Distributed Active State Estimation with User-Specified Accuracy

Charles Freundlich, *Student Member, IEEE,* Soomin Lee, *Member, IEEE,*
and Michael M. Zavlanos, *Member, IEEE*

Abstract—In this paper, we address the problem of controlling a network of mobile sensors so that a set of hidden states are estimated up to a user-specified accuracy. The sensors take measurements and fuse them online using an Information Consensus Filter (ICF). At the same time, the local estimates guide the sensors to their next best configuration. This leads to an LMI-constrained optimization problem that we solve by means of a new distributed random approximate projections method. The new method is robust to the state disagreement errors that exist among the robots as the ICF fuses the collected measurements. Assuming that the noise corrupting the measurements is zero-mean and Gaussian and that the robots are self localized in the environment, the integrated system converges to the next best positions from where new observations will be taken. This process is repeated with the robots taking a sequence of observations until the hidden states are estimated up to the desired user-specified accuracy. We present simulations of sparse landmark localization, where the robotic team achieves the desired estimation tolerances while exhibiting interesting emergent behavior.

Index Terms—Distributed optimization, Estimation, Cooperative control, Sensor networks

I. INTRODUCTION

IN this paper, we control a robotic sensor network to estimate a collection of hidden states so that desired accuracy thresholds and confidence levels are satisfied. We require that estimation and control are completely distributed, i.e., belief states and control actions are decided upon locally through pairwise communication among robots in the network.

A common assumption in problems like the one discussed herein is that observations depend linearly on the state and are corrupted by Gaussian noise as well as that the sensors are self localized [1]–[8]. Under these assumptions, a typical approach to estimate a set of hidden states is to use an Information Filter (IF). The role of the IF is to fuse new observations of the hidden states with a prior Gaussian distribution to create a minimum-variance posterior Gaussian distribution. Typically, the observations are subject to error covariance that is a function of, e.g., viewing position, angle of incidence, and the sensing modality. Several such models have been proposed in the distributed active sensing literature, see, e.g., [1]–[3], [5]–[8], but regardless of the specific model used for the measurement covariance, the IF combines the data among the nodes in the network in an additive way, making the IF

algorithm readily distributed; see, e.g., Information Consensus Filtering (ICF) [9].

The dependence of the measurement model on the sensor state has been widely used to obtain planning algorithms that, given a history of measurements, determine the next best set of observations of a hidden state. The objective is that this new set of observations optimizes an information-theoretic criterion of interest. In this paper, we choose to maximize the minimum eigenvalue of the posterior information matrix, effectively minimizing the directional variance of every hidden state. We express this requirement by reformulating the problem as a Linear Matrix Inequality (LMI) constrained optimization problem. Specifically, we introduce as many LMI constraints as the number of hidden states, that are coupled with respect to the sensors via an ICF that estimates the uncertainty of each hidden state using the sensor measurements. To solve this problem, we propose a new distributed optimization algorithm, which we call *random approximate projections*, that is robust to the state disagreement errors that exist among the robots as an ICF fuses the collected measurements. The ability to handle such errors is critical in combining distributed estimation and planning in one process.

The proposed distributed random approximate projections method falls in the class of so called consensus-based optimization algorithms, where the goal is to interleave a decentralized averaging step between local optimization steps so that the local estimates of all agents simultaneously obtain a network-wide consensus and optimality of the global problem. Existing algorithms require expensive optimization steps or projections on a complicated constraint set, e.g., sets of LMI constraints as in this paper, at every iteration. Instead, our method employs a subgradient step in the direction that minimizes the violation of randomly selected local constraints, significantly reducing computational cost compared to relevant methods that rely on projection operations. Assuming that each LMI constraint is selected with non-zero probability, we show that our proposed algorithm converges almost surely to the next best sensor positions from where a new set of observations minimizes the worst-case uncertainty of the hidden states. This process is repeated with the sensors taking a sequence of observations until all the hidden states are estimated up to the desired user-specified accuracy. For details on relevant distributed optimization methods, we refer the interested reader to [10] and references therein.

