I}, & \Delta T \times \mathbf{I} \\ \mathbf{0}, & \mathbf{I} \end{bmatrix}, \quad \mathbf{Q} = \sigma^2 \begin{bmatrix} q_1\mathbf{I}, & q_2\mathbf{I} \\ q_2\mathbf{I}, & q_3\mathbf{I} \end{bmatrix}$$

where $\mathbf{I}$ and $\mathbf{0}$ are the 2×2 identity and zero matrices, respectively. ΔT is the time difference between consecutive steps. $\mathbf{Q}$ is positive definite and parameterised with σ^2, and, $0 < q_1 < q_2 < q_3 < 1$ specifying the magnitude of the uncertainty, and, contributions of higher order terms, respectively [4].

In the measurement model, $\mathbf{R}_j$ is the measurement noise covariance, and, $\mathbf{H}_j$ is the observation matrix which we assume forms an observable pair with $\mathbf{F}$ (e.g., $\mathbf{H}_j = [\mathbf{I}, \mathbf{0}]$).

A. Local single sensor filtering

We now focus on filtering and provide explicit formulae that adopts the recursions in (13)–(18) for a single sensor in the LGSS model. We use the empirical Bayes approach explained in Section II-B for scaling with time under data association uncertainties. This approach corresponds to the single frame data association solution in multi-object tracking [35].

First, let us consider the prediction stage at time k in which we are given a filtering density evaluated at the most likely data association hypothesis $\hat{\tau}^j_{k-1}$ of the previous step [5]. The latter is a product of its marginals

$$p(\mathbf{X}_{k-1}, \tau^j_{k-1} = \hat{\tau}^j_{k-1}|\mathbf{Z}^j_{1:k-1}) =$$
$$\prod_{m=1}^{M} p_m(x_{k-1,m}, \tau^j_{k-1} = \hat{\tau}^j_{k-1}|\mathbf{Z}^j_{1:k-1})$$

where

$$p_m(x_{k-1,m}, \tau^j_{k-1} = \hat{\tau}^j_{k-1}|\mathbf{Z}^j_{1:k-1}) = p(x_{k-1,m}|z^{j,m}_{1:k-1}).$$

Here, $z^{i,m}_{1:k-1}$ denotes the measurements induced by object m from step 1 to $k-1$, i.e.,

$$z^{j,m}_{1:k-1} \triangleq \left(z^j_{k-1,\rho^j_{k-1}(m)}, z^j_{k-2,\rho^j_{k-2}(m)}, \ldots, z^j_{1,\rho^j_1(m)} \right),$$

where ρ is the inverse of τ, i.e., $\rho \circ \tau$ is the identity permutation.

Consequently, the mth marginal of the posterior at $k-1$ is a Gaussian density that can be obtained equivalently by Kalman filtering [36] over $z^{j,m}_{1:k-1}$, i.e.,

$$p_m(x_{k-1,m}|z^{j,m}_{1:k-1}) = \mathcal{N}(x_{k-1,m}; \hat{x}^j_{k-1,m}, \mathbf{P}^j_{k-1,m}), \tag{49}$$

with the mean and covariance matrices over time specifying the mth "track."

---

[4]It can easily be shown that this state transition model is invariant under the selected coordinate frame for x_ks.

[5]Note that the empirical prior on τ^j_{k-1} is selected as in (16) leading to this posterior be identically zero for all values of τ^j_{k-1} other than $\hat{\tau}^j_{k-1}$.

Hence, the prediction density in (39) (Section V-A) evaluated at $\tau_{k-1}^j = \hat{\tau}_{k-1}^j$ for the state transition in (47) is given by

$$p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}^j) = \prod_{m=1}^{M} \mathcal{N}(x_{k,m}; \hat{x}_{k|k-1,m}^j, \mathbf{P}_{k|k-1,m}^j) \quad (50)$$

$$\hat{x}_{k|k-1,m}^j = \mathbf{F}\hat{x}_{k-1,m}^j,$$

$$\mathbf{P}_{k|k-1,m}^j = \mathbf{F}\mathbf{P}_{k-1,m}^j\mathbf{F}^T + \mathbf{Q},$$

where the last two lines are Kalman prediction equations with $(.)^T$ denoting matrix transpose.

Next, let us consider the update stage in which we use the prediction density (50) with M measurements concatenated in $\mathbf{Z}_k^j$ and the measurement likelihood. This likelihood is found by substituting (48) in (6). First, we use this term within the likelihood for the association variable τ_k^j which –as it is mutually independent from $\tau_1^j, \ldots, \tau_{k-1}^j$– is given by

$$l_j(\mathbf{Z}_{1:k}^j|\tau_k^j) \propto p_{\tau_k^j}(\mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^j) \quad (51)$$

$$= \int l(\mathbf{Z}_k^j|\mathbf{X}_k, \tau_k^j)p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}^j)d\mathbf{X}_k$$

$$= \prod_{o=1}^{M} \int g_j(z_{k,o}^j|x_k')p_{\tau_k^j(o)}(x_k'|\mathbf{Z}_{1:k-1}^j)dx_k'$$

$$= \prod_{o=1}^{M} \int g_j(z_{k,o}^j|x_k')p(x_k'|z_{1:k-1}^{j,\tau_k^j(o)})dx_k'.$$

The first line above follows from that the joint distribution of $Z_{1:k-1}^j$ and τ_k^j is independent of the latter.

The prior distribution for τ_k^j is non-informative as given in (7), so, the ML estimate using (51) coincides with the MAP estimate and it is given by

$$\hat{\tau}_k^j = \arg \max_{\tau_k^j \in \mathcal{S}_M} l_j(\mathbf{Z}_{1:k}^j|\tau_k^j) \quad (52)$$

where $\mathcal{S}_M$ is the set of M-permutations.

An equivalent problem is found by taking the logarithm of the objective function in the combinatorial optimisation problem above as follows:

$$\hat{\tau}_k^j = \arg \max_{\tau_k^j \in \mathcal{S}_M} \sum_{o=1}^{M} c(o, m = \tau_k^j(o)) \quad (53)$$

$$c(o,m) \triangleq \log \int g_j(z_{k,o}^j|x_k')p(x_k'|z_{1:k-1}^{j,m})dx_k',$$

for $o, m = 1, \ldots, M$.

