Joint Estimation and Localization in Sensor Networks

Nikolay Atanasov, Roberto Tron, Victor M. Preciado, and George J. Pappas

Abstract—This paper addresses the problem of collaborative tracking of dynamic targets in wireless sensor networks. A novel distributed linear estimator, which is a version of a *distributed Kalman filter*, is derived. We prove that the filter is mean square consistent in the case of static target estimation. When large sensor networks are deployed, it is common that the sensors do not have good knowledge of their locations, which affects the target estimation procedure. Unlike most existing approaches for target tracking, we investigate the performance of our filter when the sensor poses need to be estimated by an auxiliary localization procedure. The sensors are localized via a distributed Jacobi algorithm from noisy relative measurements. We prove strong convergence guarantees for the localization method and in turn for the joint localization and target estimation approach. The performance of our algorithms is demonstrated in simulation on environmental monitoring and target tracking tasks.

I. INTRODUCTION

A central problem in networked sensing systems is the estimation and tracking of the states of dynamic phenomena of interest that evolve in the sensing field. Potential applications include environmental monitoring [1], [2], surveillance and reconnaissance [3], [4], social networks [5]. In most situations, individual sensors receive partially informative measurements which are insufficient to estimate the target state in isolation. The sensors need to engage in information exchange with one another and solve a distributed estimation problem. To complicate matters, it is often the case that the sensors need to know their own locations with respect to a common reference in order to utilize the target measurements meaningfully. Hence, in general, the sensors face a joint *localization* and *estimation* problem. Virtually all existing work in distributed target estimation assumes implicitly that the localization problem is solved, while all the literature on localization does not consider the effect of the residual errors on a common estimation task. The goal of this paper is to show that the two problems can be solved jointly, and that, with simple measurement models, the resulting estimates have strong convergence guarantees.

Assumptions and contributions. We assume that the sensors obtain linear Gaussian measurements of the target state and repeated sequential measurements of their relative positions along the edges of a graph. Our contributions are as follows:

- We derive a distributed linear estimator for tracking dynamic targets. We prove that the filter is mean-square consistent in the case of a static target.
- We provide a distributed algorithm for sensor localization from sequential relative measurements and prove mean-square and strong consistency.
- We prove mean-square consistency of the joint localization and target estimation procedure.

Related work. Our target estimation algorithm was inspired by Rahnama Rad and Tahbaz-Salehi [6], who propose an algorithm for distributed *static* parameter estimation using nonlinear sensing models. We specialize their model to heterogeneous sensors with linear Gaussian observations, show stronger convergence results (mean-square consistency instead of weak consistency), and then generalize the solution to *dynamic* targets. Our filter is similar to the Kalman-Consensus [7], [8] and the filter proposed by Khan et al. [9], [10]. Khan et al. [9] show that a dynamic target can be tracked with bounded error if the norm of the target system matrix is less than the network tracking capacity. Shahrampour et al. [10] quantify the estimation performance using a global loss function and show that the asymptotic estimation error depends on its decomposition. Kar et al. [11] study distributed static parameter estimation with nonlinear observation models and noisy inter-sensor communication. Related work also includes [12], which combines the Jacobi over-relaxation method with dynamic consensus to compute distributed weighted least squares.

Our localization algorithm follows the lines of the Jacobi algorithm, first proposed for localization in sensor networks by Barooah and Hespanha [13], [14]. In contrast with their approach, we consider repeated relative measurements and show strong convergence guarantees for the resulting sequential localization algorithm. Other work in sensor network localization considers nonlinear and less informative measurement models than those used in this paper. For instance [15], [16], [17], [18] address the problem of localization using range-only measurements, which is challenging because a graph with specified edge lengths can have several embeddings in the plane. Khan et al. [19] introduce a distributed localization (DILOC) algorithm, which uses the barycentric coordinates of a node with respect to its neighbors and show its convergence via a Markov chain. Diao et al. [20] relax the assumption of DILOC that all nodes must be inside the convex hull of the anchors. Localization has also been considered in the context of camera networks [21].

Paper organization. The joint localization and estimation problem is formulated precisely in Sec. II. The distributed linear estimator for target tracking is derived in Sec. III assuming known sensor locations. A distributed Jacobi algorithm

is introduced in Sec. IV to localize the sensors using relative measurements when the true locations are unknown. Mean-square and strong consistency are proven. In Sec. V, we show that the error of the target estimator, when combined with the localization procedure, remains arbitrarily small. All proofs are provided in the Appendix.

II. PROBLEM FORMULATION

Consider a static sensor network composed of n sensors with configurations $\{x_1, \ldots, x_n\} \subset \mathcal{X} \cong \mathbb{R}^d$. The configuration of a sensor may include its position, orientation, and other operational parameters but we will refer to it, informally, as the sensor's location. The communication network interconnecting the sensors is represented by an undirected graph $G = (V, E)$ with vertices $V := \{1, \ldots, n\}$ corresponding to the sensors and $|E| = m$ edges. An edge $(j, i) \in E$ from sensor j to sensor i exists if they can communicate. The set of nodes (neighbors) connected to sensor i is denoted by $\mathcal{N}_i$.

The task of the sensors is to estimate and track the state $y(t) \in \mathcal{Y} \cong \mathbb{R}^{d_y}$ of a phenomenon of interest (target), where $\mathcal{Y}$ is a convex set. The target evolves according to the following *target motion model*:

$$y(t+1) = Fy(t) + \eta(t), \qquad \eta(t) \sim \mathcal{N}(0, W), \quad (1)$$

where $\eta(t)$ is the process noise, whose values at any pair of times are independent. Sensor i, depending on its location x_i, can obtain a measurement $z_i(t)$ of the target state $y(t)$ at time t according to the following *sensor observation model*:

$$z_i(t) = H_i(x_i)y(t) + v_i(t, x_i), \ \ v_i(t, x_i) \sim \mathcal{N}(0, V_i(x_i)), \ (2)$$

where $v_i(t, x_i)$ is a sensor-state-dependent measurement noise specific to sensor i, which is independent at any pair of times and across different sensors. The measurement noise is independent of the target noise $\eta(t)$ as well. The signals, $z_i(t)$, observed by a single sensor, although potentially informative, do not reveal the target state completely, i.e. each sensor faces a local identification problem. We assume, however, that the target is observable if one has access to the signals received by all sensors.

The sensors need to know their locations in order to use the signals $z_i(t)$ to estimate the targets state. However, when large sensor networks are deployed, it is common that the sensors do not have good knowledge of their positions but instead only a rough estimate (prior). We suppose that each sensor has access to noisy relative measurements of the positions of its neighbors[1], which can be used to localize the sensors. In particular, at time t sensor i receives the following noisy relative configuration measurement from its neighbor j:

$$s_{ij}(t) = x_j - x_i + \epsilon_{ij}(t), \qquad \epsilon_{ij}(t) \sim \mathcal{N}(0, \mathcal{E}_{ij}), \quad (3)$$

where $\epsilon_{ij}(t)$ is a measurement noise, which is independent at any pair of times and across sensor pairs. The relative measurement noises are independent of the target measurement and

---

[1]The graphs describing the communication and the relative measurement topologies might be different in practice. However, we assume that they are the same in order to simplify the presentation.

motion noises too. Since there is translation ambiguity in the measurements (3) we assume that all sensors agree to localize themselves in the reference frame of sensor 1. The location estimates can then be used in place of the unknown sensor positions during the target estimation procedure. The joint localization and estimation problem is summarized below.

