Distributed Approximation of Joint Measurement Distributions Using Mixtures of Gaussians

Brian J. Julian*†, Stephen L. Smith‡, and Daniela Rus*

***Abstract*—This paper presents an approach to distributively approximate the continuous probability distribution that describes the fusion of sensor measurements from many networked robots. Each robot forms a weighted mixture of Gaussians to represent the measurement distribution of its local observation. From this mixture set, the robot then draws samples of Gaussian elements to enable the use of a consensus-based algorithm that evolves the corresponding canonical parameters. We show that the these evolved parameters describe a distribution that converges weakly to the joint of all the robots' unweighted mixture distributions, which itself converges weakly to the joint measurement distribution as more system resources are allocated. The major innovation of this approach is to combine sample-based sensor fusion with the notion of pre-convergence termination that results in scalable multi-robot system. We also derive bounds and convergence rates for the approximated joint measurement distribution, specifically the elements of its information vectors and the eigenvalues of its information matrices. Most importantly, these performance guarantees do not come at a cost of complexity, since computational and communication complexity scales quadratically with respect to the Gaussian dimension, linearly with respect to the number of samples, and constant with respect to the number of robots. Results from numerical simulations for object localization are discussed using both Gaussians and mixtures of Gaussians.**

I. Introduction

We wish to develop scalable approaches to state estimation tasks such as tracking, surveillance, and exploration using large teams of autonomous robots equipped with sensors. Consider the task of using many aerial robots to monitor the flow of objects into and out of a major seaport (e.g., ships, containers, ground vehicles). To collectively estimate the objects' positions, one approach is to wirelessly communicate all sensor measurements to a data fusion center, perform the estimation calculations in a centralized manner, and then globally broadcast the results to enable the robots to better position their sensors. For large systems, the central processor quickly becomes a computational and communication bottleneck, and thus is not considered to be scalable [4].

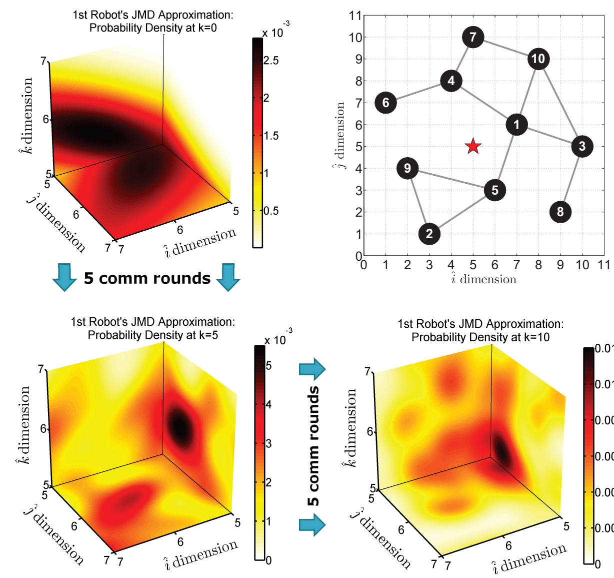

Fig. 1: This figure shows the distributed approximation of the joint measurement distribution using mixtures of Gaussians. *Top Right:* Ten robots (black circles) are located in three dimensional space with their $\hat{\imath}$ and $\hat{\jmath}$ positions shown and their $\hat{k}$ positions equal to their corresponding indices. The robots use range only sensors and communicate on an undirected graph to locate an object of interest (red star). *Top Left:* Each robot forms an unweighted mixture of 1000 Gaussians to represent its independent measurement distribution of the object's position given its local observation. Here we show the 1st robot's mixture distribution on two dimensional slices that intersect at the object's true location of $(5, 5, 5)$ meters. *Bottom:* The 1st robot's approximation of the joint measurement distribution becomes more accurate and precise as more consensus iterations are performed. Note that the color scales vary among the plots to highlight the distribution's structure.

We propose to solve this problem in a purely decentralized manner with the following approach. Each robot (i) independently makes a local observation using its imperfect sensors, (ii) represents the corresponding probability distribution with a weighted mixture of Gaussians, (iii) draws samples from this mixture set to form an unweighted mixture set, and lastly (iv) runs a consensus-based algorithm to approximate the distribution describing the joint of all robots' observations (Figure 1). This approximation can then be used in a sequential

*Brian J. Julian and Daniela Rus are with the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139, USA, bjulian@mit.edu and rus@csail.mit.edu

†Brian J. Julian is also with MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420, USA

‡Stephen L. Smith is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada, stephen.smith@uwaterloo.ca

This work is sponsored by the Department of the Air Force under Air Force contract number FA8721-05-C-0002. The opinions, interpretations, recommendations, and conclusions are those of the authors and are not necessarily endorsed by the United States Government. This work is also supported in part by ARO grant number W911NF-05-1-0219, ONR grant number N00014-09-1-1051, NSF grant number EFRI-0735953, ARL grant number W911NF-08-2-0004, and MIT Lincoln Laboratory, and the Boeing Company.

Bayesian filter to update the robots' belief of the continuous state of a finite extent of the world. Building on our prior work [9] that only considered a discrete set of probabilities, each robot uses the consensus-based algorithm to evolve its representation of the *independent measurement distribution* into an approximation for the *joint measurement distribution*. This approach allows for resource adaptive state estimation, for which the computation and communication complexities do not depend on the number of robots.

We prove for all robots on a static, connected, undirected communication graph that the approximation of the joint measurement distribution converges weakly[1] to the joint of all the robots' unweighted mixture distributions. The given restrictions on the graph are used to derive bounds and convergence rates for the approximated joint measurement distribution, specifically elements of its information vectors and eigenvalues of its information matrices. Yet, the implementation works on arbitrary networks without risk of catastrophic failures (e.g., robustness against robot failures), and without restriction on the number of communication rounds that the robots need to use for the consensus-based algorithm. An extremely attractive aspect of the approach is that expected performance provably improves as more system resources are allocated. We believe these theoretical contributions can drive the development of application specific sensor fusion algorithms that are unbiased, convergent, and scalable.

We have been inspired by over two decades worth of advancements in distributed estimation algorithms. An approach to compute locally optimal estimators from many independent sensor measurements at a central fusion center was described in detail by Gubner [6]. Concerning decentralization, the early work of Durrant-Whyte et al. [4] with decentralized Kalman filters laid the basis for the Decentralized Data Fusion architecture [10]. Extensions incorporating consensus averaging algorithms [3], [15] have been used for maximum-likelihood parameter estimation [16], maximum a-posteriori estimation [11], and distributed Kalman filtering [1], [17].

One of the most relevant works in distributed Kalman filtering is by Ren et al. [13], who showed the convergence of a filter incorporating information-based states. The proof of convergence for each Gaussian element in the joint distribution approximation closely follows in our work, even though our approach applies to a larger class of Bayes filters (e.g., map merging [2]). This generality is shared by the approach of Fraser et al. [5] using hyperparameters. However, our approach enables the early termination of the consensus-based algorithm without the risk of *double-counting* any single observation, even when the maximum in/out degree and the number of robots are unknown.