This cost for the LGSS model is explicitly found using the prediction distribution (50) and the measurements within the KF innovations [36] as

$$c(o,m) = \log \mathcal{N}(z_{k,o}^j; \hat{z}_{k,m}^j, \mathbf{S}_{k,m}^j) \quad (54)$$

$$\hat{z}_{k,m}^j = \mathbf{H}_j\hat{x}_{k|k-1,m}^j, \quad \mathbf{S}_{k,m}^j = \mathbf{R}_j + \mathbf{H}_j\mathbf{P}_{k|k-1,m}^j\mathbf{H}_j^T.$$

The optimisation in (53) is a $2-D$ assignment problem which can be solved in polynomial time with M (despite that the search space $\mathcal{S}_M$ has a factorial size) using one of the well known solvers [47] including the auction algorithm [48]. This algorithm operates over a matrix of costs obtained by $\mathbf{C} = [c(o,m)]$ to iteratively find the M pairs corresponding to the best permutation $\hat{\tau}_k^j$ in the ML problem (52). Here, com-

putation of the M^2 cost matrix usually has the predominant computational time.

Next, we consider the state distribution update (see (15) in Section II-B) and assert the empirical (model) prior in (16). As a result, the filtering density at k becomes a product of its marginals each of which is a Gaussian as in (49) found by the KF update [36], i.e.,

$$p(\mathbf{X}_k, \tau_k^j = \hat{\tau}_k^j|\mathbf{Z}_{1:k}^j) = \prod_{m=1}^{M} \mathcal{N}(x_{k,m}; \hat{x}_{k,m}, \mathbf{P}_{k,m}) \quad (55)$$

where,

$$\hat{x}_{k,m} = \hat{x}_{k|k-1,m} + \mathbf{K}_{k,m}(z_{k,\rho_k^j(m)}^j - \mathbf{H}_j\hat{x}_{k|k-1,m})$$

$$\mathbf{P}_{k,m} = (\mathbf{I} - \mathbf{K}_{k,m}\mathbf{H}_j)\mathbf{P}_{k|k-1,m}^j$$

$$\mathbf{K}_{k,m} \triangleq \mathbf{P}_{k|k-1,m}^j\mathbf{H}_j^T\mathbf{S}_{k,m}^{j-1}.$$

Note that, because the posterior density is now non-zero only for $\tau_k^j = \hat{\tau}_k^j$, the marginalisation over τ_k^j (see, e.g., (14)) in the following prediction stage reduces to (39) and (50).

B. Evaluation of the calibration quad-term in the LGSS Model

Let us consider the evaluation of s_j^k given by (42) using the formulae for the LGSS model introduced in Section VI-A. By comparison with the cost term in (53) and (54), it can easily be seen that s_j^k is the exponential of the association cost for $\hat{\tau}_k^j$, i.e.,

$$s_j^k(\mathbf{Z}_k^j) = \exp \sum_{o=1}^{M} c(o, \hat{\tau}_k^j(o)). \quad (56)$$

Next, let us consider (43) for evaluating r_{ij}^k. Evaluation of this term involves finding the object identity correspondance $\hat{\gamma}_k$ by solving a 2-D assignment as explained in Appendix E. In the LGSS model, the assignment cost matrix $\mathcal{D} = [d(o,m)]$ is found as

$$d(o,m) = \log \mathcal{N}(z_{k,o}^i; \hat{z}_{k,m}^i, \mathbf{S}_{k,m}^i) \quad (57)$$

$$\hat{z}_{k,m}^i = \mathbf{H}_i\mathcal{T}_\theta(\hat{x}_{k,m}^i),$$

$$\mathbf{S}_{k,m}^i = \mathbf{R}_i + \mathbf{H}_i\mathcal{T}_\theta(\mathbf{P}_{k,m}^j)\mathbf{H}_i^T.$$

for $o, m = 1, \ldots, M$. Here, the second order statistics $\mathbf{P}_{k,m}^j$ is also transformed by applying any rotations involved in $\mathcal{T}_\theta$ to its eigenvectors. The best assignment which here encodes $\hat{\gamma}_k$ is found using the auction algorithm [48] as well, similar to the assignment in the Bayesian filtering update (Sec. VI-A). Using this estimate, the quad-term factor is computed using

$$r_{ij}^k(\mathbf{Z}_k^i, \theta) = \exp \sum_{o=1}^{M} d\left(o, \hat{\gamma}_k(\hat{\tau}_k^i(o))\right). \quad (58)$$

In order to evaluate the other factors of the quad-term update, i.e., s_k^i and r_{ji}^k, similar computations are used. It suffices to replace i in the subscripts/superscripts of the expressions above with j, and, vice versa.

Finally, let us consider the scale factor in (46). Let us use the notation introduced in the previous section for expressing the

densities inside the integration. Starting with (44), one obtains

$$p_o(z^i, z^j | \mathbf{Z}^i_{1:k-1}, \theta) = p(z^i, z^j | z^{i,\hat{\tau}^i_{1:k-1}(o)}_{1:k-1}, \theta) \quad (59)$$
$$= \mathcal{N}([z^i, z^j]^T; \mu_1, \boldsymbol{\Sigma}_1)$$
$$\mu_1 = \begin{bmatrix} \hat{z}^i_{k,m} \\ \mathbf{H}_j \mathcal{T}^{-1}_\theta(\hat{x}^i_{k,m}) \end{bmatrix}$$
$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \mathbf{S}^i_{k,m} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_j + \mathbf{H}_j \mathcal{T}^{-1}_\theta(\mathbf{P}^i_{k,m})\mathbf{H}^T_j \end{bmatrix}$$

where $m = \hat{\tau}^i_{k-1}(o)$, and, $\hat{z}^i_{k,m}$ and $\mathbf{S}^i_{k,m}$ are computed as in (54) with i substituted in place of j.

The second density in (46) conditioned on sensor j's history, i.e., (45), is similarly found as

$$p_o(z^i, z^j | \mathbf{Z}^j_{1:k-1}, \theta) = \mathcal{N}([z^i, z^j]^T; \mu_2, \boldsymbol{\Sigma}_2)$$
$$\mu_2 = \begin{bmatrix} \mathbf{H}_i \mathcal{T}_\theta(\hat{x}^j_{k,m}) \\ \hat{z}^j_{k,m} \end{bmatrix}$$
$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} \mathbf{R}_i + \mathbf{H}_i \mathcal{T}_\theta(\mathbf{P}^j_{k,m})\mathbf{H}^T_i & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^j_{k,m} \end{bmatrix}$$

where $m = \hat{\gamma}_{k-1} \circ \hat{\tau}^i_{k-1}(o)$ and, $\hat{z}^j_{k,m}$ and $\mathbf{S}^j_{k,m}$ are given in (54).