Problem (Joint Estimation and Localization). *The task of each sensor i is to construct estimators $\hat{x}_i(t)$ and $\hat{y}_i(t)$ of its own location x_i and of the target state y in a distributed manner, i.e. using information only from its neighbors and the measurements $\{s_{ij}(t) \mid j \in \mathcal{N}_i\}$ and $\{z_i(t)\}$.*

To illustrate the results, we use two scenarios which fit our models throughout the paper. The first is environmental monitoring problem in which a sensor network of remote methane leak detectors (RMLD), based on tunable diode laser absorption spectroscopy, is deployed to estimate the methane concentration in a landfill. The methane field is assumed static (i.e. $F = I_{d_y}, W = 0$) and can be modeled by discretizing the environment into cells and representing the gas concentration with a discrete Gaussian random field, $y \in \mathbb{R}^{d_y}$ (See Fig.5). It was verified experimentally in [22] that the RMLD sensors fit the linear model in (2). Second, we consider tracking a swarm of mobile vehicles via a sensor network using range and bearing measurements (See Fig. 1). The position $(y_j^1, y_j^2) \in \mathbb{R}^2$ and velocity $(\dot{y}_j^1, \dot{y}_j^2) \in \mathbb{R}^2$ of the jth target have discretized double integrator dynamics driven by Gaussian noise:

$$y_j(t+1) = \begin{bmatrix} I_2 & \tau I_2 \\ 0 & I_2 \end{bmatrix} y_j(t) + \eta_j(t), \ \ W := q \begin{bmatrix} \frac{\tau^3}{3}I_2 & \frac{\tau^2}{2}I_2 \\ \frac{\tau^2}{2}I_2 & \tau I_2 \end{bmatrix},$$

where $y_j = [y_j^1, y_j^2, \dot{y}_j^1, \dot{y}_j^2]^T$ is the j-th target state, τ is the sampling period is sec, and q is a diffusion strength measured in $(\frac{m}{sec^2})^2 \frac{1}{Hz}$. Each sensor in the network takes noisy range and bearing measurements of the target's position:

$$z_{ij}(t) = \begin{bmatrix} \sqrt{(y_j^1 - x_i^1)^2 + (y_j^2 - x_i^2)^2} \\ \arctan\left((y_j^2 - x_i^2)/(y_j^1 - x_i^1)\right) \end{bmatrix} + v(t, x_i, y_j), \ (4)$$

where $x_i := (x_i^1, x_i^2) \in \mathbb{R}^2$ is the sensor's location and the noise v grows linearly with the distance between the sensor and the target. The observation model is nonlinear in this case so we resort to linearization in order to apply our framework.

III. DISTRIBUTED TARGET ESTIMATION

We begin with the task of estimating and tracking the dynamic state of a target via the sensor network. For now we assume that the sensors know their positions and concentrate on the estimation task. We specializing the general parameter estimation scheme of Rahnama Rad and Tahbaz-Salehi [6] to linear Gaussian observation models such as (2). We show that the resulting distributed linear filter is mean-square consistent[2] when the target is stationary. This result is

---

[2]A distributed estimator of a parameter y is *weakly consistent* if all estimates, $\hat{y}_i(t)$, converge in probability to y, i.e. $\lim_{t \to \infty} \mathbb{P}(\|\hat{y}_i(t) - y\| \geq \epsilon) = 0$ for any $\epsilon > 0$ and all i. It is *mean-square consistent* if all estimates converge in L^2 to y, i.e. $\lim_{t \to \infty} \mathbb{E}[\|\hat{y}_i(t) - y\|^2] = 0, \forall i$.
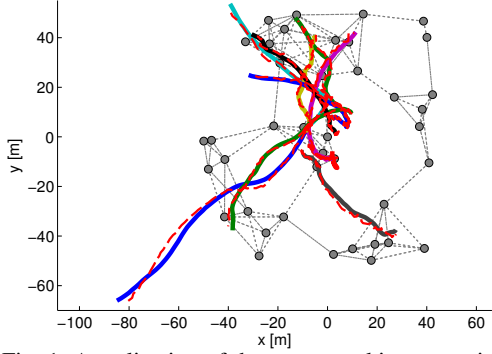
Fig. 1: A realization of the target tracking scenario in which a sensor network with 40 nodes tracks 10 mobile targets via range and bearing measurements
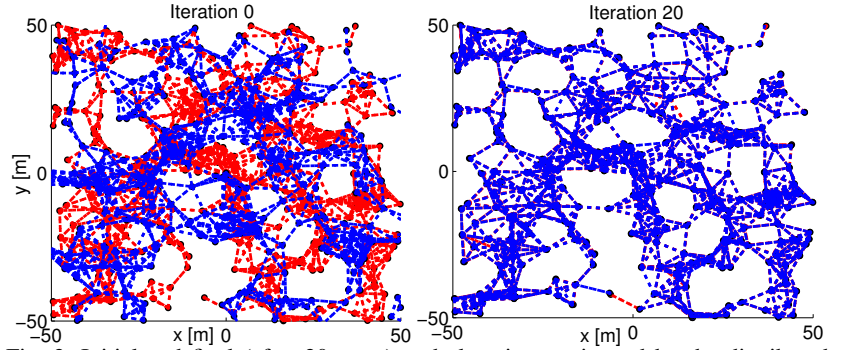
Fig. 2: Initial and final (after 20 steps) node locations estimated by the distributed localization algorithm on a randomly generated graph with 300 nodes ans 1288 edges

stronger than the weak consistency[2] shown in the general non-Gaussian case in [6, Thm.1]. Suppose for now that the target is stationary, i.e. $y := y(0) = y(1) = \ldots$. To introduce the estimation scheme from [6], suppose also that instead of the linear Gaussian measurements in (2), the sensor measurements $z_i(t)$ are drawn from a general distribution with conditional probability density function (pdf) $l_i(\cdot \mid y)$. As before, the signals observed by sensor i are iid over time and independent from the observations of all other sensors. In order to aggregate the information provided to it over time - either through observations or communication with neighbors - each sensor i holds and updates a pdf $p_{i,t}$ over the target state space $\mathcal{Y}$. Consider the following distributed estimation algorithm:

$$p_{i,t+1}(y) = \xi_{i,t} l_i(z_i(t+1) \mid y) \prod_{j \in \mathcal{N}_i \cup \{i\}} \left(p_{j,t}(y)\right)^{\kappa_{ij}},$$

$$\hat{y}_i(t) \in \arg\max_{y \in \mathcal{Y}} p_{i,t}(y),$$

(5)

where $\xi_{i,t}$ is a normalization constant ensuring that $p_{i,t+1}$ is a proper pdf and $\kappa_{ij} > 0$ are weights such that $\sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij} = 1$. The update is the same as the standard Bayes rule with the exception that sensor i does not just use its own prior but a *geometric average* of its neighbors' priors. Given a connected graph, the authors of [6] show that (5) is weakly consistent under broad assumptions on the observation models l_i.

Next, we specialize the estimator in (5) to the linear Gaussian measurement model in (2). Let $\mathcal{G}(\omega, \Omega)$ denote a Gaussian distribution (in information space) with mean $\Omega^{-1}\omega$ and covariance matrix Ω^{-1}. The quantities ω and Ω are conventionally called *information vector* and *information matrix*, respectively. Suppose that the pdfs $p_{i,t}$ of all sensors $i \in V$ at time t are that of Gaussian distributions $\mathcal{G}(\omega_{i,t}, \Omega_{i,t})$. We claim that the posteriors resulting from applying the update in (5) remain Gaussian.