This paper is organized as follows. In Section II we formalize the problem of distributively approximating the joint measurement distribution within a multi-robot system, then discuss the use of a consensus-based algorithm to calculate products of Gaussian distributions in Section III. Also in Section III are our main results on the convergence of distributively formed Gaussian mixtures to representations of the joint measurement distribution. In Section IV we derive bounds and convergence rates for the elements of the joint mixture set, followed by numerical simulations in Section V to illustrate these performance guarantees and motivate the approach.

II. Problem Formulation

A. General Setup

Consider a system of robots, where each robot has a belief of the continuous-valued state concerning the same finite extent of the world. We model the world state[2] as a random variable, X, that takes values from a continuous alphabet, $\mathcal{X}$. Each robot cannot perfectly measure the world state, but instead makes an observation with its sensors that are influenced by noise. The robots' synchronous observations together form a joint observation, which we also model as a random variable, Y. Sensing may be interpreted as using a noisy channel, and thus the relationship between the true world state and the noisy observation is described by the joint measurement distribution (JMD), $\mathbb{P}(Y = y|X)$, where y is the value the joint observation takes.

Our goal is to enable each robot to independently perform the posterior calculations,

$$\mathbb{P}^{[i]}(X|Y = y) = \frac{\mathbb{P}^{[i]}(X)\mathbb{P}(Y{=}y|X)}{\int_{x\in\mathcal{X}} \mathbb{P}^{[i]}(X{=}x)\mathbb{P}(Y{=}y|X{=}x)dx}, \qquad (1)$$

needed for sequential Bayesian filter predictions and mutual information gradient-based control [9], where $\mathbb{P}^{[i]}(X)$ is the prior distribution that describes the ith robot's belief of the world state. Since the sensors of any two robots are physically detached from one another, we assume that the errors on the observations are uncorrelated between robots. In other words, a random variable that describes the ith robot's observation, $Y^{[i]}$, is conditionally independent of any other random variable that describes another robot's observation, $Y^{[v]}$ with $v \neq i$, given the world state. This assumption for a system of n_r robots gives a JMD of $\mathbb{P}(Y = y|X) = \prod_{i=1}^{n_r} \mathbb{P}(Y^{[i]} = y^{[i]}|X)$ when we model the joint observation as an n_r-tuple random variable, $Y = (Y^{[1]}, \ldots, Y^{[n_r]})$, where $y^{[i]}$ is the value that the ith robot's observation takes and $\mathbb{P}(Y^{[i]} = y^{[i]}|X)$ is the ith robot's independent measurement distribution (IMD).

Thus, the posterior calculations from (1) become

$$\mathbb{P}^{[i]}(X|Y = y) = \frac{\mathbb{P}^{[i]}(X)\prod_{v=1}^{n_r}\mathbb{P}(Y^{[v]}{=}y^{[v]}|X)}{\int_{x\in\mathcal{X}} \mathbb{P}^{[i]}(X{=}x)\prod_{v=1}^{n_r}\mathbb{P}(Y^{[v]}{=}y^{[v]}|X)dx}, \qquad (2)$$

and our goal of approximating the JMD over all possible continuous-valued world states becomes equivalent to approximating the product of all the IMDs. It is this equivalence that enables our distributed algorithm to reasonably approximate

[1]Also known as convergence in distribution.

[2]We use the term world state as shorthand for state of the finite extent of the world.

arbitrary continuous distributions using independently formed mixtures of multivariate Gaussian distributions. By reasonably we mean that the approximation has the performance characteristics discussed in Section IV, such as convergence to the true distribution as certain system resources increase (e.g., computational capacity, network bandwidth, etc.).

B. Decentralized System

Let the n_r robots communicate according to an undirected communication graph, $\mathcal{G}$, with a corresponding unordered edge set, $\mathcal{E}$; that is, $\{i, v\} \in \mathcal{E}$ if the ith and vth robots are neighbors. We use $N^{[i]}$ to denote the set of neighbors of the ith robot, which has an in/out degree of $|N^{[i]}|$. In addition, we consider the corresponding Metropolis-Hastings weight matrix [15], $\mathbf{W}$, which has the form

$$[\mathbf{W}]_{iv} = \begin{cases} 1 - \sum_{v' \in N^{[i]}} \frac{1}{\max\{|N^{[i]}|, |N^{[v']}|\}+1}, & i = v, \\ \frac{1}{\max\{|N^{[i]}|, |N^{[v]}|\}+1}, & \{i, v\} \in \mathcal{E}, \\ 0, & \text{otherwise}, \end{cases}$$

where $[\cdot]_{iv}$ denotes the matrix entry (i, v). For vectors, $[\cdot]_m$ denotes the mth row entry.

It was previously shown [8] that by initializing the consensus states $\psi_k^{[i]}$ and $\pi_k^{[i]}$ to a basis vector and a probability vector, respectively, each robot can run the discrete-time average consensus algorithm

$$\psi_{k+1}^{[i]} = [\mathbf{W}]_{ii}\psi_k^{[i]} + \sum_{v \in N^{[i]}} [\mathbf{W}]_{iv}\psi_k^{[v]} \quad (3)$$

in parallel with its exponential form

$$\pi_{k+1}^{[i]} = (\pi_k^{[i]})^{[\mathbf{W}]_{ii}} \prod_{v \in N^{[i]}} (\pi_k^{[v]})^{[\mathbf{W}]_{iv}} \quad (4)$$

to distributively approximate a finite set of joint measurement probabilities, where $k \in \mathbb{Z}_{\geq 0}$ denotes the communication round and the above algorithms are understood to be element-wise. In this paper, we extend this approach to multivariate Gaussian distributions, then show that this extension supports the approximations of arbitrary JMDs using mixtures of Gaussians[3]. Note that many other algorithms of this form yielding asymptotic average consensus are also appropriate (see, e.g., [11]).

III. Distributed Approximations

A. Products of Gaussians

Consider for all robots $i \in \{1, \ldots, n_r\}$ the non-degenerate[4] n_g-dimensional Gaussians, $\mathcal{N}_c(\xi_0^{[i]}, \mathbf{\Omega}_0^{[i]}) \propto \exp\big(-\frac{1}{2}x^T\mathbf{\Omega}_0^{[i]}x + x^T\xi_0^{[i]}\big)$, where $\xi_0^{[i]} \in \mathbb{R}^{n_g}$ and $\mathbf{\Omega}_0^{[i]} \in \mathbb{R}^{n_g \times n_g}$ are the ith robot's information vector and information matrix, respectively. If these Gaussians represent the robots' IMDs, from (2) the JMD for the system is simply

$$\eta \prod_{i=1}^{n_r} \mathcal{N}_c(\xi_0^{[i]}, \mathbf{\Omega}_0^{[i]}) = \mathcal{N}_c(\xi, \mathbf{\Omega}), \quad (5)$$

[3]We use the term Gaussians as shorthand for multivariate Gaussian distributions.