Using the densities above and integration rules for Gaussians, the oth term of the scale factor in (46) is found as

$$\kappa_{k,o}(\theta) =$$
$$\frac{(|\boldsymbol{\Sigma}^{-1}_1||\boldsymbol{\Sigma}^{-1}_2|)^{1/4}}{\left|\frac{\boldsymbol{\Sigma}^{-1}_1+\boldsymbol{\Sigma}^{-1}_2}{2}\right|^{1/2}} \exp\left\{ -\frac{1}{4}\left(\mu^T_1 \boldsymbol{\Sigma}^{-1}_1 \mu_1 + \mu^T_2 \boldsymbol{\Sigma}^{-1}_2 \mu_2\right) \right.$$
$$+ \frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}_1 \mu_1 + \boldsymbol{\Sigma}^{-1}_2 \mu_2\right)^T \left(\boldsymbol{\Sigma}^{-1}_1 + \boldsymbol{\Sigma}^{-1}_2\right)^{-1}$$
$$\left. \times \left(\boldsymbol{\Sigma}^{-1}_1 \mu_1 + \boldsymbol{\Sigma}^{-1}_2 \mu_2\right)\right\}. \quad (60)$$

In a distributed setting, the scale factor expressions above are computed both at sensors i and j, for which the prediction stage given in (50) is carried out for both the local posterior and the posterior recevied from the other sensor at time $k-1$.

C. Sampling from the calibration marginals using non-parametric BP

In this section, we introduce particle based representations and Monte Carlo computations [49] for the realisation of (loopy) BP message passings. Note that Sections VI-A and VI-B specify the evaluation of edge potentials given in (33) for $(\tau^i_{1:t}, \tau^j_{1:t}) = (\hat{\tau}^i_{1:t}, \hat{\tau}^j_{1:t})$.

For sampling from the marginal parameter posteriors, we adopt the approach detailed in [26, Sec.VI] for carrying out LBP belief update and messaging in (25) and (24), respectively. Given L equally weighted samples from $\tilde{p}_i(\theta_i)$, i.e.,

$$\theta^{(l)}_i \sim \tilde{p}_i(\theta_i), \quad (61)$$

for $l = 1, \ldots, L$, the edge potentials are evaluated to obtain

$$\psi_{ij}(\theta^{(l)}_i, \theta^{(l)}_j) = \tilde{l}\left(\mathbf{Z}^i_{1:t}, \mathbf{Z}^j_{1:t} | \theta = (\theta^{(l)}_i, \theta^{(l)}_j)\right). \quad (62)$$

Consider the BP message from node j to i in (24). 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.,

$$\bar{\theta}^{(l)}_j \sim \tilde{p}_j(\theta_j) \prod_{i' \in ne(j)/i} m_{i'j}(\theta_j) \text{ for } l = 1, ..., L. \quad (63)$$

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 [32]. We use Gaussian kernels leading to the approximation given by

$$\hat{m}_{ji}(\theta_i) = \sum_{l=1}^{L} \omega^{(l)}_{ji} \mathcal{N}(\theta_i; \theta^{(l)}_{ji}, \Lambda_{ji}), \quad (64)$$
$$\theta^{(l)}_{ji} = T(T^{-1}(\bar{\theta}^{(l)}_j; \theta^{(l)}_j); \theta^{(l)}_i),$$
$$\omega^{(l)}_{ji} = \frac{\psi_{i,j}(\theta^{(l)}_i, \theta^{(l)}_j)}{\sum_{l'=1}^{L} \psi_{i,j}(\theta^{(l')}_i, \theta^{(l')}_j)},$$

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 [50] and find

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

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 (25). We use the weighted bootstrap (also known as sampling/importance resampling) [51] 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

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

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

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

where $p_{0,i}$ is the prior density selected for θ_i (and, the node potential in (22)). Thus, the local calibration marginal is estimated by

$$\hat{P}_i(d\theta_i) = \sum_{l=1}^{L} \omega^{(l)}_i \delta_{\theta^{(l)}_i}(d\theta_i). \quad (65)$$

Algorithm 1 Pseudo-code for estimation of θ using the quad-term separable likelihood within Belief Propagation.

1: **for all** $j \in \mathcal{V}$ **do** $\triangleright$ Local filtering
2:    **for** $k = 1, \dots, t$ **do**
3:      Find $p(X_k, \tau_k^j = \hat{\tau}_k^j | \mathbf{Z}_{1:k}^j)$ in (55) as described in Section VI-A
4:      Find $s_j^k(\mathbf{Z}_k^j)$ in (56)
5:    **end for**
6: **end for**

7: **for all** $j \in \mathcal{V}$ **do** $\triangleright$ Sample from priors
8:    Sample $\theta_i^{(l)} \sim p_{0,i}(\theta_i)$ for $l = 1, \dots, L$ as in (61)
9: **end for**

10: **for** $s = 1, \dots, S$ **do** $\triangleright$ S-steps of LBP
11:    **for all** $(i, j) \in \mathcal{E}$ **do** $\triangleright$ Evaluate edge potentials
12:      **for** $l = 1, \dots, L$ **do**
13:        Find $r_{ij}^k(\mathbf{Z}_k^i, \theta = (\theta_i^{(l)}, \theta_j^{(l)}))$ using (57), (58) for $k = 1, \dots t$
14:        Find $r_{ji}^k(\mathbf{Z}_k^j, \theta = (\theta_i^{(l)}, \theta_j^{(l)}))$ for $k = 1, \dots t$
15:        Find $\kappa_k(\theta = (\theta_i^{(l)}, \theta_j^{(l)}))$ using (46), (59)–(60) for $k = 1, \dots, t$
16:        Find $q(\mathbf{Z}_k^i, \mathbf{Z}_k^j | \mathbf{Z}_{1:k-1}^i, \mathbf{Z}_{1:k-1}^j, \theta = (\theta_i^{(l)}, \theta_j^{(l)}))$ using (34) for $k = 1, \dots, t$
17:        Find $\psi_{i,j}(\theta_i^{(l)}, \theta_j^{(l)})$ in (62) as the quad-term likelihood in (33)
18:      **end for**
19:    **end for**
20:    **for all** $(i, j) \in \mathcal{E}$ **do** $\triangleright$ Find LBP message
21:      Find the kernel representation $\hat{m}_{ji}(\theta_i)$ in (64)
22:    **end for**
23:    **for all** $i \in \mathcal{V}$ **do** $\triangleright$ Update local marginals
24:      Find the updated $\hat{P}_i$ in (65) and sample $\theta_i^{(l)} \sim \tilde{p}_i(\theta_i)$
25:      $\hat{\theta}_i \leftarrow \frac{1}{L} \sum_{l=1}^{L} \theta_i^{(l)}$
26:    **end for**
27: **end for**

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 (63).