Lemma 1 ([23, Thm.2]). *Let $Y_i \sim \mathcal{G}(\omega_i, \Omega_i)$ for $i = 1, \ldots, n$ be a collection of random Gaussian vectors with associated weights κ_i. The weighted geometric mean, $\prod_{i=1}^n p_i^{\kappa_i}$, of their pdfs p_i is proportional to the pdf of a random vector with distribution $\mathcal{G}\left(\sum_{i=1}^n \kappa_i \omega_i, \sum_{i=1}^n \kappa_i \Omega_i\right)$.*

Lemma 2 ([23, Thm.2]). *Let $Y \sim \mathcal{G}(\omega, \Omega)$ and $\mathcal{V} \sim \mathcal{G}(0, V^{-1})$ be random vectors. Consider the linear transformation $Z = HY + \mathcal{V}$. The conditional distribution of $Y \mid Z = z$ is proportional to $\mathcal{G}(\omega + H^T V^{-1} z, \Omega + H^T V^{-1} H)$.*

Lemma 1 says that if the sensor priors are Gaussian $\mathcal{G}(\omega_{i,t}, \Omega_{i,t})$, then after applying the geometric averaging in (5) the resulting distribution will still be Gaussian and its information vector and information matrix will be weighted averages of the prior ones. Lemma 2 says that after applying Bayes rule the distribution will remain Gaussian. Combining the two allows us to derive the following linear Gaussian version of the estimator in (5):

$$\omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij} \omega_{j,t} + H_i^T V_i^{-1} z_i(t),$$

$$\Omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij} \Omega_{j,t} + H_i^T V_i^{-1} H_i,$$

(6)

where $H_i := H_i(x_i)$ and $V_i := V_i(x_i)$. The estimate of sensor i at time t of the true target state y is:

$$\hat{y}_i(t) := \Omega_{i,t}^{-1} \omega_{i,t}.$$

(7)

In this linear Gaussian case, we prove a strong result about the quality of the estimates.

Theorem 1. *Suppose that the communication graph G is connected and the matrix $\begin{bmatrix} H_1^T & \ldots & H_n^T \end{bmatrix}^T$ has rank d_y. Then, the estimates (7) of all sensors converge in mean square to y, i.e. $\lim_{t \to \infty} \mathbb{E}\left[\|\hat{y}_i(t) - y\|_2^2\right] = 0$ for all i.*

The estimation procedure in (6), (7) can be extended to track a dynamic target as in (1) by adding a local prediction step, same as that of the Kalman filter, at each sensor. The distributed linear filter is summarized in Alg. 1 and Thm. 1 guarantees its mean-square consistency for stationary targets. Its performance on dynamic targets was studied in the target tracking scenario introduced in Sec. II and the results are presented in Fig. 1 and Fig. 3.

IV. LOCALIZATION FROM RELATIVE MEASUREMENTS

Target tracking via the distributed estimator in Alg. 1 requires knowledge of the true sensor locations. As mentioned
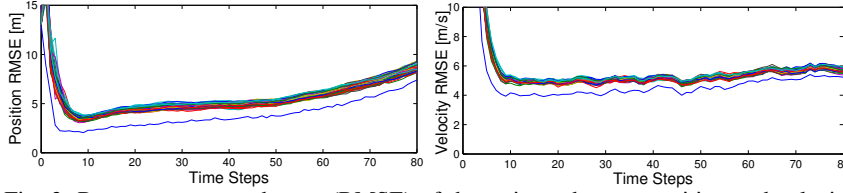
Fig. 3: Root mean squared error (RMSE) of the estimated target position and velocity obtained from averaging 50 simulated runs of the distributed linear estimator in the target tracking scenario (Fig. 1). The error increases because as targets move away from the sensor network, the covariance of the measurement noise grows linearly with distance. The errors of node 1 (blue) are lower because it was always placed at the origin and thus close to the starting target positions.
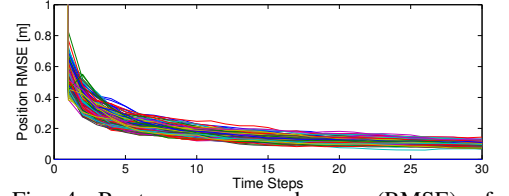

Fig. 4: Root mean squared error (RMSE) of the location estimates obtained from averaging 50 simulated runs of the distributed localization alogorithm with randomly generated graphs with 300 nodes (e.g. Fig. 2)

Algorithm 1 Distributed Linear Estimator

Input: Prior $(\omega_{i,t}, \Omega_{i,t})$, messages $(\omega_{j,t}, \Omega_{j,t}), \forall j \in \mathcal{N}_i$, and measurement $z_i(t)$
Output: $(\omega_{i,t+1}, \Omega_{i,t+1})$

Update Step:
$$\omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij} \omega_{j,t} + H_i^T V_i^{-1} z_i(t)$$
$$\Omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij} \Omega_{j,t} + H_i^T V_i^{-1} H_i$$
$$\hat{y}_i(t+1) = \Omega_{i,t+1}^{-1} \omega_{i,t+1}$$

Prediction Step:
$$\Omega_{i,t+1} = (F \Omega_{i,t+1}^{-1} F^T + W)^{-1}$$
$$\omega_{i,t+1} = \Omega_{i,t+1} F \hat{y}_i(t+1)$$

---

earlier this is typically not the case, especially when large sensor networks are deployed. This section describes a method for localization from relative measurements (3), whose strong convergence guarantees can be used to analyze the convergence of a joint localization and estimation procedure. The relative measurements, received by all sensors at time t, can be written in matrix form as follows:

$$s(t) = (B \otimes I_d)^T x + \epsilon(t),$$

where $B \in \mathbb{R}^{n \times m}$ is the incidence matrix of the communication graph G. All sensors agree to localize relative to node 1 and know that $x_1 = 0$. Let $\tilde{B} \in \mathbb{R}^{(n-1) \times m}$ be the incidence matrix with the row corresponding to sensor 1 removed. Further, define $\mathcal{E} := \mathbb{E}[\epsilon(t)\epsilon(t)^T] = \mathbf{diag}(\mathcal{E}_1, \ldots, \mathcal{E}_m)$, where $\{\mathcal{E}_k\}$ is an enumeration of the noise covariances associated with the edges of G. Given t measurements, the least squares estimate of x leads to the classical Best Linear Unbiased Estimator (BLUE), given by:

$$\hat{x}(t) := \left(\tilde{B}\mathcal{E}^{-1}\tilde{B}^T\right)^{-1} \tilde{B}\mathcal{E}^{-1} \sum_{\tau=0}^{t-1} s(\tau), \qquad (8)$$

where the inverse of $\tilde{B}\mathcal{E}^{-1}\tilde{B}^T$ exists as long as the graph G is connected [13]. Among all linear estimators of x, BLUE has the smallest variance for the estimation error [24]. The computation in (8) can be distributed via a Jacobi algorithm for solving a linear system as follows. Each sensor maintains an estimate $\hat{x}_i(t)$ of its own state at time t and a history of the averaged measurements, $\sigma_i(t) := \frac{1}{t+1} \sum_{\tau=0}^{t} \sum_{j \in \mathcal{N}_i} \mathcal{E}_{ij}^{-1} s_{ij}(\tau)$, received up to time t. Given prior estimates $(\hat{x}_i(t), \sigma_i(t))$, the

update of the distributed Jacobi algorithm at sensor i is:

$$\hat{x}_i(t+1) = \left(\sum_{j \in \mathcal{N}_i} \mathcal{E}_{ij}^{-1}\right)^{-1} \left(\sum_{j \in \mathcal{N}_i} \mathcal{E}_{ij}^{-1} \hat{x}_j(t) - \sigma_i(t)\right),$$
$$\sigma_i(t+1) = \frac{1}{t+1} \left(t\sigma_i(t) + \sum_{j \in \mathcal{N}_i} \mathcal{E}_{ij}^{-1} s_{ij}(t)\right). \qquad (9)$$

Barooah and Hespanha [13], [14] show that, with a *single round* of relative measurements, the the Jacobi algorithm provides an unbiased estimate of x. Here, we incorporate repeated sequential measurements and prove much stronger performance guarantee.

Theorem 2. *Suppose that the communication graph G is connected. Then, the estimates $\hat{x}_i(t)$ of the sensor configurations in (9) are mean-square and strongly consistent estimators of the true sensor states, i.e.:*

$$\lim_{t \to \infty} \mathbb{E}\left[\|\hat{x}_i(t) - x_i\|_2^2\right] = 0, \ \mathbb{P}\left(\lim_{t \to \infty} \|\hat{x}_i(t) - x_i\|_2 = 0\right) = 1, \forall i$$

The performance of our distributed localization algorithm was analyzed on randomly generated graphs with 300 nodes. The location priors were chosen from a normal distribution with standard deviation of 5 meters from the true node positions. An instance of the localization task is illustrated in Fig. 2, while the estimation error is shown in Fig. 4.