[4]By non-degenerate we mean that the information matrix of a Gaussian is a real positive-definite symmetric matrix.

where $\xi = \sum_{i=1}^{n_r} \xi_0^{[i]}$ is the joint information vector, $\mathbf{\Omega} = \sum_{i=1}^{n_r} \mathbf{\Omega}_0^{[i]}$ is the joint information matrix, and η is a normalizing constant.

For a given world state $x \in \mathcal{X}$, let $\pi_k^{[i]} \in \mathbb{R}_{\geq 0}$ and $\psi_k^{[i]} \in [0, 1]^{n_r}$ be initialized to $\mathcal{N}_c(\xi_0^{[i]}, \mathbf{\Omega}_0^{[i]})$ and $\mathbf{e}_i$, respectively, where $\mathbf{e}_i$ is the standard basis pointing in the ith direction in $\mathbb{R}^{n_r}$. On a connected graph $\mathcal{G}$, we can use (4) and (3) at each communication round to have $(\pi_k^{[i]})^{\beta_k^{[i]}}$ converge to $\prod_{i=1}^{n_r} \mathcal{N}_c(\xi_0^{[i]}, \mathbf{\Omega}_0^{[i]})$ in the limit as $k \to \infty$, where $\beta_k^{[i]} = \|\psi_k^{[i]}\|_\infty^{-1}$ is a scalar exponential factor that converges to n_r [8]. The expansion of $(\pi_k^{[i]})^{\beta_k^{[i]}}$ leads to the following.

Theorem 1 (Consensus of a Product of Gaussians)**.** *Let $\xi_k^{[i]} \in \mathbb{R}^{n_g}$ and $\mathbf{\Omega}_k^{[i]} \in \mathbb{R}^{n_g \times n_g}$ be initialized to $\xi_0^{[i]}$ and $\mathbf{\Omega}_0^{[i]}$, respectively, and have both evolve according to* (3) *on a connected graph $\mathcal{G}$. Then for all robots we have that*

$$\mathcal{N}_c(\beta_k^{[i]}\xi_k^{[i]}, \beta_k^{[i]}\mathbf{\Omega}_k^{[i]}) \to \mathcal{N}_c(\xi, \mathbf{\Omega}), \quad \forall x \in \mathbb{R}^{n_g} \quad (6)$$

as $k \to \infty$, or in other words, that $\mathcal{N}_c(\beta_k^{[i]}\xi_k^{[i]}, \beta_k^{[i]}\mathbf{\Omega}_k^{[i]})$ converges weakly to $\mathcal{N}_c(\xi, \mathbf{\Omega})$.

Proof (Theorem 1)**.** *We first note for all robots and communication rounds that $\pi_k^{[i]}$ is a product of values taken from (possibly unnormalized) Gaussians. Hence $(\pi_{k+1}^{[i]})^{\beta_{k+1}^{[i]}}$ is itself a value that is taken from an unnormalized Gaussian proportional to*

$$\prod_{v \in \{\{i\} \cup N^{[i]}\}} \exp\Big(\beta_{k+1}^{[i]}[\boldsymbol{W}]_{iv}\big(\tfrac{1}{2}x^T\mathbf{\Omega}_k^{[i]}x + x^T\xi_k^{[v]}\big)\Big)$$

which gives us the desired consensus update expressions for $\xi_{k+1}^{[i]}$ and $\mathbf{\Omega}_{k+1}^{[i]}$. Lastly, from [15] and [8] we have for every $x \in \mathcal{X}$ that $\pi_k^{[i]}$ and $\beta_k^{[i]}$ converge to $\prod_{i=1}^{n_r} \mathcal{N}_c(\xi_0^{[i]}, \mathbf{\Omega}_0^{[i]})^{1/n_r}$ and n_r, respectively. □

Remark 1 (Consensus of the Canonical Parameters)**.** *Even though $\pi_0^{[i]}$ is dependent on a given world state, we have that $\xi_0^{[i]}$ and $\mathbf{\Omega}_0^{[i]}$ are not. Thus,* (6) *implies that we can run our consensus-based algorithm on canonical parameters of sizes $O(n_g)$ and $O(n_g^2)$ to reasonably approximate a JMD of Gaussian form over all world states.*

B. Mixtures of Gaussians

As previously stated, our goal is to enable each robot to independently perform the posterior calculations (2) by distributively approximating the product of all the IMDs. With Theorem 1, we have presented sufficient machinery to enable these approximations if each IMD can be accurately represented by a single Gaussian. Performance guarantees for this particular case will be discussed in Section IV. For arbitrary continuous distributions, we now complete the approach using mixtures of Gaussians.

Let each robot form a weighted mixture set $\mathcal{M}^{[i]} := \{(w^{[i,j]}, \xi_0^{[i,j]}, \mathbf{\Omega}_0^{[i,j]}) : j \in \mathcal{I}(i)\}$ composed of triples that have scalar weights[5] $w^{[i,j]} \in [0,1]$, information vectors $\xi_0^{[i,j]} \in \mathbb{R}^{n_g}$, and information matrices $\mathbf{\Omega}_0^{[i,j]} \in \mathbb{R}^{n_g \times n_g}$. For simplicity we assume that the weighted summation of the corresponding n_g-dimensional Gaussians perfectly represents the IMD, but also note that an approximation of this form converges weakly[6] to the true IMD in the limit as the size of the weighted mixture set tends to infinity. Hence, the JMD is represented by the normalized product of these weighted summations across all robots,

$$\mathcal{N}(\mathcal{M}) := \eta \prod_{i=1}^{n_r} \sum_{j \in \mathcal{I}(i)} w^{[i,j]} \mathcal{N}_c(\xi_0^{[i,j]}, \mathbf{\Omega}_0^{[i,j]}).$$

Unfortunately, the computational complexity of computing the JMD scales exponentially with respect to the number of robots, and thus is intractable even for moderately sized systems. We now describe a technique that forms an unbiased[7] approximation of the JMD, for which computation is tractable and readily distributed among the robots. Let each robot draw n_m samples from its weighted mixture set with probabilities proportional to the corresponding weights. The ordered pairing of drawn information vectors and information matrices form an unweighted mixture set, $\check{\mathcal{M}}_0^{[i]} := \{(\check{\xi}_0^{[i,j]}, \check{\mathbf{\Omega}}_0^{[i,j]}) : j \in \{1, \ldots, n_m\}\}$, from which the normalized summation

$$\mathcal{N}(\check{\mathcal{M}}_0^{[i]}) := \sum_{j=1}^{n_m} \frac{\mathcal{N}_c(\check{\xi}_0^{[i,j]}, \check{\mathbf{\Omega}}_0^{[i,j]})}{n_m} \approx \mathbb{P}(Y^{[i]} = y^{[i]} | X)$$

approximates the robot's IMD. We then define the joint mixture set, $\check{\mathcal{M}}$, to be the unweighted set of canonical parameter pairs resulting from the product of the robots' unweighted independent mixture distributions having equal indices. More formally, we have $\check{\mathcal{M}} := \{(\check{\xi}^{[j]}, \check{\mathbf{\Omega}}^{[j]}) : j \in \{1, \ldots, n_m\}\}$, where $\check{\xi}^{[j]} = \sum_{i=1}^{n_r} \check{\xi}_0^{[i,j]}$ and $\check{\mathbf{\Omega}}^{[j]} = \sum_{i=1}^{n_r} \check{\mathbf{\Omega}}_0^{[i,j]}$. We are interested in each robot independently forming the joint mixture set to approximate the JMD, leading to the following.