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 1.

VII. EXAMPLE: SELF-LOCALISATION IN LGSS MODELS

In this example, we demonstrate the quad-term node-wise separable likelihood in sensor self-localisation. The LGSS model given by (47) and (48) is used with process noise parameters selected as $\sigma = 0.5$, $q_1 = 1/4$, $q_2 = q_3 = 1/2$, $q_4 = 1$. The measurement model for sensor i is given by $\mathbf{H}_i = [\mathbf{I}, \mathbf{0}]$ and $\mathbf{R}_i = \sigma_n^2 \mathbf{I}$ with $\sigma_n = 10$ modelling noisy position measurements in the local coordinate frame.

Let us consider the multi-object multi-sensor scenario depicted in Fig. 2. 16 sensors observe 4 objects moving with data association uncertainties. The locations of the sensors are to be estimated with respect to sensor 1 which is selected as the origin of the network coordinate system. Therefore
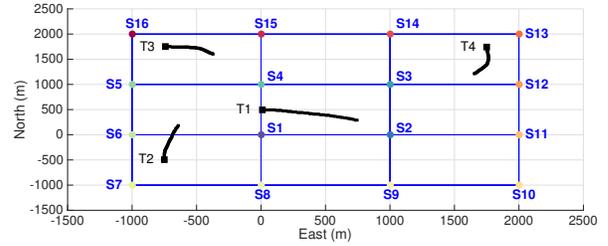

Fig. 2. Example scenario: 16 Sensors collect measurements from 4 objects (T1-T4) with association uncertainties. Initial positions of the objects are denoted by black squares. Trajectories for 60 time steps are depicted. The blue lines depict the edges of the MRF model used for estimation.

$\theta = [\theta_1, \dots, \theta_{16}]$ with the prior distribution for θ_1 selected as Dirac's delta, i.e., $p_{0,1}(\theta_1) = \delta(\theta_1)$. For the other nodes, the localisation prior, i.e., $p_{0,i}(\theta_i)$ for $i = 2, \dots, 16$, is a uniform distribution over the sensing region.

We use Algorithm 1 specified in Section VI for estimating θ. The MRF model we consider is specified by the pairwise graph $\mathcal{G}$ in Fig. 2 (blue edges). We use $L = 100$ points in (61) to represent the local belief densities. We select the sensor data time window length as $t = 10$ (starting at time 21 until 30 in the scenario in Fig. 2). We follow the steps in Algorithm 1 for $S = 16$ iterations.

A typical run is illustrated in Fig. 3. Here, the scatter plot of particles from marginal posteriors are given over iterations. Note that the network coordinate system is established by sensor 1 through its informative prior, and, LBP emanates this information towards the outer nodes while learning the edge potentials using node-wise separable likelihood evaluations $\psi_{i,j}(\theta_i^{(l)}, \theta_j^{(l)})$ given in (23) and (33).

For performance assesment, first, we consider the mean squared error for $\hat{\theta}$ output by our algorithm, as we have built our discussion on MMSE estimators in (12). We find this value empirically by taking the average of the squared norm of estimation errors over 100 Monte Carlo simulations. In Fig. 4, we present a semi-log plot of this quantity over iterations (blue line). Note that, convergence occurs in less
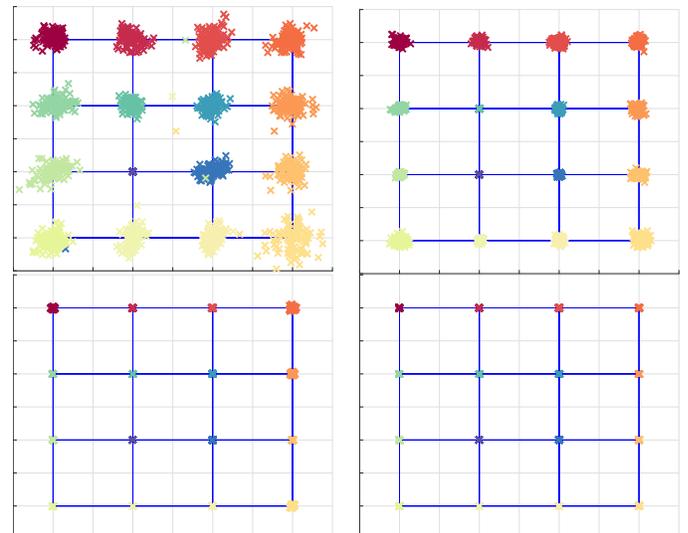

Fig. 3. Node beliefs in LBP iterations: Marginal posterior estimates after iteration 4 (upper-left), 6 (upper-right), 8 (lower-left), and, 10 (lower-right).
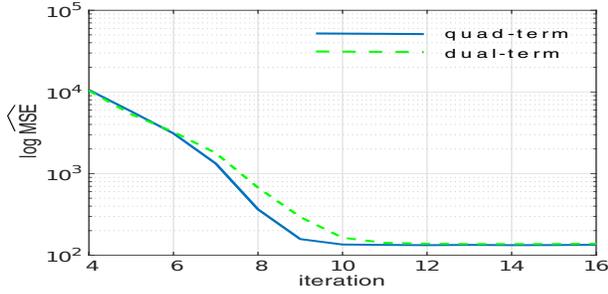
Fig. 4. Log-normalised error margin versus the iteration number n.

than ten iterations which is a favourable feature. We compare this algorithm with the RFS based dual-term pseudo-likelihood proposed in [26]. When evaluating this term, we use a Poisson multi-object model output by using the Gaussian mixture probability hypothesis density (GM-PHD) filter [52] with the LGSS model. The averaged MSE performance for the case that dual-term likelihoods are used as edge potentials is depicted with the green dashed line in Fig. 4. The quad-term approximation is seen to provide faster convergence with both pseudo-likelihoods leading to an on par accuracy in the steady regime, in this example. The edge update time for the quad-term update averages to 0.601 per edge per particle compared to 1.312 for the dual-term update demonstrating its relative efficiency.