V. JOINT LOCALIZATION AND ESTIMATION

Having derived separate estimators for the sensor locations and the target state, we are ready to return to the original problem of joint localization and estimation. At time t, the location estimates $\{\hat{x}_i(t)\}$ in (9) can be used in the target estimator (6), (7) instead of the true sensor positions. It is important to analyze the evolution of the coupled estimation procedure because it is not clear that the convergence result in Thm. 1 will continue to hold. Define the *sensor information matrix* $M_i(x) := H_i(x)^T V_i(x)^{-1} H_i(x)$. In an analogy with the centralized Kalman filter, the sensor information matrix captures the amount of information added to the inverse of the covariance matrix during an update step of the Riccati map. From this point of view, it is natural to describe sensor properties in terms of the sensor information matrix. A regularity assumption which stipulates that nearby sensing locations provide similar information gain is necessary.
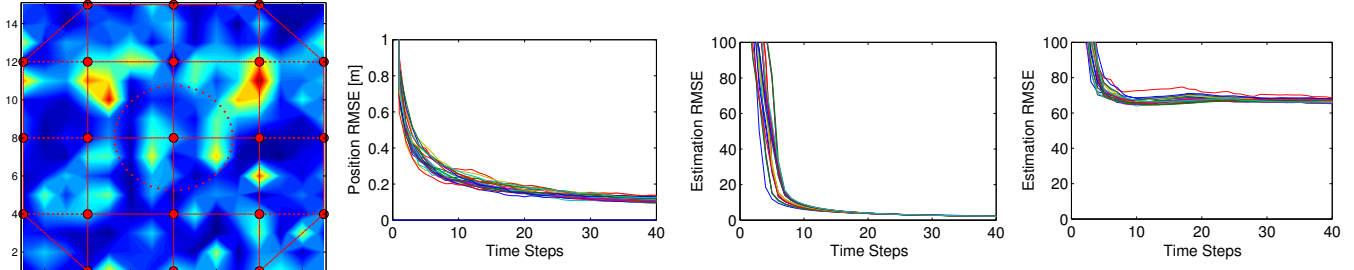
Fig. 5: Methane emission monitoring via a sensor network. The true (unknown) sensor locations (red dots), the sensing range (red circle), and a typical realization of the methane field are shown on the left. The root mean squared error (RMSE) of the location estimates and of the field estimates obtained from averaging 50 simulated runs of the joint localization and estimation algorithm with continuous sensor observation models are shown in the two middle plots. In an additional experiment, the sensors were placed on the boundaries of the cells of the discretized field. As the observation model for each sensor was defined in terms of the proximal environment cells, this made the model discontinuous. The rightmost plot illustrates that the field estimation error does not vanish when discontinuities are present.

Assumption (Observation Model Continuity). *The sensor information matrices $M_i(x)$ are bounded[3] continuous functions of x for all i.*

The following theorem ensures that the target state estimator retains its convergence properties when used jointly with the distributed localization procedure.

Theorem 3. *Let $\{\hat{x}_i(t)\}$ be strongly consistent estimators of the sensor configurations, i.e. $\hat{x}_i(t) \xrightarrow{a.s.} x_i, \forall i$. Suppose that the communication graph G is connected and the matrix $\begin{bmatrix} H_1(x_1)^T & \ldots & H_n(x_n)^T \end{bmatrix}^T$ has rank d_y. Let $\delta > 0$ be arbitrary. If each sensor i updates its target estimate $(\omega_{i,t}, \Omega_{i,t})$ as follows:*

$$\omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij}\omega_{j,t} + \widehat{H}_{i,t}^T \widehat{V}_{i,t}^{-1} z_i(t),$$

$$\Omega_{i,t+1} = \sum_{j \in \mathcal{N}_i \cup \{i\}} \kappa_{ij}\Omega_{j,t} + \widehat{H}_{i,t}^T \widehat{V}_{it}^{-1} \widehat{H}_{i,t}, \quad (10)$$

$$\hat{y}_i(t+1) = (\Omega_{i,t+1} + (t+1)\delta I_d)^{-1}\omega_{i,t+1},$$

where $\widehat{H}_{i,t} := H_i(\hat{x}_i(t))$ and $\widehat{V}_{i,t} := V_i(\hat{x}_i(t))$, then the asymptotic mean-square error of target estimates is $O(\delta^2)$:

$$\lim_{t \to \infty} \mathbb{E}\big[\|\hat{y}_i(t) - y\|_2^2\big] = \delta^2 y^T \Big(\sum_{j=1}^{n} \pi_j M_j(x_j) + \delta I\Big)^{-2} y,$$

for all i, where y is the true target state and x_j is the true position of sensor j.

The combined procedure specified by (9) and (10) provides a mean-square consistent way to estimate the sensor locations and the target state jointly. The performance of the joint algorithm was evaluated on the methane concentration estimation problem and the results are summarized in Fig. 5.

VI. CONCLUSION

This paper studied the problem of joint target tracking and node localization in sensor networks. A distributed linear estimator for tracking dynamic targets was derived. It was

[3]There exists a constant q such that $\|M_i(x)\| \leq q < \infty$ for all i and x.

proven that the filter is mean-square consistent when estimating static states. Next, a distributed Jacobi algorithm was proposed for localization and its mean-square and almost sure consistency were shown. Finally, the combined localization and target estimation procedure was shown to have arbirarily small asymptotic estimation error.

Future work will focus on strengthening the result in Thm. 3 to mean-square consistency and relaxing the assumption of a strongly consistent localization procedure. Studying the relationship between our distributed linear estimator, the Kalman-Consensus filter [8], and the filter proposed by Khan et al. [9] is of interest as well.

APPENDIX A: PROOF OF THEOREM 1

Define the following:

$$\omega_t := \begin{bmatrix} \omega_{1t}^T & \ldots & \omega_{nt}^T \end{bmatrix}^T \qquad \Omega_t := \begin{bmatrix} \Omega_{1t}^T & \ldots & \Omega_{nt}^T \end{bmatrix}^T$$

$$M_i := H_i(x_i)^T V_i^{-1}(x_i)H_i(x_i) \qquad M := \begin{bmatrix} M_1^T & \ldots & M_n^T \end{bmatrix}^T$$

$$\zeta(t) := \begin{bmatrix} H_1 V_1^{-T} v_1(t)^T & \ldots & H_n V_n^{-T} v_n(t)^T \end{bmatrix}^T.$$

The update equations of the filter (6) in matrix form are:

$$\omega_{t+1} = (\mathcal{K} \otimes I_{d_y})\omega_t + My + \zeta(t),$$
$$\Omega_{t+1} = (\mathcal{K} \otimes I_{d_y})\Omega_t + M, \quad (11)$$

where $\mathcal{K} = [\kappa_{ij}]$ with $\kappa_{ij} = 0$ if $j \notin \mathcal{N}_i \cup \{i\}$ is a stochastic matrix. The solutions of the linear systems are:

$$\omega_t = (\mathcal{K} \otimes I_{d_y})^t \omega_0 + \sum_{\tau=0}^{t-1} (\mathcal{K} \otimes I_{d_y})^{t-1-\tau}\Big(My + \zeta(\tau)\Big),$$

$$\Omega_t = (\mathcal{K} \otimes I_{d_y})^t \Omega_0 + \sum_{\tau=0}^{t-1} (\mathcal{K} \otimes I_{d_y})^{t-1-\tau} M.$$

Looking at the i-th components again, we have:

$$\frac{\omega_{it}}{t+1} := \frac{1}{t+1}\sum_{j=1}^{n}[\mathcal{K}^t]_{ij}\omega_{j0} +$$

$$\frac{1}{t+1}\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}[\mathcal{K}^{t-\tau-1}]_{ij}(M_j y + H_j^T V_j^{-1} v_j(\tau)),$$

$$\frac{\Omega_{it}}{t+1} := \frac{1}{t+1}\sum_{j=1}^{n}[\mathcal{K}^t]_{ij}\Omega_{j0} + \frac{1}{t+1}\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}[\mathcal{K}^{t-\tau-1}]_{ij}M_j.$$