Lemma 1 (Properties of the Joint Mixture Distribution)**.** *Define the joint mixture distribution to be the normalized summation of Gaussians formed from the joint mixture set,*

$$\mathcal{N}(\check{\mathcal{M}}) := \sum_{j=1}^{n_m} \frac{\mathcal{N}_c(\check{\xi}^{[j]}, \check{\mathbf{\Omega}}^{[j]})}{n_m}.$$

Then the joint mixture distribution is an unbiased approximation of the JMD that converges weakly as the number of samples n_m tends to infinity.

Proof (Lemma 1)**.** *We first prove that the joint mixture distribution converges weakly to the JMD. Consider any tuple $a \in \prod_{i=1}^{n_r} \mathcal{I}(i)$, where for each $i \in \{1, \ldots, n_r\}$ the ith entry is $[a]_i \in \mathcal{I}(i)$. Let us define $w^{[a]} := \prod_{i=1}^{n_r} w^{[i,[a]_i]}$. For a given unweighted sample $j \in \{1, \ldots, n_m\}$, let $\check{A}^{[j]} = (\check{A}^{[1,j]}, \ldots, \check{A}^{[n_r,j]})$ be an n_r-tuple random variable, for which each element $\check{A}^{[i,j]}$ takes samples from $\mathcal{I}(i)$ with probability $\mathbb{P}(\check{A}^{[i,j]} = [a]_i) = w^{[i,[a]_i]}$. Hence, the probability that the jth sample is a is given by*

$$\mathbb{P}(\check{A}^{[j]} = a) = \prod_{i=1}^{n_r} \mathbb{P}(\check{A}^{[i,j]} = [a]_i) = \prod_{i=1}^{n_r} w^{[i,a_{[i]}]} = w^{[a]}.$$

Next, let us define an indicator random variable $\chi_{\check{A}^{[j]}}$ for the event $\{\check{A}^{[j]} = a\}$. We have that

$$\begin{aligned}\mathcal{N}(\check{\mathcal{M}}) &= \sum_{j=1}^{n_m} \frac{\mathcal{N}_c(\check{\xi}^{[j]}, \check{\mathbf{\Omega}}^{[j]})}{n_m} \\ &= \eta \sum_{j=1}^{n_m} \sum_{a \in \prod_{i=1}^{n_r} \mathcal{I}(i)} \frac{\chi_{\check{A}^{[j]}}}{n_m} \prod_{i=1}^{n_r} \mathcal{N}_c(\xi_0^{[i,[a]_i]}, \mathbf{\Omega}_0^{[i,[a]_i]}), \qquad (7)\end{aligned}$$

where η is again a normalization constant.

By the Strong Law of Large Numbers [14], we have that

$$\lim_{n_m \to \infty} \sum_{j=1}^{n_m} \frac{\chi_{\check{A}^{[j]}}}{n_m} = \mathbb{E}(\chi_{\check{A}^{[j]}}) = w^{[a]},$$

with probability one. Therefore, exchanging the order of the summations in (7) and taking the limit as the number of samples tend to infinity, we have with probability one that

$$\begin{aligned}\lim_{n_m \to \infty} \mathcal{N}(\check{\mathcal{M}}) &= \eta \sum_{j=1}^{n_m} w^{[a]} \sum_{a \in \prod_{i=1}^{n_r} \mathcal{I}(i)} \mathcal{N}_c(\xi_0^{[i,a(i)]}, \mathbf{\Omega}_0^{[i,a(i)]}) \\ &= \eta \sum_{a \in \prod_{i=1}^{n_r} \mathcal{I}(i)} \prod_{i=1}^{n_r} w^{[i,[a]_i]} \mathcal{N}_c(\xi_0^{[i,[a]_i]}, \mathbf{\Omega}_0^{[i,[a]_i]}) \\ &= \eta \prod_{i=1}^{n_r} \sum_{j \in \mathcal{I}(i)} w^{[i,j]} \mathcal{N}_c(\xi_0^{[i,j]}, \mathbf{\Omega}_0^{[i,j]}) = \mathcal{N}(\mathcal{M}).\end{aligned}$$

We now prove that the joint mixture distribution is an unbiased approximation of the JMD. Let the following random variables take values according to the corresponding distributions:

$$\begin{aligned}\check{Z} &\sim \mathcal{N}(\check{\mathcal{M}}), \\ \check{Z}^{[j]} &\sim \mathcal{N}_c(\check{\xi}^{[j]}, \check{\mathbf{\Omega}}^{[j]}), \\ Z^{[a]} &\sim \eta \prod_{i}^{n_r} \mathcal{N}_c(\xi_0^{[i,j_i]}, \mathbf{\Omega}_0^{[i,j_i]}), \\ Z &\sim \mathcal{N}(\mathcal{M}).\end{aligned}$$

Considering the expected value of $\check{Z}$, we have that

$$\mathbb{E}(\check{Z}) = \tfrac{1}{n_m} \mathbb{E}(\sum_{j=1}^{n_m} \check{Z}^{[j]}) = \mathbb{E}(\check{Z}^{[1]}) = \sum_{a \in \prod_{i=1}^{n_r} \mathcal{I}(i)} w^{[a]} \mathbb{E}(Z^{[a]}) = \mathbb{E}(Z),$$

where the equalities in order are due to (i) the linearity of the expected value function, (ii) the independence of the drawn samples, (iii) the conditional independence of the robots' observations, and (iv) the definition of the joint mixture set. Thus, the joint mixture distribution is unbiased. □

[5]We have for all robots that $\sum_{j \in \mathcal{I}(i)} w^{[i,j]} = 1$.

[6]For weak convergence, we are assuming that the robot's IMD belongs to a certain reproducing kernel Hilbert space. See [12] for more detail.

[7]By unbiased we mean that the expected first moment of the approximation and the true distribution are equal.

C. Approximation of the Joint Mixture Distribution

Revisiting Remark 1, we can run a consensus-based algorithm on distribution parameters of sizes $O(n_g)$ and $O(n_g^2)$ to reasonably approximate a JMD of Gaussian form over all world states. This capability combined with the independence of the joint mixture set size with respect to the number of robots is the key to enabling distributed and scalable approximations of the JMD. The following formalizes the approach and its convergence, while the subsequent remarks discuss its significance and limitations.