Note that, the MSE is a network-wide term and the local error norms are smaller. The localisation miss-distance averaged over sensors is given in Fig. 5. The average error ($\pm$ one standard deviation) in the final step is $2.60 \pm 0.70m$ with a maximum value of 4.96 which is less than 0.5% of the edge distances of $1000m$. These results demonstrate that the proposed scheme is capable of providing self-localisation with favourable accuracy and small error margins.

VIII. Conclusions

In this work, we have addressed the prohibitive complexity of latent parameter estimation in state space models when there are many sensors collecting measurements. We proposed a pseudo-likelihood, namely the quad-term node-wise separable likelihood, as an accurate surrogate to the actual likelihood which is extremely costly to evaluate. The separable structure of this quad-term approximation makes it possible to evaluate
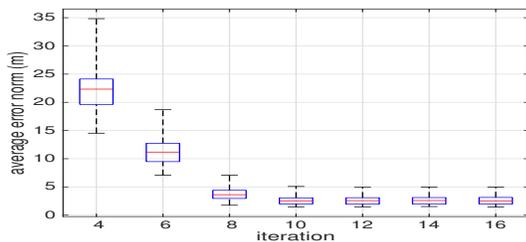

Fig. 5. Localisation miss-distance averaged over nodes versus the iteration number. 100 Monte Carlo runs displayed with the boxes centered at the median (red). Edges (blue) indicate the 25th and 75th percentiles.

it using local filtering operations, hence, scale with the number of sensors.

In order to use the proposed approximation in the case of multiple objects, we employed a parameterised multi-object state space model in which different configurations of the parameter specify different hypothesis of object-to-measurement and object-to-object associations. Specifically, we introduced an empirical Bayesian perspective for evaluating separable likelihoods in this model, using only local Bayesian filtering. This approach substitutes the estimates of the hypothesis variables when evaluating the likelihood of the latent parameter, and, decouples these two inference tasks if the hypothesis variables can be estimated locally. Therefore, it can be extended to general hypothesis variables that capture, for example, variable number of objects M_k, less than one probability of detection, i.e., $P_D < 1$, and, measurements with false alarm, by adapting the assignment problems involved accordingly (see, for example [35]).

The associated posterior distribution is a MRF over which distributed inference is possible using message passing algorithms such as LBP. We specified a particle message passing algorithm for sampling from latent parameter marginals for a linear Gaussian state space model with multiple objects. This algorithm is demonstrated in simulations for sensor self-localisation using point measurements from non-cooperative objects in a potentially GPS denying environment. It is possible to estimate unkown orientation angles as well, by appropriately defining the transform in (37) and taking into account the rotations implied by θ_is in the LBP steps of Algorithm 1 (additional details can be found in [53]).

Appendix

A. Empirical Bayes parameter likelihood

The empirical Bayes parameter posterior follows from the decomposition of the posterior in (11) using the chain rule of probabilities as

$$p(\theta|\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N) = \qquad (66)$$
$$\sum_{\tau_{1:t}^1} \cdots \sum_{\tau_{1:t}^N} p(\theta|\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N, \tau_{1:t}^{1:N}) p(\tau_{1:t}^{1:N}|\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N).$$

The first term inside the summations is the parameter posterior conditioned on the association variables and the second term is similar to a prior with respect to the first term. The fact that this term is conditioned on the measurements makes an empirical selection possible as discussed in Section II-B. Let us use a similar empirical prior selection approach as used in (16), i.e.,

$$p(\tau_{1:t}^{1:N}|\mathbf{Z}_{1:t}^1, \ldots, \mathbf{Z}_{1:t}^N) = \prod_{k=1}^{t} p(\tau_k^{1:N}|\mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N)$$
$$p(\tau_k^{1:N}|\mathbf{Z}_k^1, \ldots, \mathbf{Z}_k^N) \leftarrow \delta_{\bar{\tau}_{k-1}^{1:N}}(\tau_{k-1}^{1:N}). \qquad (67)$$

After substituting from (67) in (66), the parameter posterior is found as

$$p(\theta|\mathbf{Z}^1_{1:t}, \dots, \mathbf{Z}^N_{1:t})$$
$$= p(\theta|\mathbf{Z}^1_{1:t}, \dots, \mathbf{Z}^N_{1:t}, \tau^{1:N}_{1:t} = \bar{\tau}^{1:N}_{1:t})$$
$$\propto l(\mathbf{Z}^1_{1:t}, \dots, \mathbf{Z}^N_{1:t}|\theta, \tau^{1:N}_{1:t} = \bar{\tau}^{1:N}_{1:t})p(\theta)$$

where the likelihood in the last line is given by (17) and (18) in Section II-B.

B. Proof of Proposition 4.1

Proof. Let us expand the KLD term in (31) by substituting its arguments given in (9), (29) and (30):

$$D\left(p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta)||q(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta)\right)$$

$$= \int d\mathbf{Z}^i_{1:k} d\mathbf{Z}^j_{1:k} d\theta\, p(\mathbf{Z}^i_{1:k}, \mathbf{Z}^j_{1:k}, \theta)$$

$$\times \log \frac{p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta)}{q(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta)}$$

$$= \int d\mathbf{Z}^i_{1:k} d\mathbf{Z}^j_{1:k} d\theta\, p(\mathbf{Z}^i_{1:k}, \mathbf{Z}^j_{1:k}, \theta)$$

$$\times \frac{1}{2}\left(\log \frac{p(\mathbf{Z}^i_k, \mathbf{Z}^j_k, \mathbf{Z}^i_{1:k-1}|\mathbf{Z}^j_{1:k-1}, \theta)}{p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^j_{1:k-1}, \theta)p(\mathbf{Z}^i_{1:k-1}|\mathbf{Z}^j_{1:k-1}, \theta)}\right.$$

$$+ \log \frac{p(\mathbf{Z}^i_k, \mathbf{Z}^j_k, \mathbf{Z}^j_{1:k-1}|\mathbf{Z}^i_{1:k-1}, \theta)}{p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \theta)p(\mathbf{Z}^j_{1:k-1}|\mathbf{Z}^i_{1:k-1}, \theta)}$$

$$\left.+ 2\log \kappa_k(\theta)\right) \tag{68}$$

$$= \frac{1}{2}\left(I(\mathcal{Z}^j_k, \mathcal{Z}^i_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta)\right.$$

$$\left.+ I(\mathcal{Z}^j_k, \mathcal{Z}^i_k; \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k-1}, \Theta)\right) + E\{\log \kappa_k(\theta)\}. \tag{69}$$