Related to our random projection technique is the Markov incremental random subgradient method [11]. The key difference is in the mechanism that is used to distribute the

computations. In the Markov incremental algorithm, the agents sequentially update a single iterate sequence by passing the iterate to each other following a time inhomogeneous Markov chain. In our algorithm, each agent maintains a local iterate sequence and updates this by communicating with its neighbors. In two recent papers [12], [13], a dual-decomposition method is proposed that can be used for problems similar to the distributed control problem considered here, with lower communication overhead and increased privacy. Generally, dual methods are well suited for network optimization problems, they come at increased local cost because of the need to compute the dual function at each node. The approaches described in [12], [13] additionally require a special problem structure; specifically, a decomposable objective and/or a star network topology.

Applications in the area of distributed sensor planning and estimation that are directly relevant to our work include SLAM [14], localization [4]–[6], coverage [7], [8], mobile target tracking [1]–[3], [15], classification [16], and inverse problems for PDE-constrained systems [17], [18]. Compared to state estimation, inverse problems for PDE-constrained systems address both state and source identification and are particularly difficult to solve. A common characteristic of the applications discussed above is that the sensors must collectively reason over the possible posterior distributions resulting from (i) multiple hidden states being observed by a single sensor or (ii) individual hidden states being observed by several sensors at once. Without assuming Gaussian distributions, optimization over the possible posteriors can only be done approximately [16]. In this paper, and in much of the relevant literature [1]–[6], it is assumed that the posteriors for the hidden states are Gaussian. Under this assumption, typical choices of an objective function to minimize in order to obtain the next best set of observations include the trace, the determinant, or the maximum eigenvalue of the posterior covariance matrix. Intuitively, the trace minimizes the sum of the uncertainties of the hidden states, the determinant the volume of the confidence ellipsoids of the hidden states, and the maximum eigenvalue the worst-case error. To our knowledge, in the context of the distributed active localization problem, existing works that control the maximum eigenvalue of the error covariance matrix either focus on sensor selection from a discrete set [12], [13], [19]–[21], are limited to two sensors [6], or do not factor in simultaneous measurements by other sensors when deciding where to go [2], [4], [22].

Methods that account for nonlinear dynamics and arbitrary error distributions [15], [23] are also relevant to the problem addressed herein. In these cases, updates to the hidden state pdfs may no longer be available in closed form, and so numerical methods are needed to evaluate the expectation integrals. To account for uncertainty in the sensor location and actions, decentralized Partially Observable Markov Decision Process (dec-POMDP) can be used to solve these problems. Dec-POMDPs have coupled value functions among the agents, and typically require a centralized planner [24]. At the same time, comparing the possible posterior distributions can be computationally expensive, so these methods do not usually scale well with the number of targets and agents. Although many of these works assume that the robots are self localized with respect to a common global reference frame [1]–[4], [6], [15], [16], this is certainly not always the case; see, e.g., work on partially observable systems [23], [24] or the SLAM problem [14].

Contributions: In this paper we propose a distributed estimation and sensor planning method that can ensure desired estimation tolerances for large numbers of hidden states. To our knowledge, even though the distributed active sensing literature is well-developed, the ability to control worst-case estimation uncertainty in a decentralized way is new. This is possible by combining an ICF for distributed data fusion with an LMI-constrained optimization problem for sensor planning, for which we propose a new computationally inexpensive distributed method that is robust to the inexact pre-consensus filter data. Coordination among the sensors requires only weak connectivity of the communication graph and can, thus, handle occasional disruption of communication links.

The distributed planning algorithm itself is an extension of the authors' previous work [25] and [26]. The difference is that the method developed in this paper is robust to state disagreement errors that appear as a result of the information filter. The ability to handle such errors is critical in integrating distributed estimation and planning. To the best of our knowledge, there are no distributed optimization methods that can handle efficiently LMI constraints under inexact data. It is also worth mentioning that a wide variety of convex constraints arising in control, such as quadratic inequalities, inequalities involving matrix norms, Lyapunov inequalities and quadratic matrix inequalities, can be cast as LMIs. Therefore, our developed algorithm can be used for other, potentially unrelated, control problems involving inexact data.