Corollary 1 (Distributed Approximation and Convergence)**.** *For all robots and samples $j \in \{1, \ldots, n_m\}$, let $\check{\xi}_k^{[i,j]}$ and $\check{\mathbf{\Omega}}_k^{[i,j]}$ be initialized to $\check{\xi}_0^{[i,j]}$ and $\check{\mathbf{\Omega}}_0^{[i,j]}$, respectively, and have both evolve according to* (3) *on a connected graph $\mathcal{G}$. Define the ith robot's approximation of the joint mixture set to be $\check{\mathcal{M}}_k^{[i]} := \{(\beta_k^{[i]}\check{\xi}_k^{[i,j]}, \beta_k^{[i]}\check{\mathbf{\Omega}}_k^{[i,j]}) : j \in \{1, \ldots, n_m\}\}$. We have that*

$$\mathcal{N}(\check{\mathcal{M}}_k^{[i]}) := \sum_{j=1}^{n_m} \frac{\mathcal{N}_c(\beta_k^{[i]}\check{\xi}_k^{[i,j]}, \beta_k^{[i]}\check{\mathbf{\Omega}}_k^{[i,j]})}{n_m}$$

converges weakly to the joint mixture distribution as $k \to \infty$.

Proof (Corollary 1)**.** *The proof follows directly from Theorem 1 and Lemma 1.*

Remark 2 (Complexity and Scalability)**.** *We have that the communication complexity is $O(n_g^2 n_m)$ for each communication round, while the memory and computational complexity for calculating $\mathcal{N}(\check{\mathcal{M}}_k^{[i]})$ is also $O(n_g^2 n_m)$. Thus, this distributed algorithm scales quadratically with respect to Gaussian dimension and linearly with respect to number of samples. Most importantly, overall complexity remains constant with respect to number of robots.*

Remark 3 (Significance and Interpretation)**.** *The concatenation of the robot mixture sets (versus a Cartesian product) is allowed due to the fact that the corresponding samples are both unweighted (i.e., all samples are equally likely) and conditionally independent across the robots. Without these properties, Lemma 1 would not necessarily be true, and thus the robot's approximation of the JMD would be an arbitrarily poor representation. Instead, this approximation is guaranteed to converge weakly to the JMD as the communication round and the number of samples tend to infinity, or in words, as certain system resources increase.*

Remark 4 (Limitations)**.** *One should immediately recognize that as the number of robots increases, each mixture sample becomes less informative about the JMD of the entire system. Simultaneously increasing the robot mixture set size to retain the approximation's global precision can be exponential with respect to the number of robots. In our work using gradient-based information theoretic controllers, this limitation is not significant since we typically want to retain precision with respect to the JMD of a local neighborhood in physical proximity to the robot.*

IV. Performance Guarantees

A. Non-Degeneracy and Disagreement

We begin to characterize the joint mixture approximation by proving that the corresponding Gaussians make sense. Since we are forming these distributions from the scaled canonical parameters $\beta_k^{[i]}\check{\xi}_k^{[i,j]}$ and $\beta_k^{[i]}\check{\mathbf{\Omega}}_k^{[i,j]}$, this making sense objective is equivalent to proving for all robots, samples, and communication rounds that $\beta_k^{[i]}\check{\xi}_k^{[i,j]}$ is a real vector and $\beta_k^{[i]}\check{\mathbf{\Omega}}_k^{[i,j]}$ is a real positive-definite symmetric matrix. Since the collection of real vectors and the collection of positive-definite symmetric matrices are both closed under addition and positive scalar multiplication ($\beta_k^{[i]} \in [1, n_r]$ from the upcoming Lemma 3), it holds that the joint mixture set is composed of non-degenerate Gaussians.

The guarantee of non-degeneracy is fundamental to many of the claims to come. More interestingly, the mathematical structure of (3) that allows this guarantee also allows for intuitive interpretations of how the approximations evolve over time, especially concerning the rate of convergence of the scaled canonical parameters. These will be discussed shortly, but first we review the concept of exponentially decreasing disagreement [11].

Lemma 2 (Exponentially Decreasing Disagreement)**.** *For all robots and communication rounds, we have that*

$$\left\|\psi_k^{[i]} - \mathbf{1}\tfrac{1}{n_r}\right\|_2 \leq U_k^{\psi} := \left\|\boldsymbol{W} - \mathbf{1}\mathbf{1}^T\tfrac{1}{n_r}\ \right\|_2^k \left(1 - \tfrac{1}{n_r}\right)^{\frac{1}{2}},$$

where lefthand side of the inequality is termed disagreement and $\|\cdot\|_2$ for a matrix denotes the spectral norm.

Proof (Lemma 2)**.** *The proof follows from Xiao et al. [15] with $\psi_0^{[i]} = \boldsymbol{e}_i$, since*

$$\|\psi_0^{[i]} - \mathbf{1}\tfrac{1}{n_r}\|_2^2 = (1 - \tfrac{1}{n_r})^2 + \tfrac{n_r - 1}{n_r^2} = 1 - \tfrac{1}{n_r}.$$

Lemma 3 (Properties of $\psi_k^{[i]}$)**.** *For all robots and communication rounds, we have that $\psi_k^{[i]} \in [0,1]^{n_r}$, $\|\psi_k^{[i]}\|_1 = 1$, and $\|\psi_k^{[i]}\|_\infty \geq 1/n_r$.*

Proof (Lemma 3)**.** *Since for all robots $\|\psi_0^{[i]}\|_1 = \|\boldsymbol{e}_i\|_1 = 1$, we have that*

$$\|\psi_1^{[i]}\|_1 = [\boldsymbol{W}]_{ii}\|\psi_0^{[i]}\|_1 + \textstyle\sum_{v\in N^{[i]}}[\boldsymbol{W}]_{iv}\|\psi_0^{[v]}\|_1$$
$$= [\boldsymbol{W}]_{ii} + \sum_{v\in N^{[i]}}[\boldsymbol{W}]_{iv} = 1 - \sum_{v\in N^{[i]}}[\boldsymbol{W}]_{iv} + \sum_{v\in N^{[i]}}[\boldsymbol{W}]_{iv} = 1.$$

In addition, $\psi_1^{[i]}$ is nonnegative since it is an element-wise summation of nonnegative terms, which from the previous equation implies $\psi_1^{[i]} \in [0,1]^{n_r}$. Lastly, we have from the relationship between 1- *and* ∞*-norms that $\|\psi_1^{[i]}\|_\infty \geq \|\psi_1^{[i]}\|_1/n_r = 1/n_r$. The proof follows by induction on k.* □

B. Bounds on Scaled Canonical Parameters

In the following subsection, we simplify notation by dropping the overhead check $\check{\sqcup}$ and the sample index j. For example, we have that $\xi_k^{[i]}$ and $\mathbf{\Omega}_k^{[i]}$ denote $\check{\xi}_k^{[i,j]}$ and $\check{\mathbf{\Omega}}_k^{[i,j]}$, respectively.