Define the following to simplify the notation:

$$g_{it} := \frac{1}{t+1}\sum_{j=1}^{n}[\mathcal{K}^t]_{ij}\omega_{j0},$$
$$G_{it} := \frac{1}{t+1}\sum_{j=1}^{n}[\mathcal{K}^t]_{ij}\Omega_{j0},$$
$$\phi_{it} := \frac{1}{t+1}\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}[\mathcal{K}^{t-\tau-1}]_{ij}H_j^T V_j^{-1}v_j(\tau), \quad (12)$$
$$C_{it} := \frac{1}{t+1}\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}[\mathcal{K}^{t-\tau-1}]_{ij}M_j,$$
$$b_{it} := g_{it} - G_{it}y, \qquad B_{it} := \frac{1}{t+1}\Omega_{it}.$$

With the shorthand notation:

$$\frac{\omega_{it}}{t+1} = g_{it}+\phi_{it}+C_{it}y, \qquad B_{it} = \frac{\Omega_{it}}{t+1} = G_{it}+C_{it}, \quad (13)$$

where ϕ_{it} is the only random quantity. Its mean is zero because the measurement noise has zero mean, while its covariance is:

$$\mathbb{E}[\phi_{it}\phi_{it}^T] = \frac{1}{(t+1)^2}\mathbb{E}\left[\left(\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}[\mathcal{K}^{t-\tau-1}]_{ij}H_j^T V_j^{-1}v_j(\tau)\right)\right.$$
$$\left. \times \left(\sum_{s=0}^{t-1}\sum_{\eta=1}^{n}[\mathcal{K}^{t-s-1}]_{i\eta}H_\eta^T V_\eta^{-1}v_\eta(s)\right)^T\right]$$
$$= \frac{1}{(t+1)^2}\sum_{j=1}^{n}\sum_{\tau=0}^{t-1}[\mathcal{K}^{t-\tau-1}]_{ij}^2 H_j^T V_j^{-1}\mathbb{E}[v_j(\tau)v_j(\tau)^T]V_j^{-1}H_j$$
$$= \frac{1}{(t+1)^2}\sum_{j=1}^{n}\sum_{\tau=0}^{t-1}[\mathcal{K}^{t-\tau-1}]_{ij}^2 M_j \preceq \frac{1}{t+1}C_{it}, \quad (14)$$

where the second equality uses the fact that $v_j(\tau)$ and $v_\eta(s)$ are independent unless the indices coincide, i.e. $\mathbb{E}v_j(\tau)v_\eta(s)^T = \delta_{\tau s}\delta_{j\eta}V_j$. The Löwner ordering inequality in the last step uses that $0 \le [\mathcal{K}^{t-\tau-1}]_{ij} \le 1$ and $M_j \succeq 0$.

Since G is connected, $\mathcal{K}$ corresponds to the transition matrix of an aperiodic irreducible Markov chain with a unique stationary distribution π so that $\mathcal{K}^t \to \pi\mathbf{1}^T$ with $\pi_j > 0$. This implies that, as $t \to \infty$, the numerators of g_{it} and G_{it} remain bounded and therefore $g_{it} \to 0$ and $G_{it} \to 0$. Since Cesáro means preserve convergent sequences and their limits:

$$\frac{1}{t+1}\sum_{\tau=0}^{t-1}[\mathcal{K}^{t-\tau-1}]_{ij} \to \pi_j, \quad \forall i,$$

which implies that $C_{it} \to \sum_{j=1}^{n}\pi_j M_j$. The full-rank assumption on $[H_1^T \ \dots \ H_n^T]^T$ and $\pi_j > 0$ guarantee that $\sum_{j=1}^{n}\pi_j M_j$ is positive definite. Finally, consider the mean squared error:

$$\mathbb{E}\left[(\hat{y}_i(t) - y)^T(\hat{y}_i(t) - y)\right]$$
$$= \mathbb{E}\left\|\left(\frac{\Omega_{it}}{t+1}\right)^{-1}\frac{\omega_{it}}{t+1} - \left(\frac{\Omega_{it}}{t+1}\right)^{-1}\left(\frac{\Omega_{it}}{t+1}\right)y\right\|_2^2$$
$$= \mathbb{E}\left\|B_{it}^{-1}\left(g_{it}+C_{it}y+\phi_{it}-(G_{it}+C_{it})y\right)\right\|_2^2$$
$$= \mathbb{E}\left\|B_{it}^{-1}(b_{it}+\phi_{it})\right\|_2^2$$
$$= \mathbb{E}\left[b_{it}^T B_{it}^{-T}B_{it}^{-1}b_{it} + 2b_{it}^T B_{it}^{-T}B_{it}^{-1}\phi_{it} + \phi_{it}^T B_{it}^{-T}B_{it}^{-1}\phi_{it}\right]$$
$$\overset{(a)}{=} b_{it}^T B_{it}^{-T}B_{it}^{-1}b_{it} + \text{tr}(B_{it}^{-1}\mathbb{E}[\phi_{it}\phi_{it}^T]B_{it}^{-T})$$
$$\overset{(b)}{\le} b_{it}^T B_{it}^{-T}B_{it}^{-1}b_{it} + \frac{1}{t+1}\text{tr}(B_{it}^{-1}C_{it}B_{it}^{-T}) \to 0,$$

where (a) holds because the first term is deterministic, while the cross term contains $\mathbb{E}[\phi_{it}] = 0$. Inequality (b) follows from (14). In the final step, as shown before $B_{it}^{-1} \to \left(\sum_{j=1}^{n}\pi_j M_j\right)^{-1}$ and $C_{it} \to \sum_{j=1}^{n}\pi_j M_j \succ 0$ remain bounded, while $b_{it} \to 0$ and $1/(t+1) \to 0$. □

APPENDIX B: PROOF OF THEOREM 2

Define the *generalized (matrix-weighted) degree matrix* $D \in \mathbb{R}^{nd\times nd}$ of the graph G as a block-diagonal matrix with $D_{ii} := \sum_{j\in\mathcal{N}_i}\mathcal{E}_{ij}^{-1}$. Since $\mathcal{E}_{ij} \succ 0$ for all $\{i,j\} \in E$, the generalized degree matrix is positive definite, $D \succ 0$. Define also the *generalized adjacency matrix* $A \in \mathbb{R}^{nd\times nd}$ as follows:

$$A_{ij} := \begin{cases} \mathcal{E}_{ij}^{-1} & \text{if } \{i,j\} \in E, \\ 0 & \text{else.} \end{cases}$$

The *generalized Laplacian* and the *generalized signless Laplacian* of G are defined as $L := D - A$ and $|L| := D + A$, respectively. Further, let $R := (B \otimes I_d)^T \in \mathbb{R}^{md\times nd}$ and define the block-diagonal matrix $\mathcal{E} \in \mathbb{R}^{md\times md}$ with blocks $\mathcal{E}_{ij}$ for $\{i,j\} \in E$. It is straightforward to verify that $L = R^T\mathcal{E}^{-1}R \succeq 0$ and $|L| = (|B| \otimes I_d)\mathcal{E}^{-1}(|B| \otimes I_d)^T \succeq 0$, where $|B| \in \mathbb{R}^{n\times m}$ is the signless incidence matrix of G. Let $\tilde{B} \in \mathbb{R}^{(n-1)\times m}$ and $\tilde{R} \in \mathbb{R}^{md\times(n-1)d}$ be the matrices resulting after removing the row corresponding to sensor 1 from B. Similarly, let $\tilde{D}, \tilde{A}, \tilde{L}, |\tilde{L}| \in \mathbb{R}^{(n-1)d\times(n-1)d}$ denote the generalized degree, adjacency, Laplacian, and signless Laplacian matrices with the row and column corresponding to sensor 1 removed. Thm. 2.2.1 in [14] shows that $\tilde{L} \succ 0$ provided that G is connected. The same approach can be used to show that $|\tilde{L}| \succ 0$. Let $\tilde{x} \in \mathbb{R}^{(n-1)d}$ be the locations of sensors $2,\dots,n$ in the reference frame of sensor 1 and $\hat{x}(t) \in \mathbb{R}^{(n-1)d}$ be their estimates at time t obtained from (9). The update in (9) can be written in matrix form as follows:

$$\tilde{D}\hat{x}(t+1) = \tilde{A}\hat{x}(t) + \tilde{R}^T\mathcal{E}^{-1}\left(\tilde{R}\tilde{x} + \frac{1}{t+1}\sum_{\tau=0}^{t}\epsilon(\tau)\right). \quad (15)$$