The rest of the paper is organized as follows. Section II formulates the problem under consideration. Then, Section III develops the proposed distributed method to maximize the minimum eigenvalue of the information matrix. Section IV states assumptions and the main convergence result of our proposed algorithm. Section V shows that the proposed algorithm converges to the optimal point. Section VI provides simulations, and Section VII concludes the paper.

II. PROBLEM FORMULATION

Consider the problem of estimating m stationary hidden states $\{\mathbf{x}_i \in \mathbb{R}^p\}_{i \in \mathcal{I}}$ using noisy observations from n mobile sensors, where p is the dimension of the hidden states and $\mathcal{I} = \{1, \ldots, m\}$. We assume that the hidden states are sparse so that there are no correlations between them. Denote the locations of the mobile sensors at time t by $\{\mathbf{r}_s(t) \in \mathbb{R}^q\}_{s \in \mathcal{S}}$, where q is the dimension of the sensors' configuration space and $\mathcal{S} = \{1, \ldots, n\}$. The observation of state i by sensor s at time t is given by

$$\mathbf{y}_{i,s}(t) = \mathbf{x}_i + \boldsymbol{\zeta}_{i,s}(t). \tag{1}$$

We assume that the measurement errors are Normally distributed so that $\boldsymbol{\zeta}_{i,s}(t) \sim N\big(\mathbf{0}, Q(\mathbf{r}_s(t), \mathbf{x}_i)\big)$, where $Q \colon \mathbb{R}^{q+p} \to \mathrm{Sym}_+(p, \mathbb{R})$ denotes the measurement precision matrix, or information matrix (not the covariance), and

$\mathrm{Sym}_+(p, \mathbb{R})$ is the set of $p \times p$ symmetric positive semidefinite real matrices. We also assume that if signal $\mathbf{x}_i$ is out of range of sensor s, then the function Q returns the information matrix $\mathbf{0}_{p \times p}$, corresponding to infinite variance. We provide an example of the measurement information function Q in Section VI. Denote the set of all observations at time t by $\mathcal{O}(t) \triangleq \{(\mathbf{y}_{i,s}(t), Q(\mathbf{r}_s(t), \mathbf{x}_i))\}_{i \in \mathcal{I}, s \in \mathcal{S}}$

Moreover, let $\hat{\mathbf{x}}_i(t)$ denote the estimate of $\mathbf{x}_i$ at time t and $S_i(t)$ denote the information matrix corresponding to the estimate $\hat{\mathbf{x}}_i(t)$. We assume that initial values for the pair $(\hat{\mathbf{x}}_i(0), S_i(0))$ are available *a priori*. Let also $\hat{\mathbf{x}}(t) \in \mathbb{R}^{pm}$ denote the stack vector of all estimates $\hat{\mathbf{x}}_i(t)$ and $S(t) \in \mathrm{Sym}_+(p, \mathbb{R}) \times \cdots \times \mathrm{Sym}_+(p, \mathbb{R}) := (\mathrm{Sym}_+(p, \mathbb{R}))^m$ denote the collection of all information matrices $S_i(t)$ at time t. Then, the global data at time t can be captured by the set $\mathcal{D}(t) \triangleq (\hat{\mathbf{x}}(t), S(t)) \in \mathbb{R}^{pm} \times (\mathrm{Sym}_+(p, \mathbb{R}))^m$.

Given prior data $\mathcal{D}(t-1)$ known by all sensors, and observations $\mathcal{O}(t-1)$, the current set of data $\mathcal{D}(t)$ can be determined in a distributed way using an Information Consensus Filter (ICF) [9] as[1]

$$(\mathcal{D}(t-1), \mathcal{O}(t-1)) \overset{\mathrm{ICF}}{\longmapsto} \mathcal{D}(t). \qquad (2)$$