It was discussed in [8] that the exponential factor $\beta_k^{[i]}$ accounts for the fact that the approximation of the JMD may be used before the consensus-based algorithm has converged. In our case, we expect this algorithm to prematurely terminate before the Gaussian parameters converge, and thus the exponential factor indicates how *close* the approximated information and information are to the true joint canonical parameters. In the following, we provide a strictly increasing lower bound for the exponential factor that equals one at $k=0$ and converges to n_r in the limit as k tends to infinity.

Theorem 2 (Lower Bound for the Exponential Factor)**.** *For all robots and communication rounds, we have that*

$$\beta_k^{[i]} \geq L_k^\beta := \big(U_k^\psi \sqrt{1-\tfrac{1}{n_r}} + \tfrac{1}{n_r}\big)^{-1}.$$

Proof (Theorem 2)**.** *Consider the optimization problem of maximizing $\|\psi_k^{[i]}\|_\infty$ with $\psi_k^{[i]}$ being a free variable subject to the constraints in Lemma 3 and subject to*

$$\|\psi_k^{[i]} - \boldsymbol{1}\tfrac{1}{n_r}\|_2 \leq U_k^\psi \in [0, \sqrt{1-\tfrac{1}{n_r}}] = [0, U_0^\psi].$$

Note that an optimal solution ψ_k^* always exists. Put $c \geq 0$, and without loss of generality suppose $\|\psi_k^*\|_\infty = [\psi_k^*]_1$ and $\|\psi_k^* - \boldsymbol{1}\frac{1}{n_r}\|_2^2 = c^2$. We define $f(\psi_k^{[i]}, \mu_1, \mu_2)$ to be

$$[\psi_k^{[i]}]_1 + \mu_1\big(\|\psi_k^{[i]} - \boldsymbol{1}\tfrac{1}{n_r}\|_2^2 - c^2\big) + \mu_2\big(\|\psi_k^{[i]}\|_1 - 1\big),$$

and by using Lagrange multipliers obtain

$$\tfrac{([\psi_k^{[i]}]_1 - 1/n_r)^2}{n_r - 1} + ([\psi_k^{[i]}]_1 - \tfrac{1}{n_r})^2 - c^2 = 0.$$

Thus, we have that $[\psi_k^{[i]}]_1 = c\sqrt{1-1/n_r} + 1/n_r$ and $c \leq \sqrt{1-1/n_r}$ since $[\psi_k^{[i]}]_1 \in [0,1]$. By the last equality, $[\psi_k^{[i]}]_1$ is proportional to c, and by the last inequality we conclude that $c = U_k^\psi$. Thus, the corresponding value of $[\psi_k^{[i]}]_1 = \|\psi_k^*\|_\infty$ is $U_k^\psi\sqrt{1-1/n_r} + 1/n_r$.

Lastly, consider $\psi_k^{[i]}$ as consensus term rather than a free variable. From above, we can interpret $\|\psi_k^*\|_\infty^{-1}$ as a lower bound for $\beta_k^{[i]}$ given U_k^ψ, which gives L_k^β. □

We now shift our attention to the geometric interpretation of the scaled information matrix $\beta_k^{[i]}\mathbf{\Omega}_k^{[i]}$, which describes ellipsoidal contours of equal density for the corresponding Gaussian. The squared lengths of the contours' principal axes are given by the inverse of the scaled information matrix eigenvalues, with larger values representing distribution axes of higher certainty. As more communication rounds are performed and the scaled information matrix converges element-wise, we expect this certainty to increase and also converge. This is in fact the case, and by using the lower bound for the exponential factor, we provide a strictly increasing lower bound for the scaled information matrix eigenvalues.

Theorem 3 (Lower Bound for the Scaled Information Matrix Eigenvalues)**.** *Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n_g}$. Then for all robots, samples, communication rounds, and $m \in \{1, \ldots, n_g\}$, we have that*

$$\lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]}) \geq L_{k,m}^{\mathbf{\Omega}} := \max\{L_{k,m}^{\mathbf{\Omega}-}, L_{k,m}^{\mathbf{\Omega}+}\},$$

where

$$L_{k,m}^{\mathbf{\Omega}-} := \textstyle\sum_{\ell=1}^{\lfloor L_k^\beta \rfloor} \lambda_{n_g}(\mathbf{\Omega}_0^{[\ell]}) + \big(L_k^\beta - \lfloor L_k^\beta \rfloor\big)\lambda_{n_g}(\mathbf{\Omega}_0^{[\lceil L_k^\beta \rceil]})$$

with the robot indices ordered such that $\lambda_1(\mathbf{\Omega}_0^{[1]}) \leq \lambda_1(\mathbf{\Omega}_0^{[2]}) \leq \cdots \leq \lambda_1(\mathbf{\Omega}_0^{[n_r]})$, and where

$$L_{k,m}^{\mathbf{\Omega}+} := \lambda_m(\mathbf{\Omega}) - \textstyle\sum_{\ell=\lceil L_k^\beta \rceil + 1}^{n_r} \lambda_{n_g}(\mathbf{\Omega}_0^{[\ell]}) - \big(\lceil L_k^\beta \rceil - L_k^\beta\big)\lambda_{n_g}(\mathbf{\Omega}_0^{[\lceil L_k^\beta \rceil]})$$

with $\lambda_{n_g}(\mathbf{\Omega}_0^{[1]}) \leq \lambda_{n_g}(\mathbf{\Omega}_0^{[2]}) \leq \cdots \leq \lambda_{n_g}(\mathbf{\Omega}_0^{[n_r]})$.

Proof (Theorem 3)**.** *We first prove that $\lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]})$ is bounded below by $L_{k,m}^{\mathbf{\Omega}-}$. Note that*

$$\beta_k^{[i]}\mathbf{\Omega}_k^{[i]} = \beta_k^{[i]} \textstyle\sum_{v=1}^{n_r} [\psi_k^{[i]}]_v \mathbf{\Omega}_0^{[v]}.$$

Recursively applying Weyl's Theorem [7], we have that

$$\lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]}) \geq \beta_k^{[i]} \textstyle\sum_{v=1}^{n_r} [\psi_k^{[i]}]_v \lambda_1(\mathbf{\Omega}_0^{[v]}). \tag{8}$$

Ordering the robot indices for λ_1 and using the lower bound from Theorem 2, we have that

$$\beta_k^{[i]} \sum_{v=1}^{n_r} [\psi_k^{[i]}]_v \lambda_1(\mathbf{\Omega}_0^{[v]}) \geq \sum_{\ell=1}^{\lfloor L_k^\beta \rfloor} \lambda_1(\mathbf{\Omega}_0^{[\ell]}) + \big(L_k^\beta - \lfloor L_k^\beta \rfloor\big)\lambda_1(\mathbf{\Omega}_0^{[\lceil L_k^\beta \rceil]}).$$