In the equations above, $\kappa_k(\theta)$ is a normalisation constant given by (30). Eq.(68) is obtained after multiplying both the numerator and the denominator of the quotient inside the logarithm by $p(\mathbf{Z}^i_{1:k-1}|\mathbf{Z}^j_{1:k-1}, \theta)p(\mathbf{Z}^j_{1:k-1}|\mathbf{Z}^i_{1:k-1}, \theta)$ and a rearrangement of the terms. The definition of MI [31] results with the first two terms in Eq.(69). The last term is the expectation of the normalisation constant over the joint distribution of the sensor histories $\mathcal{Z}^i_{1:k-1}$ and $\mathcal{Z}^j_{1:k-1}$, and, Θ.

Let us now consider the normalisation constant:

$$\kappa_k(\theta) = \int d\mathbf{Z}^i_k d\mathbf{Z}^j_k \left(p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^j_{1:k-1}, \theta)p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \theta)\right)^{1/2}$$

$$\leqslant \left(\int d\mathbf{Z}^i_k d\mathbf{Z}^j_k p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^j_{1:k-1}, \theta)\right)^{1/2}$$

$$\times \left(\int d\mathbf{Z}^i_k d\mathbf{Z}^j_k p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \theta)\right)^{1/2}$$

$$= 1.$$

The inequality (70) follows from Hölder's Inequality. Consequently, the last term in (69) is non-positive, and, (31) is obtained. ∎

C. Proof of Corollary 4.2

Proof. We apply the chain rule of information to the MI terms on the RHS of (31) leading to

$$I(\mathcal{Z}^i_k, \mathcal{Z}^j_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta) =$$
$$I(\mathcal{Z}^j_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta) + I(\mathcal{Z}^i_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k}, \Theta), \tag{71}$$

and,

$$I(\mathcal{Z}^i_k, \mathcal{Z}^j_k; \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k-1}, \Theta) =$$
$$I(\mathcal{Z}^i_k; \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k-1}, \Theta) + I(\mathcal{Z}^j_k; \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k}, \Theta). \tag{72}$$

The MI terms on the RHSs of the equations above are for random variables which form Markov chains with the current the state variable $\mathcal{X}_k$. Consider the (conditional) chains $\mathcal{Z}^j_k \leftrightarrow \mathcal{X}_k \leftrightarrow \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta$ and $\mathcal{Z}^i_k \leftrightarrow \mathcal{X}_k \leftrightarrow \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k-1}, \Theta$ for the RHS of Eq.(71). The Data Processing Inequality [31] applied to these terms lead to

$$I(\mathcal{Z}^i_k, \mathcal{Z}^j_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta)$$
$$\leqslant I(\mathcal{X}_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta) + I(\mathcal{X}_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_k, \mathcal{Z}^j_{1:k-1}, \Theta)$$
$$= H(\mathcal{X}_k|\mathcal{Z}^j_{1:k-1}, \Theta) - H(\mathcal{X}_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta) \tag{73}$$
$$+ H(\mathcal{X}_k|\mathcal{Z}^j_k, \mathcal{Z}^j_{1:k-1}, \Theta) - H(\mathcal{X}_k|\mathcal{Z}^j_k, \mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta)$$

A similar break down of Eq.(72) results with

$$I(\mathcal{Z}^i_k, \mathcal{Z}^j_k; \mathcal{Z}^j_{1:k-1}|\mathcal{Z}^i_{1:k-1}, \Theta)$$
$$\leqslant H(\mathcal{X}_k|\mathcal{Z}^i_{1:k-1}, \Theta) - H(\mathcal{X}_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta) \tag{74}$$
$$+ H(\mathcal{X}_k|\mathcal{Z}^i_k, \mathcal{Z}^i_{1:k-1}, \Theta) - H(\mathcal{X}_k|\mathcal{Z}^i_k, \mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta).$$

Substituting from (73) and (74) into (31) results with (32). ∎

D. Comparison of the quad-term and dual-term updates

Let us compare the KLDs of the quad-term and dual-term updates. The KLD of the dual-term in (36) is given by [26]

$$D(p(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta)||u(\mathbf{Z}^i_k, \mathbf{Z}^j_k|\mathbf{Z}^i_{1:k-1}, \mathbf{Z}^j_{1:k-1}, \theta))$$
$$= I(\mathcal{Z}^j_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta) + I(\mathcal{Z}^i_k; \mathcal{Z}^i_{1:k-1}|\mathcal{Z}^j_{1:k-1}, \Theta)$$
$$+ I(\mathcal{Z}^i_k; \mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta). \tag{75}$$

The MI terms on the RHS of the above equation can be expanded using that $I(\mathcal{A}; \mathcal{B}|\mathcal{C}) = H(\mathcal{A}|\mathcal{C}) - H(\mathcal{A}|\mathcal{C}, \mathcal{B})$ [31]. The third term can be expanded in alternative ways as follows:

$$I(\mathcal{Z}^i_k; \mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta)$$
$$\overset{(a)}{=} H(\mathcal{Z}^i_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta) - H(\mathcal{Z}^i_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k}, \Theta) \tag{76}$$
$$\overset{(b)}{=} H(\mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta) - H(\mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^i_{1:k}, \Theta) \tag{77}$$

After decomposing the first two terms on the RHS of (75) similarly and adding to the average of (76) and (77) (which equals to the third term), we obtain

$$D(p||u) = \Delta u^+ - \Delta u^- \tag{78}$$
$$\Delta u^+ \triangleq H(\mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \Theta) + H(\mathcal{Z}^i_k|\mathcal{Z}^j_{1:k-1}, \Theta)$$
$$\Delta u^- \triangleq -\frac{1}{2}\left[H(\mathcal{Z}^i_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta)\right.$$
$$+ H(\mathcal{Z}^j_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k-1}, \Theta)$$
$$\left.+ H(\mathcal{Z}^i_k|\mathcal{Z}^i_{1:k-1}, \mathcal{Z}^j_{1:k}, \Theta) + H(\mathcal{Z}^j_k|\mathcal{Z}^i_{1:k}, \mathcal{Z}^j_{1:k-1}, \Theta)\right]$$