Then, the goal is to control the next set of observations $\mathcal{O}(t)$ in order to reduce the future uncertainty in $\mathcal{D}(t+1)$. We can achieve this goal by controlling the mobile sensors and, therefore, the expected information $\{Q(\mathbf{r}_s(t), \mathbf{x}_i(t))\}_{i \in \mathcal{I}, s \in \mathcal{S}}$ associated with the observations $\{\mathbf{y}_{i,s}(t)\}_{i \in \mathcal{I}, s \in \mathcal{S}}$. Specifically, let $\mathbf{r}_s(t) = \mathbf{r}_s(t-1) + \mathbf{u}_s$ denote the motion model of sensor s, where $\mathbf{u}_s \in \mathbb{R}^q$ is a candidate control input used to drive sensor s to a new position $\mathbf{r}_s(t)$. Then, our goal is to maximize the minimum eigenvalues of the information matrices in $\mathcal{D}(t+1)$ that, for each hidden state i, are given by

$$S_i(t+1) = \overbrace{S_i(t)}^{\text{from } \mathcal{D}(t)} + \underbrace{\sum_{s \in \mathcal{S}} Q(\mathbf{r}_s(t-1) + \mathbf{u}_s, \overbrace{\hat{\mathbf{x}}_i(t)}^{\text{from } \mathcal{D}(t)})}_{\text{from } \mathcal{O}(t)}. \qquad (3)$$

We assume that $\mathbf{u}_s$ lies in a compact set containing the origin. Denote this set by $\mathcal{U}_s$. Note that there is a $\delta > 0$ such that the closed ball of radius δ contains $\mathcal{U}_s$. Define $\mathcal{U} \triangleq \prod_{s \in \mathcal{S}} \mathcal{U}_s$ and let $\mathbf{u} \in \mathbb{R}^{qn}$ denote the stack of all controllers $\mathbf{u}_s$. Additionally, define the set $\Gamma \triangleq \prod_{i \in \mathcal{I}} [0, \tau_i]$, where τ_i represent the given user-specified estimation thresholds for each hidden state i. Define also $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m) \in \mathbb{R}^m$. With the new notation, the problem we consider in this paper can be defined as

$$\max_{(\boldsymbol{\gamma}, \mathbf{u}) \in \Gamma \times \mathcal{U}} \sum_{i \in \mathcal{I}} \gamma_i \qquad (4\mathrm{a})$$

$$\text{s.t} \quad \gamma_i I \preceq S_i + \sum_{s \in \mathcal{S}} Q(\mathbf{r}_s + \mathbf{u}_s, \hat{\mathbf{x}}_i) \ \forall i \in \mathcal{I}, \qquad (4\mathrm{b})$$

where we have dropped dependence on the time t for notational convenience. For the explicit references to time in (4b) see (3). The symbol $\preceq$ denotes an ordering on the negative semidefinite cone, i.e., $A \preceq B$ if and only if $A - B$ is negative semidefinite. The great challenge in solving (4) is that the

[1]Explicit details on the mapping (2) are provided in Section III.

constraints (4b) require knowledge of $\mathcal{D}(t)$, however, only $\mathcal{D}(t-1)$ and those parts of $\mathcal{O}(t-1)$ corresponding to local sensor measurements are available to any given sensor. In what follows, we propose an integrated approach that combines the ICF to determine the global data $\mathcal{D}(t)$ and the solution of the optimization (4) in one process. We show that our proposed method can handle pre-consensus disagreement errors due to the ICF and converges to the next best set of observations $\mathcal{O}(t)$, as desired.

III. CONCURRENT ESTIMATION AND SENSOR PLANNING

Problem (4) is a nonlinear semidefinite program. To simplify this problem, we replace the nonlinear function $Q(\mathbf{r}_s + \mathbf{u}_s, \hat{\mathbf{x}}_i)$ in (4b) with its Taylor expansion about the input $\mathbf{u}_s = 0$. Specifically, define the function

$$h : \mathbb{R}^{m+qn} \times (\mathbb{R}^{pm} \times (\mathrm{Sym}_+(p, \mathbb{R}))^m) \times \mathcal{I} \to \mathrm{Sym}(p, \mathbb{R}),$$

where $\mathrm{Sym}(p, \mathbb{R})$ denotes the set of symmetric $p \times p$ dimensional matrices, by

$$h((\boldsymbol{\gamma}, \mathbf{u}); \mathcal{D}, i) \triangleq \gamma_i I - S_i \qquad (5)$$
$$- \sum_{s \in \mathcal{S}} \left[Q(\mathbf{r}_s, \hat{\mathbf{x}}_i) + \sum_{j=1}^q \nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)[\mathbf{u}_s]_j \right].$$