Substituting this inequality into (8) *gives $L_{k,m}^{\mathbf{\Omega}-}$.*

Lastly we prove in similar fashion that $\lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]})$ is bounded above by $L_{k,m}^{\mathbf{\Omega}+}$. Note that

$$\beta_k^{[i]}\mathbf{\Omega}_k^{[i]} = \mathbf{\Omega} - \textstyle\sum_{v=1}^{n_r} (1 - \beta_k^{[i]}[\psi_k^{[i]}]_v)\mathbf{\Omega}_0^{[v]}.$$

Recursively applying Weyl's Theorem, we have that

$$\lambda_m(\mathbf{\Omega}) \leq \lambda_m(\mathbf{\Omega}_k^{[i]}) + \textstyle\sum_{v=1}^{n_r} (1 - \beta_k^{[i]}[\psi_k^{[i]}]_v)\lambda_{n_g}(\mathbf{\Omega}_0^{[v]}). \tag{9}$$

Ordering the robot indices for λ_{n_g} and using the upper bound from Theorem 2, we have that

$$\textstyle\sum_{v=1}^{n_r} (1 - \beta_k^{[i]}[\psi_k^{[i]}]_v)\lambda_{n_g}(\mathbf{\Omega}_0^{[v]}) \leq \big(\lceil L_k^\beta \rceil - L_k^\beta\big)\lambda_{n_g}(\mathbf{\Omega}_0^{[\lceil L_k^\beta \rceil]}) + \textstyle\sum_{\ell=\lceil L_k^\beta \rceil + 1}^{n_r} \lambda_{n_g}(\mathbf{\Omega}_0^{[\ell]}).$$

Subtracting the summation term from both sides of (9)*, substituting the result into the previous inequality gives* $L^{\mathbf{\Omega}+}_{k,m}$*.* □

Remark 5 (Maximum of Two Bounds)**.** *The use of both* $L^{\mathbf{\Omega}-}_{k,m}$ *and* $L^{\mathbf{\Omega}+}_{k,m}$ *yields an intuitive bound on* $\lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]})$ *in the instances where* $k = 0$ *and* $k \to \infty$*, respectively. The former implies* $\lambda_m(\mathbf{\Omega}_0^{[i]}) \leq \max_v \lambda_m(\mathbf{\Omega}_0^{[v]})$ *and the latter with Lemma 3 implies* $\lim_{k\to\infty} \lambda_m(\beta_k^{[i]}\mathbf{\Omega}_k^{[i]}) = \lambda_m(\mathbf{\Omega})$*, both of which are obvious. In addition, the two bounds are equivalent for univariate Gaussians (i.e.,* $n_g = 1$*).*

Lastly, we derive the strictly shrinking range for the information vector elements, which when combined with the bounds on the information matrix eigenvalues well characterizes the convergence behavior of the resulting Gaussians. We believe such characterizations can lead to bounds on such information theoretic metrics such as Kullback-Leibler divergence of the mixture of Gaussians, however, such efforts are reserved for future work.

Corollary 2 (Range of the Scaled Information Vector Elements)**.** *For all robots, samples, communication rounds, and* $m \in \{1, \ldots, n_g\}$*, we have that* $L^{\xi}_{k,m} \leq [\beta_k^{[i]}\xi_k^{[i]}]_m \leq U^{\xi}_{k,m}$*, where*

$$L^{\xi}_{k,m} := \sum_{v=1}^{\lfloor L^{\beta}_k \rfloor} [\xi_0^{[v]}]_m + (L^{\beta}_k - \lfloor L^{\beta}_k \rfloor)[\xi_0^{[\lceil L^{\beta}_k \rceil]}]_m$$

with the robot indices arranged such that $[\xi_0^{[1]}]_m \leq [\xi_0^{[2]}]_m \leq \cdots \leq [\xi_0^{[n_r]}]_m$*, and where* $U^{\xi}_{k,m}$ *is defined the same as* $L^{\xi}_{k,m}$ *but with* $[\xi_0^{[1]}]_m \geq [\xi_0^{[2]}]_m \geq \cdots \geq [\xi_0^{[n_r]}]_m$*.*

Proof (Corollary 2)**.** *The proof follows from applying Theorem 2 to Theorem 1.*

V. Numerical Simulations

A. Consensus of Univariate Gaussians

We first consider an object localization task using ten robots with IMDs that can be accurately represented by univariate Gaussians. Such a simplified task best illustrates how each robot's JMD approximation converges weakly to the true one, which is analogous to how each Gaussian element of a mixture would converge. Figure 2 shows the probability distributions with corresponding parameters for the ten IMDs. Note that we selected the canonical parameters to separate the distributions for illustrative purposes, as one should not expect such initial disagreement within a fielded robot system. Figure 2 also shows the JMD of nonzero mean, since the assumption of zero mean can potentially lead to misleadingly tight bounds (e.g., bounds that are not invariant under translation).

We evaluated the performance of our consensus-based algorithm on the connected communication graph shown in Figure 1. Figure 3 shows the evolution of each robot's JMD approximation with respect to a strictly shrinking envelope

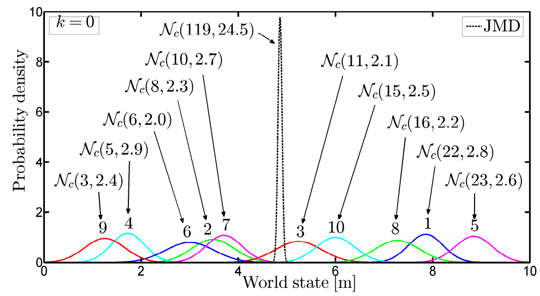

Fig. 2: This figure shows the one dimensional IMDs of the ten robots with respect to the JMD.

derived from bounds given in Theorem 3 and Corollary 2. These envelopes can be interpreted as feasible regions within which the peaks of all the robots' JMD approximations must lie, intuitively highlighting the performance guarantees discussed in Section IV. We note that these bounds for this particular communication graph are conservative; we found that graphs with higher algebraic connectivity tend to produce tighter bounds.

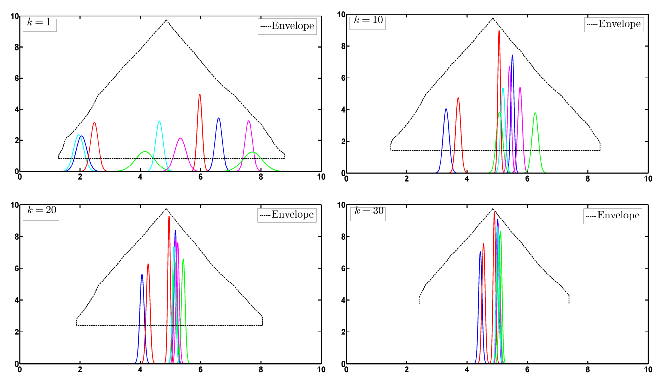

Fig. 3: This figure shows the evolution of each robot's JMD approximation on a connected communication graph topology at communication rounds of $k = \{1, 10, 20, 30\}$. The dashed envelope in each plot represents the feasible region within which the peak of every robot's JMD approximation must lie.