Define the estimation error at time t as $e(t) := \tilde{x} - \hat{x}(t)$ and let $u(t) := \frac{1}{t+1}\sum_{\tau=0}^{t}\epsilon(\tau)$. The dynamics of the error state can

be obtained from (15):

$$e(t+1) = \tilde{x} - \tilde{D}^{-1}\tilde{A}\hat{x}(t) - \tilde{D}^{-1}\tilde{L}\tilde{x} - \tilde{D}^{-1}\tilde{R}^T\mathcal{E}^{-1}u(t)$$
$$= \tilde{x} - \tilde{D}^{-1}\tilde{A}\hat{x}(t) - \tilde{D}^{-1}\left(\tilde{D} - \tilde{A}\right)\tilde{x} - \tilde{D}^{-1}\tilde{R}^T\mathcal{E}^{-1}u(t)$$
$$= \tilde{D}^{-1}\tilde{A}e(t) - \tilde{D}^{-1}\tilde{R}^T\mathcal{E}^{-1}u(t).$$

The error dynamics are governed by a stochastic linear time-invariant system, whose internal stability depends on the eigenvalues of $\tilde{D}^{-1}\tilde{A}$. To show that the error dynamics are stable, we resort to the following lemma.

Lemma 3 ([25, Lemma 4.2])**.** *Let $L = D - A \in \mathbb{C}^{n \times n}$ be such that $D + D^* \succ 0$ and $L_\theta = D + D^* - (e^{i\theta}A + e^{-i\theta}A^*) \succ 0$ for all $\theta \in \mathbb{R}$. Then $\rho(D^{-1}A) < 1$.*

Consider $\tilde{L}_\theta := 2(\tilde{D} - \cos(\theta)\tilde{A})$. If $\cos\theta = 0$, then $\tilde{L}_\theta = 2\tilde{D} \succ 0$. If $\cos\theta \in (0,1]$, then $\tilde{L}_\theta \succeq 2\cos\theta\tilde{L} \succ 0$. Finally, if $\cos\theta \in [-1,0)$, then $\tilde{L}_\theta \succeq 2|\cos\theta||\tilde{L}| \succ 0$. Thus, we can conclude that $\rho(\tilde{D}^{-1}\tilde{A}) < 1$. The proof of the theorem is completed by the following lemma with $\mathsf{F} := \tilde{D}^{-1}\tilde{A}$ and $\mathsf{G} := -\tilde{D}^{-1}\tilde{R}^T\mathcal{E}^{-1}$.

Lemma 4. *Consider the discrete-time stochastic linear time-invariant system:*

$$e(t+1) = \mathsf{F}e(t) + \mathsf{G}\frac{1}{t+1}\sum_{\tau=0}^{t}\epsilon(\tau) \qquad (16)$$

driven by Gaussian noise $\epsilon(\tau) \sim \mathcal{N}(0,\mathcal{E})$, which is independent at any pair of times. If the spectral radius of F satisfies $\rho(\mathsf{F}) < 1$, then $e(t) \xrightarrow{a.s.,L^2} 0$.

Proof. By the strong law of large numbers [26, Thm.2.4.1], $u(t) := \frac{1}{t+1}\sum_{\tau=0}^{t}\epsilon(\tau)$ converges to 0 almost surely. Let Ω be the set with measure 1 on which $u(t)$ converges so that for any $\gamma > 0$, $\exists T \in \mathbb{N}$ such that $\forall t \geq T$, $\|u(t)\| \leq \gamma$. For realizations in Ω, the solution to (16) with initial time T is:

$$e(t) = \mathsf{F}^{t-T}e(T) + \sum_{\tau=T}^{t-1}\mathsf{F}^{t-\tau-1}\mathsf{G}u(\tau).$$

Then, $\|e(t)\| \leq \|\mathsf{F}^{t-T}e(T)\| + \sum_{\tau=T}^{t-1}\|\mathsf{F}^{t-\tau-1}\|\|\mathsf{G}\|\gamma$. Taking the limit of t and using that F is stable, we have

$$\lim_{t\to\infty}\|e(t)\| \leq \left(\sum_{\tau=0}^{\infty}\|\mathsf{F}^\tau\|\right)\|\mathsf{G}\|\gamma.$$

Since $\rho(\mathsf{F}) < 1$, the system is internally (uniformly) exponentially stable, which is equivalent to $\sum_{\tau=0}^{\infty}\|\mathsf{F}^\tau\| \leq \beta$ for some finite constant β [27, Ch.22]. Thus, $\lim_{t\to\infty}\|e(t)\| \leq \beta\|\mathsf{G}\|\gamma$, which can be made arbitrarily small by choice of γ. We conclude that $e(t) \to 0$ on Ω and consequently $e(t) \xrightarrow{a.s.} 0$.

Next, we show convergence in L^2. First, consider the propagation of the cross term $C(t) := (t+1)\mathbb{E}e(t)u(t)^T$. Note that $\mathbb{E}u(t) = 0$ and $\mathbb{E}u(t)u(t)^T = \frac{\mathcal{E}}{t+1}$. Using the fact that $\epsilon(t+1)$ is independent of $e(t)$ and $u(t)$ we have

$$C(t+1) = \mathbb{E}\big(\mathsf{F}e(t) + \mathsf{G}u(t)\big)\big((t+1)u(t) + \epsilon(t+1)\big)^T$$
$$= \mathsf{F}C(t) + (t+1)\mathsf{G}\mathbb{E}u(t)u(t)^T = \mathsf{F}C(t) + \mathsf{G}\mathcal{E}.$$

The solution of the above linear time-invariant system is:

$$C(t) = \mathsf{F}^t C(0) + \sum_{\tau=0}^{t-1}\mathsf{F}^{t-\tau-1}\mathsf{G}\mathcal{E}$$

and since F is stable: $\lim_{t\to\infty}\mathbb{E}e(t)u(t)^T = \lim_{t\to\infty}\frac{1}{t+1}\sum_{\tau=0}^{t-1}\mathsf{F}^\tau\mathsf{G}\mathcal{E} = 0$.