The notation $\nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)$ is the partial derivative of the matrix-valued function Q with respect to the j-th coordinate of $\mathbb{R}^{q+p}$, evaluated at $(\mathbf{r}_s, \hat{\mathbf{x}}_i)$. Also, $[\mathbf{u}_s]_j$ denotes the j-th coordinate of the vector $\mathbf{u}_s$, and the semicolon notation in the arguments of h separates the decision variables $(\boldsymbol{\gamma}, \mathbf{u})$ from the parameters $\mathcal{D}$ and i. Note that the value of h is negative semidefinite at a point $((\boldsymbol{\gamma}, \mathbf{u}); \mathcal{D}, i)$ if and only if the linearized version of the constraint (4b) is satisfied for γ_i using the data $\mathcal{D} = (\hat{\mathbf{x}}, S)$. The effect of the linear approximation is investigated in Section VI, Fig. 4.

To simplify notation, collect all decision variables in the vector $\mathbf{z} = (\boldsymbol{\gamma}, \mathbf{u}) \in \mathbb{R}^{m+qn}$. Then, to decompose the linearized version of (4) over the set of sensors so that the problem can be solved in a distributed way, let $\mathbf{z}_s$ denote a copy of $\mathbf{z}$ that is local to sensor $s \in \mathcal{S}$. Additionally, let $f(\mathbf{z}_s) = -\sum_{i \in \mathcal{I}} \gamma_{i,s}$ so that we may express the global objective as $\sum_{s \in \mathcal{S}} f(\mathbf{z}_s)$. Then, problem (4) with linearized constraints can be expressed as

$$\min_{\{\mathbf{z}_s\}_{s \in \mathcal{S}} \subset \mathcal{X}_0} \sum_{s \in \mathcal{S}} f(\mathbf{z}_s) \qquad (6\mathrm{a})$$

$$\text{s.t} \quad h(\mathbf{z}_s; \mathcal{D}, i) \preceq \mathbf{0} \quad \forall i \in \mathcal{I}, \forall s \in \mathcal{S}, \qquad (6\mathrm{b})$$

where $\mathcal{X}_0 = \Gamma \times \mathcal{U}$. In problem (6), the objective function (6a) is separable among the sensors and the constraints (6b) are linear in $\mathbf{z}_s$ and local to every sensor. In this form, problem (6) can be decomposed and solved in a distributed way. A necessary requirement is that, at the solution, the local variables $\mathbf{z}_s$ need to agree for all sensors s, i.e., $\mathbf{z}_s = \mathbf{z}_v$ for all $s, v \in \mathcal{S}$. As discussed in Section II, the main challenge in solving problem (6) is that the global data in $\mathcal{D}$ are not known to the sensors and need to be computed using the ICF concurrently with the optimization of the decision variables $\mathbf{z}_s$. The ICF is an iterative process by which every sensor updates a local copy of the global data until all sensors agree on the true values

Algorithm 1 Concurrent Filtering and Optimization for $s \in \mathcal{S}$ during the interval $[t-1, t]$