Now, let us consider (31). After substituting (71) and (72) in the RHS and expanding the MI terms as above, one obtains

$$D(p||q) \leqslant \Delta q^+ - \Delta q^- \tag{79}$$
$$\Delta q^+ \triangleq \frac{1}{2}\left[H(\mathcal{Z}^j_k|\mathcal{Z}^j_{1:k-1}, \Theta) + H(\mathcal{Z}^i_k|\mathcal{Z}^i_{1:k-1}, \Theta)\right.$$

$$+H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k}^i,\Theta)+H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k}^j,\Theta)\Big]$$

Now, let us compare (79) and (78): The negative weighted terms are equal, so, the difference of the positive weighted terms are considered:

$$\Delta u^+ - \Delta q^+ = H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k-1}^i,\Theta) - H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k-1}^j,\Theta) \quad (80)$$
$$+H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k-1}^j,\Theta) - H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k-1}^i,\Theta) + \epsilon$$
$$\epsilon \triangleq H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k-1}^i,\Theta) - H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k}^i,\Theta)$$
$$+H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k-1}^j,\Theta) - H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k}^j,\Theta)$$

Properties of differential entropy [31] suggest that $\epsilon \geqslant 0$ regardless of the problem setting (as conditioning reduces entropy). Under normal sensing conditions it is reasonable to expect that $\epsilon > 0$ holds as sensor measurements are highly informative on all the variables at time k. For sensors with similar sensing capabilities, it is also reasonable to expect that their measurement history are interchangeable. In other words,

$$H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k-1}^j,\Theta)+H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k-1}^i,\Theta)=$$
$$H(\mathcal{Z}_k^j|\mathcal{Z}_{1:k-1}^j,\Theta)+H(\mathcal{Z}_k^i|\mathcal{Z}_{1:k-1}^i,\Theta) \quad (81)$$

holds in which case (80) is nonzero and $D(p||u) > D(p||q)$. This condition can be relaxed for a difference of ϵ between the RHS and LHS of (81).

Comparison in terms of entropy upper bounds is more straightforward. Let us consider (32) and Corollary 4.2 in [26] which we repeat here for convenience:

$$D(p||u) \leqslant H_u^+ - H_u^-, \quad (82)$$
$$H_u^+ = H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^j,\Theta) + H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^i,\Theta),$$
$$H_u^- = H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^i,\mathcal{Z}_{1:k-1}^j,\Theta) +$$
$$\max\{H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^i,\mathcal{Z}_{1:k}^j,\Theta), H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^j,\mathcal{Z}_{1:k}^i,\Theta)\}.$$

For sensors of identical capabilities, both terms in the maximisation should be identical as the accuracy of the state estimate should not differ for using either of the sensors' current measurement in addition to the histories of both. As a result, we can replace H_u^- with

$$H_u^- = H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^i,\mathcal{Z}_{1:k-1}^j,\Theta)+$$
$$\frac{1}{2}\left(H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^i,\mathcal{Z}_{1:k}^j,\Theta) + H(\mathcal{X}_k|\mathcal{Z}_{1:k-1}^j,\mathcal{Z}_{1:k}^i,\Theta)\right).$$

Now, note that the sum of the positive weighted terms in the RHS of (32) (let us denote by H_q^+) is smaller than H_u^+ owing to that conditioning reduces entropy [31]. The sum of the negative weighted terms (let us denote by H_q^-), for the case ,equals to H_u^-, leading to

$$H_u^+ - H_u^- - (H_q^+ - H_q^-) > 0,$$

which indicates that the entropy bound of the quad-term update is smaller than that for the dual-term update.

E. Local ML estimate of object correspondances

The semantic of the object correspondance γ_k is slightly different from that of the random permutation variables used in identity management [45] in a way closer to data association

because local (track) identities are synonymous with the measurements they are associated with instead of signal features etc. Therefore, in our model, γ_k is uniquely defined when given τ_k^i and τ_k^j, and, $\gamma_1,\ldots,\gamma_k$ are mutually independent. The ML estimate of γ_k using the likelihood for data set $(\mathbf{Z}_k^i,\mathbf{Z}_{1:k}^j)$ is also conditioned on the association configuration $(\hat{\tau}_k^i,\hat{\tau}_k^i)$ and θ is given by

$$\hat{\gamma}_k = \arg \max_{\gamma_k \in \mathcal{S}_M} \log l(\mathbf{Z}_k^i,\mathbf{Z}_{1:k}^j|\gamma_k,\tau_k^i=\hat{\tau}_k^i,\tau_k^j=\hat{\tau}_k^j,\theta) \quad (83)$$

where $\mathcal{S}_M$ is the set of M-permutations. This likelihood can be decomposed using the chain rule of probability as

$$l(\mathbf{Z}_k^i,\mathbf{Z}_{1:k}^j|\gamma_k,\tau_k^i,\tau_k^j,\theta)$$
$$= p(\mathbf{Z}_k^i,\mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^j,\gamma_k,\tau_k^i,\tau_k^j,\theta)p(\mathbf{Z}_{1:k-1}^j|\gamma_k,\tau_k^i,\tau_k^j,\theta),$$
$$= p(\mathbf{Z}_k^i,\mathbf{Z}_k^j|\mathbf{Z}_{1:k-1}^j,\gamma_k,\tau_k^i,\tau_k^j,\theta)p(\mathbf{Z}_{1:k-1}^j|\theta), \quad (84)$$

where the second equality follows from the independence of sensor j's measurements up to time $k-1$ from the association variables at time k. The first term on the RHS is easily identified as (43) evaluated for γ_k when $m' = \gamma_k(\tau_k^i(o))$. This term, when subsituted in (83), leads to a 2-D assignment problem given by

$$\hat{\gamma}_k = \arg \max_{\gamma_k \in \mathcal{S}_M} \sum_{o=1}^{M} d(o, m' = \gamma_k(\hat{\tau}_k^i(o)))$$
$$d(o, m') \triangleq \log r_{ij,o}^k(z_{k,o}^i,\theta),$$

for $o, m' = 1, \ldots, M$, where $r_{ij,o}^k$ is given in (43).

After finding the M^2 costs above, this problem can be solved using the auction algorithm [48] in polynomial time with M. This algorithm finds the M pairs corresponding to the best permutation $\hat{\gamma}_k$. We use a similar approach for the data association problem in Bayesian filtering in Section VI-A.