Now, consider the second moment of the error:

$$\Sigma(t+1) := \mathbb{E}e(t+1)e(t+1)^T =$$
$$\mathsf{F}\Sigma(t)\mathsf{F}^T + \mathsf{F}\left(\mathbb{E}e(t)u(t)^T\right)\mathsf{G}^T + \mathsf{G}\left(\mathbb{E}u(t)e(t)^T\right)\mathsf{F}^T + \frac{1}{t+1}\mathsf{G}\mathcal{E}\mathsf{G}^T$$
$$= \mathsf{F}\Sigma(t)\mathsf{F}^T + Q(t),$$

where $Q(t) := \frac{1}{t+1}\left(\mathsf{F}C(t)\mathsf{G}^T + \mathsf{G}C(t)^T\mathsf{F}^T + \mathsf{G}\mathcal{E}\mathsf{G}^T\right)$. As shown above $Q(t) \to 0$ as $t \to \infty$, i.e. for any $\delta > 0$, $\exists T' \in \mathbb{N}$ such that $\forall t \geq T'$, $\|Q(t)\| \leq \delta$. With initial time T',

$$\Sigma(t) = \mathsf{F}^{t-T'}\Sigma(T')(\mathsf{F}^T)^{t-T'} + \sum_{\tau=T'}^{t-1}\mathsf{F}^{t-\tau-1}Q(\tau)(\mathsf{F}^T)^{t-\tau-1}$$

for $t \geq T'$. Then:

$$\|\Sigma(t)\| \leq \|\mathsf{F}^{t-T'}\|^2\|\Sigma(T')\| + \sum_{\tau=0}^{t-T'-1}\|\mathsf{F}^\tau\|^2\delta$$
$$\leq \alpha^2\mu^{2(t-T')} + \delta\alpha^2\sum_{\tau=0}^{t-T'-1}\mu^{2\tau},$$

where the existence of the constants $\alpha > 0$ and $0 \leq \mu < 1$ is guaranteed by the stability of F. We conclude that $\lim_{t\to\infty}\|\Sigma(t)\| \leq \frac{\delta\alpha^2}{1-\mu^2}$, which can be made arbitrarily small by choice of δ. In other words, $e(t) \xrightarrow{L^2} 0$. $\square$

APPENDIX C: PROOF OF THEOREM 3

We use the same notation and follow the same steps as in the proof of Thm. 1, except that now the terms $H_i, V_i, M_i, M, \zeta(t), \phi_{it}, C_{it}, B_{it}$ are time-varying and stochastic because they depend on the location estimates $\hat{x}_i(t)$. To emphasize this, we denote them by $\widehat{H}_{it}, \widehat{V}_{it}, \widehat{M}_{it}, \widehat{M}_t, \widehat{\zeta}(t), \widehat{\phi}_{it}, \widehat{C}_{it}, \widehat{B}_{it}$, where for example $\widehat{M}_{it} := M_i(\hat{x}_i(t))$. The same linear systems (11) describe the evolutions of ω_t and Ω_t except that they are stochastic now and (13) becomes:

$$\frac{\omega_{it}}{t+1} = g_{it} + \widehat{C}_{it}y + \widehat{\phi}_{it}, \qquad \widehat{B}_{it} := \frac{\Omega_{it}}{t+1} = G_{it} + \widehat{C}_{it}.$$

We still have that $\mathcal{K}^t \to \pi\mathbf{1}^T$ with $\pi_j > 0$. Also, g_{it}, G_{it}, and b_{it} are still deterministic and converge to zero as $t \to \infty$. The following observations are necessary to conclude that $\widehat{C}_{it}$ still converges to $\sum_{j=1}^{n}\pi_j M_j$.

Lemma 5. *If $\hat{x}_i(t) \xrightarrow{a.s.} x_i$, then $\widehat{M}_{it} \xrightarrow{a.s.,L^2} M_i$.*

Proof. Almost sure convergence follows from the continuity of $M_i(\cdot)$ and the continuous mapping theorem [26, Thm.3.2.4]. L^2-convergence follows from the boundedness of $M_i(\cdot)$ and the dominated convergence theorem [26, Thm.1.6.7]. $\square$

Lemma 6. *If $a_t \to a$ and $b_t \to b$, then $\frac{1}{t}\sum_{\tau=0}^{t-1} a_{t-\tau}b_\tau \to ab$.*

Proof. The convergence of a_t implies its boundedness, $|a_t| \le q < \infty$. Then, notice $ab = \frac{1}{t}\sum_{\tau=0}^{t-1} ab$ and

$$\left|\frac{1}{t}\sum_{\tau=0}^{t-1} a_{t-\tau}b_\tau - ab\right| = \left|\frac{1}{t}\sum_{\tau=0}^{t-1}\big(a_{t-\tau}(b_\tau - b) + (a_{t-\tau} - a)b\big)\right|$$

$$\le \left|\frac{1}{t}\sum_{\tau=0}^{t-1} a_{t-\tau}(b_\tau - b)\right| + \left|\frac{1}{t}\sum_{\tau=0}^{t-1}(a_{t-\tau} - a)b\right|$$

$$\le \left|q\left(\frac{1}{t}\sum_{\tau=0}^{t-1} b_\tau - b\right)\right| + \left|\left(\frac{1}{t}\sum_{\tau=1}^{t} a_\tau - a\right)b\right|,$$

where both terms converge to zero since Cesáro means preserve convergent sequences and their limits. $\square$

Combining Lemma 5, $[\mathcal{K}^t]_{ij} \to \pi_j$, and Lemma 6, we have:

$$\frac{1}{t+1}\sum_{\tau=0}^{t-1}[\mathcal{K}^{t-\tau-1}]_{ij}\widehat{M}_{j\tau} \xrightarrow{a.s.} [\pi\mathbf{1}^T]_{ij}M_j = \pi_j M_j.$$

Moreover, $0 \le [\mathcal{K}^t]_{ij} \le 1$ and the boundedness of $\widehat{M}_{jt}$ imply, by the bounded convergence theorem [26, Thm.1.6.7], that the sequence above converges in L^2 as well:

$$\widehat{C}_{it} \xrightarrow{a.s.,L^2} \sum_{j=1}^{n}\pi_j M_j \succ 0. \tag{17}$$

In turn, (17) guarantees that:

$$\widehat{B}_{it}^{-2} = (G_{it} + \widehat{C}_{it})^{-2} \xrightarrow{a.s.} \left(\sum_{j=1}^{n}\pi_j M_j\right)^{-2} \tag{18}$$

but is not enough to ensure that $\mathbb{E}\big[\widehat{B}_{it}^{-2}\big]$ remains bounded as $t \to \infty$. The parameter $\delta > 0$ is needed to guarantee the boundedness. In particular, define $\widehat{B}_{it}(\delta) := \widehat{B}_{it} + \delta I_{d_y}$. Then

$$\widehat{B}_{it}(\delta)^{-2} = (G_{it} + \widehat{C}_{it} + \delta I_{d_y})^{-2} \prec \frac{1}{\delta^2}I_{d_y}$$

and by the bounded convergence theorem and (18):

$$\widehat{B}_{it}(\delta)^{-2} \xrightarrow{a.s.,L^1} \left(\sum_{j=1}^{n}\pi_j M_j + \delta I_{d_y}\right)^{-2}, \tag{19}$$

so that $\lim_{t\to\infty}\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-2}\big] < \infty$. From (17) and the boundedness of $\widehat{B}_{it}(\delta)^{-1}$ and $\widehat{C}_{it}$, we also have:

$$\widehat{B}_{it}(\delta)^{-1}\widehat{C}_{it}\widehat{B}_{it}(\delta)^{-T} \xrightarrow{a.s.,L^2} \tag{20}$$

$$\left(\sum_{j=1}^{n}\pi_j M_j + \delta I_{d_y}\right)^{-1}\left(\sum_{j=1}^{n}\pi_j M_j\right)\left(\sum_{j=1}^{n}\pi_j M_j + \delta I_{d_y}\right)^{-T}.$$

Since $\widehat{H}_{it}$ and $\widehat{V}_{it}$ depend solely on $\hat{x}_i(t)$, they are independent of $v_i(t)$. Because $\mathbb{E}[v_j(\tau)] = 0$, $\mathbb{E}[\widehat{H}_{j\tau}^T\widehat{V}_{j\tau}^{-1}v_j(\tau)] = 0$ and as before $\mathbb{E}[\widehat{\phi}_{it}] = 0$. Since $\widehat{B}_{it}(\delta)$ is independent of $v_i(t)$ as well, $\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-2}\widehat{\phi}_{it}\big] = 0$ and a result equivalent to (14) holds:

$$\mathbb{E}[\widehat{B}_{it}(\delta)^{-1}\widehat{\phi}_{it}\widehat{\phi}_{it}^T\widehat{B}_{it}(\delta)^{-T}]$$

$$= \mathbb{E}\left[\widehat{B}_{it}(\delta)^{-1}\left(\frac{1}{(t+1)^2}\sum_{j=1}^{n}\sum_{\tau=0}^{t-1}[\mathcal{K}^{t-\tau-1}]_{ij}^2\widehat{M}_{j\tau}\right)\widehat{B}_{it}(\delta)^{-T}\right]$$

$$\preceq \frac{1}{t+1}\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-1}\widehat{C}_{it}\widehat{B}_{it}(\delta)^{-T}\big]. \tag{21}$$