Require: Data $\mathcal{D}_s(t-1)$, measurements $\{\mathbf{y}_{i,s}(t-1)\}_{i \in \mathcal{I}}$, and $\mathbf{z}_{s,0}$ feasible for (6).
 1: $(\hat{\mathbf{x}}, S) \leftarrow \mathcal{D}_s(t-1)$
 2: $\Xi_{i,s,0} \leftarrow \frac{1}{n} S_i + Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)$ for all $i \in \mathcal{I}$
 3: $\boldsymbol{\xi}_{i,s,0} \leftarrow \left(\frac{1}{n} S_i + Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)\right) \mathbf{y}_{i,s}$ for all $i \in \mathcal{I}$
 4: **for** $k = 1, 2, \ldots$ **do**
 5: $\quad$ Broadcast $\mathbf{z}_{s,k-1}, \{\Xi_{i,s,k-1}, \boldsymbol{\xi}_{i,s,k-1} \mid i \in \mathcal{I}\}$ to $\mathcal{N}_{s,k}$
 6: $\quad$ $\mathbf{p}_{s,k} \leftarrow \sum_{j \in \mathcal{N}_{s,k}} [W_k]_{sj} \mathbf{z}_{s,k-1}$
 7: $\quad$ $\Xi_{i,s,k} \leftarrow \sum_{j \in \mathcal{N}_{s,k}} [W_k]_{sj} \Xi_{i,s,k-1}$ for all $i \in \mathcal{I}$
 8: $\quad$ $\boldsymbol{\xi}_{i,s,k} \leftarrow \sum_{j \in \mathcal{N}_{s,k}} [W_k]_{sj} \boldsymbol{\xi}_{i,s,k-1}$ for all $i \in \mathcal{I}$
 9:
$$\mathcal{D}_{s,k} \leftarrow \Big(\big((n\Xi_{1,s,k})^{-1} \boldsymbol{\xi}_{1,s,k}, \ldots, (n\Xi_{m,s,k})^{-1} \boldsymbol{\xi}_{m,s,k} \big),$$
$$(n\Xi_{1,s,k}, \ldots, n\Xi_{m,s,k}) \Big)$$

10: $\quad$ $\mathbf{v}_{s,k} \leftarrow \Pi_{\mathcal{X}_0} \left(\mathbf{p}_{s,k} - \alpha_k f'(\mathbf{p}_{s,k}) \right)$
11: $\quad$ Choose random $\omega_{s,k} \in \mathcal{I}$, compute $h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$ and $h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$
12: $\quad$ $\beta_{s,k} \leftarrow \frac{h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{\left\| h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \right\|^2}$
13: $\quad$ $\mathbf{z}_{s,k} \leftarrow \Pi_{\mathcal{X}_0} \left(\mathbf{v}_{s,k} - \beta_{s,k} h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \right)$
14: **end for**
15: **return** Updated data $\mathcal{D}_s(t) \leftarrow \mathcal{D}_{s,k}$ and control input $\mathbf{u}_s$, taken from $\mathbf{z}_{s,k}$

---

in $\mathcal{D}$. Therefore, before the ICF has converged, disagreement errors on the local copies of the data set $\mathcal{D}$ exist, which means that the gradients and objective function evaluations at every sensor node will not depend on the same data. In other words, gradients and objective function evaluations will be based on inexact data. An additional challenge is that the number of LMI constraints can grow very large with the number of sensors and hidden states. We address these challenges in the remainder of this section.

A. Distributed Information Filtering

We solve problem (6) using an iterative approach. Let $\mathcal{D}_{s,k}$ denote the copy of the data set $\mathcal{D}$ that is local to sensor s at iteration $k \in \mathbb{N}$, where $\mathbb{N} = \{1, 2, 3, \ldots\}$ denotes the natural numbers. Additionally, let $\mathbf{z}_{s,k}$ denote the decision variables of the sensor s at iteration k. The first step of the algorithm is for sensor s to broadcast $\mathbf{z}_{s,k-1}$ and a summary of $\mathcal{D}_{s,k-1}$. In Algorithm 1, the *data summary for state i* is captured by the intermediate information matrix $\Xi_{i,s,k}$ and information vector $\boldsymbol{\xi}_{i,s,k}$, defined in lines 2 and 3. We refer the reader to [9] for a detailed discussion of Information Consensus Filtering, but we note briefly here that if the sensors each compute $\Xi_{i,s,0}$ and $\boldsymbol{\xi}_{i,s,0}$, and carry out the weighted averaging that is standard in all consensus algorithms, given in lines 7 and 8, then $n\Xi_{i,s,k}$ and $\boldsymbol{\xi}_{i,s,k}$ converge geometrically to the consensus posterior information. In other words, under the ICF, we have that, for all $s \in \mathcal{S}$, $\{(n\Xi_{1,s,k}, \ldots, n\Xi_{m,s,k})\}_{k \in \mathbb{N}} \to S$ for all $i \in \mathcal{I}$ and $\left\{ \left((n\Xi_{1,s,k})^{-1} \boldsymbol{\xi}_{1,s,k}, \ldots, (n\Xi_{m,s,k})^{-1} \boldsymbol{\xi}_{m,s,k} \right) \right\}_{k \in \mathbb{N}} \to \hat{\mathbf{x}}$. In particular, let $\mathcal{G}_k = (\mathcal{S}, \mathcal{E}_k)$ represent a communication