ACKNOWLEDGMENT

REFERENCES

[1] S. Sarkka, A. Solin, and J. Hartikainen, "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, July 2013.

[2] B.-N. Vo, M. Mallick, Y. Bar-Shalom, S. Coraluppi, R. Osborne, R. Mahler, B.-T. Vo, and J.G. Webster, *Wiley Encyclopedia of Electrical and Electronics Eng.*, Multitarget Tracking, J. Wiley & Sons, Inc., 2015.

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

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

[5] S. S. Singh, N. Whiteley, and S. J. Godsill, "Approximate likelihood estimation of static parameters in multi-target models," in *Bayesian Time Series Models*, D. Barber, A. T. Cemgil, and S. Chiappa, Eds., pp. 225–244. Cambridge University Press, 2011.

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

[7] 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.

[8] M. Erol-Kantarci, H.T. Mouftah, and S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 3, pp. 487–502, Third Quarter 2011.

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

[10] M. Uney, B. Mulgrew, and D. Clark, "Target aided online sensor localisation in bearing only clusters," in *SSPD'14*, Sept 2014, pp. 1–5.

[11] D. Devarajan, Z. Cheng, and R.J. Radke, "Calibrating distributed camera networks," *Proc. of the IEEE*, vol. 96, no. 10, pp. 1625–1639, 2008.

[12] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, "On particle methods for parameter estimation in state-space models," *Statistical Science*, vol. 30, no. 3, pp. 328–351, 08 2015.

[13] L. Chen, M.J. Wainwright, M. Çetin, and A.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] L. Jiang, S. S. Singh, and S.Yildirim, "Bayesian tracking and parameter learning for non-linear multiple target tracking models," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5733–5745, Nov 2015.

[15] C. Andrieu, A. Doucet, and R. 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.

[16] J. Ala-Luhtala, N. Whiteley, K. Heine, and R. Piche, "An introduction to twisted particle filters and parameter estimation in non-linear state-space models," *IEEE Trans. on Signal Proc.*, vol. PP, no. 99, pp. 1–1, 2016.

[17] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, 2012.

[18] C. Varin, N. Reid, and D. Firth, "An overview of composite likelihood methods," *Statistica Sinica*, vol. 21, pp. 5–42, 2011.

[19] E. Ehrlich, A. Jasra, and N. Kantas, "Gradient free parameter estimation for hidden Markov models with intractable likelihoods," *Methodology and Computing in Applied Prob.*, vol. 17, no. 2, pp. 315–349, 2015.

[20] S. Yildirim, S. S. Singh, T. Dean, and A. Jasra, "Parameter estimation in hidden Markov models with intractable likelihoods using sequential Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 24, no. 3, pp. 846–865, 2015.

[21] C. Andrieu, A. Doucet, and V. B. Tadic, "On-line parameter estimation in general state-space models," in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec 2005, pp. 332–337.

[22] 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.

[23] Q. Liu and A. 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.

[24] Y. D. Mizrahi, M. Denil, and N. D. Freitas, "Distributed parameter estimation in probabilistic graphical models," in *Advances in NIPS 27*, Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, Eds., pp. 1700–1708. Curran Assoc., Inc., 2014.

[25] M. Uney, B. Mulgrew, and D. Clark, "Cooperative sensor localisation in distributed fusion networks by exploiting non-cooperative targets," in *IEEE Workshop on Stat. Signal Proc. (SSP)'14*, June 2014, pp. 516–519.

[26] 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.

[27] S. Yildirim, L. Jiang, S. S. Singh, and T. A. Dean, "Calibrating the Gaussian multi-target tracking model," *Statistics and Computing*, vol. 25, no. 3, pp. 595–608, May 2015.

[28] B. P. Carlin and T. A. Louis, *Bayes and empirical Bayes methods for data analysis*, Chapman and Hall/CRC, 2nd edition, 2000.

[29] D. Hall, C.-Y. Chong, J. Llinas, and M. Liggins II, Eds., *Distributed Data Fusion for Network-Centric Operations*, CRC Press, 2013.

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

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[32] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," *Commun. ACM*, vol. 53, no. 10, pp. 95–103, Oct. 2010.

[33] 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.

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

[35] A. B. Poore and S. 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 Applications.

[36] B. Ristic, Arulampalam S., and N. Gordon, *Beyond the Kalman Filter; particle filters for tracking applications*, Artech House, 2004.

[37] J. Kokkala and S. Sarkka, "Combining particle MCMC with Rao-Blackwellized Monte Carlo data association for parameter estimation in multiple target tracking," *Digital Signal Processing*, vol. 47, pp. 84 – 95, 2015, Special Issue in Honour of W. J. (Bill) Fitzgerald.

[38] I. Schlangen, D. E. Clark, and E. D. Delande, "Single-cluster PHD filter methods for joint multi-object filtering and parameter estimation," *arXiv preprint arXiv:1705.05312*, 2017.

[39] 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.

[40] 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.

[41] 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.

[42] M.J. Wainwright, T.S. Jaakkola, and A.S. Willsky, "Tree-based reparameterization framework for analysis of sum-product and related algorithms," *IEEE Trans. on Info. Theory*, vol. 49, no. 5, pp. 1120–1146, May 2003.

[43] J. S. Yedidia, W.T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. on Info. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[44] M. Üney, B. Mulgrew, and D. Clark, "Distributed localisation of sensors with partially overlapping field-of-views in fusion networks," in *2016 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1340–1347.

[45] L. J. Guibas, "The identity management problem- a short survey," in *11th Int. Conf. on Info. Fusion (Fusion'08)*, June 2008, pp. 1–7.

[46] J. Huang, C. Guestrin, and L. Guibas, "Fourier theoretic probabilistic inference over permutations," *Journal of Machine Learning Research*, vol. 10, pp. 997–1070, May 2009.

[47] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems,*, Society for Industrial and Applied Mathematics, 2012.

[48] 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.

[49] G. Casella and C. P. Robert, *Monte Carlo Statistical Methods*, Springer, second edition, 2005.

[50] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.

[51] 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.

[52] B. N. Vo and W. K. Ma, "The gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov 2006.

[53] M. Üney, K. Copsey, S. Page, B. Mulgrew, and P. Thomas, "Enabling self-configuration of fusion networks via scalable opportunistic sensor calibration," in *Proceedings of the SPIE Defence+Security Symposium, Signal Processing, Sensor/Information Fusion, and Target Recognition XXVII*, 2018, to appear.