B. Consensus of Mixtures of Gaussians

We now focus on an object localization task where the world state is three dimensional and each robot's IMD cannot be accurately represented by a single Gaussian. Consider the ten robots in Figure 1 task to localize an object by taking range only measurements that obey a Gaussian sensor model. The resulting IMDs are three dimensional probability distributions with contours of equal probability being spheres centered at the corresponding robot (Figure 4). To represent the IMD with a weighted Gaussian mixture given an observation, each robot forms a weighted mixture set of three dimensional Gaussian elements using a dodecahedron-based geometric layout.

Using a computer cluster, Monte Carlo simulations employing various mixture sizes were performed in a distributed fashion, meaning that every robot was simulated on an independent computer cluster node and communicated using

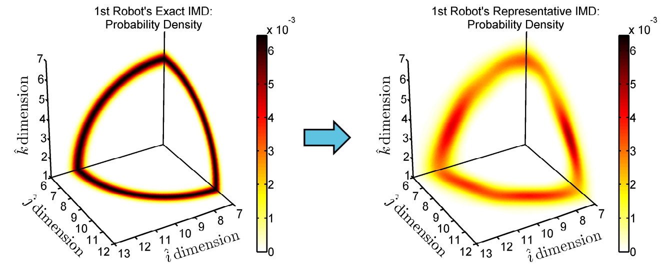

Fig. 4: This figure shows the process of representing a robot's IMD with a weighted mixture of Gaussians. A range only sensor model of Gaussian distribution was assumed to have one standard deviation accuracy equal to five percent of the received sensor measurement. *Left:* Given a relative observation distance of 4.5 meters, the robot's three dimensional IMD is generated for all relative world states. Here we show the mixture distribution on two dimensional slices that intersect at the robot's location of $(7, 6, 1)$ meters. *Right:* A weighted mixture of 32 Gaussian elements is formed to represent the IMD, where each element is located at a vertex or a face centroid of a dodecahedron concentric with the robot.

MatlabMPI. For a given simulation run, each robot (i) drew an observation from the previously described range only measurement model, (ii) represented the resulting IMD with a weighted Gaussian mixture, (iii) drew unweighted samples to form the unweighted mixture set, and finally (iv) ran the consensus-based algorithm to approximate the JMD. Figure 1 illustrates one particular evolution of the 1st robot's three dimensional JMD approximation, which becomes more accurate and precise as more consensus rounds are performed.

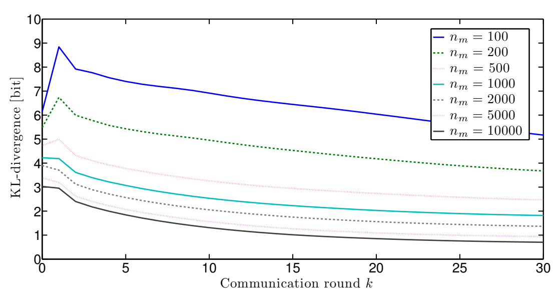

Fig. 5: This figure shows the average Kullback-Leibler divergence of the robots' JMD approximations with respect to the joint mixture distribution for various mixture sizes.

Figure 5 shows the average Kullback-Leibler divergence over 1000 simulations with respect to the joint mixture distribution. Not surprisingly, the divergence at all communication rounds is smaller for larger mixture sizes; however, this behavior clearly exhibits diminishing returns. In addition, more than 500 samples are needed to prevent the divergence from initially increasing, although again all JMD approximations by Corollary 1 are guaranteed to converge weakly to the joint mixture distribution as the number of communication rounds tend to infinity.

VI. Conclusions

We present a scalable, decentralized approach that enables robots within a large team to independently perform posterior calculations needed for sequential Bayesian filter predictions and mutual information-based control. We focused on distributively approximating the joint of continuous-valued measurement distributions, providing performance guarantees and complexity analyses. We are currently investigating the concepts of sample deprivation and diminishing returns that were highlighted in our numerical simulations. Lastly, we wish to adapt the approach for specific types of Bayes filters (e.g., Kalman) for which we can benchmark against much prior work. Our paradigm of sample-based sensor fusion combined with the notion of pre-convergence termination has the potential to impact how the research community perceives scalability in multi-robot systems.

References

[1] M. Alighanbari and J. P. How. An unbiased Kalman consensus algorithm. In *Proceedings of the American Control Conference*, pages 4753–4766, 2006.

[2] R. Aragués, J. Cortés, and C. Sagués. Dynamic consensus for merging visual maps under limited communications. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3032–3037, May 2010.

[3] J. Cortés. Distributed algorithms for reaching consensus on general functions. *Automatica*, 44(3):727–737, 2008.

[4] H. F. Durrant-Whyte, B. Y. S. Rao, and H. Hu. Toward a fully decentralized architecture for multi-sensor data-fusion. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pages 1331–1336, 1990.

[5] C. S. R. Fraser, L. F. Bertuccelli, H. L. Choi, and J. P. How. A hyperparameter consensus method for agreement under uncertainty. *Automatica*, 48(2):374–380, February 2012.

[6] J. A. Gubner. Distributed estimation and quantization. *IEEE Transactions on Information Theory*, 39(4):1456–1459, 1993.

[7] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[8] B. J. Julian, M. Angermann, and D. Rus. Nonparametric inference and coordination for distributed robotics. In *Proceedings of the IEEE Conference on Decision and Control*, December 2012. Submitted.

[9] B. J. Julian, M. Angermann, M. Schwager, and D. Rus. A scalable information theoretic approach to distributed robot coordination. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5187–5194, 2011.

[10] A. Manyika and H. F. Durrant-Whyte. *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Prentice Hall, 1998.

[11] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. In *Proceedings of the Network Embedded Sensing and Control Workshop*, pages 169–182, 2005.

[12] R. B. Platte and T. A. Driscoll. Polynomials and potential theory for Gaussian radial basis function interpolation. *SIAM Journal on Numerical Analysis*, 43(2):750–766, 2006.

[13] W. Ren, R. W. Beard, and D. B. Kingston. Multi-agent Kalman consensus with relative uncertainty. In *Proceedings of the American Control Conference*, pages 1865–1870, 2005.

[14] D. Williams. *Probability With Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, 1991.

[15] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67:33–46, January 2007.

[16] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Symposium on Information Processing of Sensor Networks*, pages 63–70, 2005.

[17] P. Yang, R. A. Freeman, and K. M. Lynch. Distributed cooperative active sensing using consensus filters. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, pages 405–410, May 2006.