graph such that $(s, v) \in \mathcal{E}_k$ if and only if sensors s and v can communicate at iteration k. This defines the *neighbor set*, $\mathcal{N}_{s,k} \triangleq \{s\} \cup \{v \in \mathcal{S} \mid (v, s) \in \mathcal{E}_k\}$. Sensor s receives $\left\{ \mathbf{z}_{v,k-1}, \{\boldsymbol{\xi}_{i,v,k}, \Xi_{i,v,k}\}_{i \in \mathcal{I}} \mid v \in \mathcal{N}_{s,k} \right\}$. With this new information, the sensor updates its own optimization variable and data summary via using weighted averaging. In particular, let W_k be an $n \times n$ row stochastic matrix such that $[W_k]_{sv} > 0 \iff v \in \mathcal{N}_s$. The exact assumptions for W_k will be introduced later, when they are necessary for the proof of our main result. The consensus steps are depicted in lines 6, 7, and 8 of Algorithm 1. Then, the agent can compute $\mathcal{D}_{s,k}$ as shown in line 9, concluding the IF portion of iteration k. Note that, given initial conditions $\{\mathcal{D}_{s,0}\}_{s \in \mathcal{S}}$, the data for any sensor s at time k is bounded by the initial conditions as

$$\|\mathcal{D}_{s,k}\| \leq \max_{v \in \mathcal{S}} \|(\mathcal{D}_{v,0})\|, \tag{7}$$

where $\|\cdot\|$ is any norm in the space of data $\mathbb{R}^{pm} \times \left(\mathrm{Sym}_{++}(p, \mathbb{R}) \right)^m$. This implies that the set of possible data is bounded with respect to any norm in the vector space $\mathbb{R}^{pm} \times (\mathrm{Sym}(p, \mathbb{R}))^m$. Note that, by the same arguments as can be found in [9], under any norm, all sensors' data converge to the same consensus value, i.e., $\|\mathcal{D}_{s,k} - \mathcal{D}_{v,k}\| \to 0$ and $\|\mathcal{D}_{s,k} - \mathcal{D}\| \to 0$ for all $s, v \in \mathcal{S}$.

B. Random Approximate Projections

The optimization part of the algorithm begins in line 10 by computing a negative subgradient of the f at the local iterate, which we denote by $-f'(\mathbf{z}_{s,k})$, taking a gradient descent step, and projecting the iterate back to the simple constraint set $\mathcal{X}_0$. The positive step size α_k is diminishing, non-summable and square-summable, i.e, $\alpha_k \to 0$, $\sum_{k \in \mathbb{N}} \alpha_k = \infty$ and $\sum_{k \in \mathbb{N}} \alpha_k^2 < \infty$. The last step of our method is not a projection to the feasible set, but instead a subgradient step in the direction that minimizes the violation of a *single* LMI constraint. Essentially, each node selects one of the LMIs randomly, measures the violation of that constraint, and takes an additional gradient descent step with step size $\beta_{s,k}$ in order to minimize this violation. We call this process *random approximate projection*.