Finally, we can consider the mean squared error:

$$\mathbb{E}\big[\|\hat{y}_i(t) - y\|_2^2\big] = \mathbb{E}\left\|\widehat{B}_{it}(\delta)^{-1}\frac{\omega_{it}}{t+1} - \widehat{B}_{it}(\delta)^{-1}\widehat{B}_{it}(\delta)y\right\|_2^2$$

$$= \mathbb{E}\left\|\widehat{B}_{it}(\delta)^{-1}\left(g_{it} + \widehat{C}_{it}y + \widehat{\phi}_{it} - (G_{it} + \widehat{C}_{it} + \delta I_{d_y})y\right)\right\|_2^2$$

$$= \mathbb{E}\|\widehat{B}_{it}(\delta)^{-1}(b_{it} + \widehat{\phi}_{it} + \delta y)\|_2^2$$

$$= \mathbb{E}\Big[b_{it}^T\widehat{B}_{it}(\delta)^{-2}b_{it} + \widehat{\phi}_{it}^T\widehat{B}_{it}(\delta)^{-2}\widehat{\phi}_{it} + \delta^2 y^T\widehat{B}_{it}(\delta)^{-2}y$$

$$+ 2b_{it}^T\widehat{B}_{it}(\delta)^{-2}\widehat{\phi}_{it} + 2\delta y^T\widehat{B}_{it}(\delta)^{-2}\widehat{\phi}_{it} + 2\delta b_{it}^T\widehat{B}_{it}(\delta)^{-2}y\Big]$$

$$= b_{it}^T\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-2}\big]b_{it} + \mathrm{tr}(\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-1}\widehat{\phi}_{it}\widehat{\phi}_{it}^T\widehat{B}_{it}(\delta)^{-T}\big])$$

$$+ \delta^2 y^T\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-2}\big]y + 2\delta b_{it}^T\mathbb{E}\big[\widehat{B}_{it}(\delta)^{-2}\big]y$$

$$\overset{(21)}{\le} b_{it}^T\mathbb{E}\widehat{B}_{it}(\delta)^{-2}b_{it} + 2\delta b_{it}^T\mathbb{E}\widehat{B}_{it}(\delta)^{-2}y + \delta^2 y^T\mathbb{E}\widehat{B}_{it}(\delta)^{-2}y$$

$$+ \frac{1}{t+1}\mathrm{tr}\left(\mathbb{E}\left[\widehat{B}_{it}(\delta)^{-1}\widehat{C}_{it}\widehat{B}_{it}(\delta)^{-T}\right]\right)$$

$$\to \delta^2 y^T\Big(\sum_{j=1}^{n}\pi_j M_j + \delta I_{d_y}\Big)^{-2}y.$$

In the final step, the first two terms go to zero because $b_{it} \to 0$ and $\lim_t \mathbb{E}\widehat{B}_{it}(\delta)^{-2} < \infty$ from (19), the third term converges in view of (19) again, while the last term goes to zero because the trace is bounded in the limit in view of (20). $\square$

REFERENCES

[1] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless Sensor Networks for Habitat Monitoring," in *Int. Workshop on Wireless Sensor Networks and Applications*, 2002.

[2] N. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. Fratantoni, and R. Davis, "Collective Motion, Sensor Networks, and Ocean Sampling," *Proceedings of the IEEE*, vol. 95, no. 1, 2007.

[3] T. Yan, T. He, and J. Stankovic, "Differentiated Surveillance for Sensor Networks," in *Int. Conf. on Embedded Networked Sensor Systems*, 2003.

[4] J. Lu and T. Suda, "Differentiated Surveillance for Static and Random Mobile Sensor Networks," *IEEE Trans. on Wireless Communications*, vol. 7, no. 11, 2008.

[5] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian Social Learning," *Games and Economic Behavior*, vol. 76, no. 1, 2012.

[6] K. Rahnama Rad and A. Tahbaz-Salehi, "Distributed Parameter Estimation in Networks," in *Conf. on Decision and Control (CDC)*, 2010.

[7] R. Olfati-Saber, "Distributed Kalman Filtering for Sensor Networks," in *Conf. on Decision and Control (CDC)*, 2007.

[8] ——, "Kalman-Consensus Filter: Optimality, Stability, and Performance," in *Joint Conf. on Decision and Control (CDC) and Chinese Control Conference*, 2009.

[9] U. Khan, S. Kar, A. Jadbabaie, and J. Moura, "On connectivity, observability, and stability in distributed estimation," in *Conf. on Decision and Control (CDC)*, 2010.

[10] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Online Learning of Dynamic Parameters in Social Networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[11] S. Kar, J. Moura, and K. Ramanan, "Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication," *Trans. on Information Theory*, vol. 58, no. 6, 2012.

[12] J. Cortés, "Distributed Kriged Kalman filter for spatial estimation," *Trans. on Automatic Control (TAC)*, vol. 54, no. 12, 2009.

[13] P. Barooah and J. Hespanha, "Estimation on Graphs from Relative Measurements," *IEEE Control Systems*, vol. 27, no. 4, 2007.

[14] P. Barooah, "Estimation and Control with Relative Measurements: Algorithms and Scaling Laws," Ph.D. dissertation, University of California Santa Barbara, 2007.

[15] J. Aspnes, T. Eren, D. Goldenberg, A. Morse, W. Whiteley, Y. Yang, B. Anderson, and P. Belhumeur, "A Theory of Network Localization," *IEEE Trans. on Mobile Computing*, vol. 5, no. 12, 2006.

[16] D. Moore, J. Leonard, D. Rus, and S. Teller, "Robust Distributed Network Localization with Noisy Range Measurements," in *Int. Conf. on Embedded Networked Sensor Systems*, 2004.

[17] A. So and Y. Ye, "Theory of semidefinite programming for Sensor Network Localization," *Mathematical Programming*, vol. 109, no. 2-3, 2007.

[18] N. Priyantha, H. Balakrishnan, E. Demaine, and S. Teller, "Anchor-Free Distributed Localization in Sensor Networks," in *Int. Conf. on Embedded Networked Sensor Systems*, 2003.

[19] U. Khan, S. Kar, and J. Moura, "Distributed Sensor Localization in Random Environments Using Minimal Number of Anchor Nodes," *Trans. on Signal Processing*, vol. 57, no. 5, 2009.

[20] Y. Diao, Z. Lin, M. Fu, and H. Zhang, "A New Distributed Localization Method for Sensor Networks," in *Asian Control Conf. (ASCC)*, 2013.

[21] R. Tron and R. Vidal, "Distributed Image-Based 3-D Localization in Camera Sensor Networks," in *Joint Conf. on Decision and Control (CDC) and Chinese Control Conference*, 2009.

[22] V. Hernandez Bennetts, A. Lilienthal, A. Khaliq, V. Pomareda Sese, and M. Trincavelli, "Towards Real-World Gas Distribution Mapping and Leak Localization Using a Mobile Robot with 3D and Remote Gas Sensing Capabilities," in *Int. Conf. on Robotics and Automation*, 2013.

[23] A. Barker, D. Brown, and W. Martin, "Bayesian estimation and the Kalman filter," *Computers and Mathematics with Applications*, vol. 30, no. 10, 1995.

[24] J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Prentice-Hall, 1995.

[25] L. Elsner and V. Mehrmann, "Convergence of block iterative methods for linear systems arising in the numerical solution of Euler equations," *Numerische Mathematik*, vol. 59, no. 1, 1991.

[26] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge University Press, 2010.

[27] W. Rugh, *Linear System Theory*, 2nd ed. Prentice-Hall, 1996.