Specifically, line 11 randomly assigns $\omega_{s,k} \in \mathcal{I}$, effectively choosing to activate $h(\cdot; \cdot, \omega_{s,k})$. The idea behind line 13 is to minimize a scalar metric that measures the amount of violation of $h(\cdot; \cdot, \omega_{s,k})$, which we denote by $h_+(\cdot; \cdot, \omega_{s,k})$, while remaining feasible. Let $\Pi_+ : \mathrm{Sym}(p, \mathbb{R}) \to \mathrm{Sym}_+(p, \mathbb{R})$ denote the projection operator onto the set of positive semidefinite matrices, given by

$$\Pi_+ X = E \mathrm{diag}\left[(\lambda_1)_+, \ldots, (\lambda_p)_+ \right] E^\top, \tag{8}$$

where $(\cdot)_+ \triangleq \max\{\cdot, 0\}$ and $\lambda_1, \ldots, \lambda_p$ are the eigenvalues and E is a matrix with columns $\mathbf{e}_1, \ldots, \mathbf{e}_p$ that are the corresponding eigenvectors of X. A measure of the "positive definiteness" of X is thus $\|\Pi_+ X\|_F \equiv \sqrt{(\lambda_1)_+^2 + \cdots + (\lambda_p)_+^2}$. This defines our measure of constraint violation,

$$h_+(\mathbf{z}; \mathcal{D}, i) \triangleq \|\Pi_+ h(\mathbf{z}; \mathcal{D}, i)\|_F. \tag{9}$$

We use special notation to denote the vector of derivatives of h_+ with respect to the decision variables $\mathbf{z}$

evaluated at $(\mathbf{z}; \mathcal{D}, i)$. In particular, define the gradient $h'_+(\mathbf{z}; \mathcal{D}, i) = (\nabla_1 h_+(\mathbf{z}; \mathcal{D}, i), \ldots, \nabla_{m+qn} h_+(\mathbf{z}; \mathcal{D}, i))^\top = \nabla_{\mathbf{z}} h_+(\mathbf{z}; \mathcal{D}, i)$. Defining the function $\psi_{ij} \colon \operatorname{Sym}(p, \mathbb{R}) \times \mathbb{R}^{m+qn} \times (\mathbb{R}^{pm} \times (\operatorname{Sym}_+(p, \mathbb{R}))^m) \to \mathbb{R}$ for notational convenience, the j-th entry of $h'_+(\mathbf{z}; \mathcal{D}, i)$ is given by

$$\nabla_j h_+(\mathbf{z}; \mathcal{D}, i) = \psi_{ij}\big(\Pi_+\big(h\left(\mathbf{z}\left(\mathcal{D}, i\right)\right)\big), \mathbf{z}, \mathcal{D}\big), \tag{10}$$

where

$$\psi_{ij}(X, \mathbf{z}, \mathcal{D}) \triangleq \begin{cases} \frac{\operatorname{Tr}(\nabla_j h(\mathbf{z}, \mathcal{D}, i) X)}{\|X\|_F} & \text{if } \|X\|_F > 0 \\ d & \text{else} \end{cases} \tag{11}$$

and $d > 0$ is some arbitrary constant. The function ψ_{ij} is not continuous at $X = \mathbf{0}$, and we set $d \neq 0$ for technical reasons discussed below. The step size

$$\beta_{s,k} = \frac{h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{\big\|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\big\|^2} \tag{12}$$

given in line 12 is a variant of the Polyak step size [27]. Note that this is a well defined step size as we let $h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) = d\mathbf{1}_{m+qn}$ (cf. Eq (11)) whenever $h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) = 0$ and it is nonzero elsewhere by construction.

IV. MAIN RESULTS

In this section, we discuss our assumptions and state the main result. The detailed proofs are deferred to Section V. The feasible set is $\mathcal{X} = \mathcal{X}_0 \cap \{\mathbf{z} \in \mathbb{R}^{m+qn} \mid h(\mathbf{z}; \mathcal{D}, i) \preceq 0, \ \forall i \in \mathcal{I}\}$. The optimal value and solution set of the problem (6) is $f^* = \min_{\mathcal{X}} f(\mathbf{z})$ and $\mathcal{X}^* = \{\mathbf{z} \in \mathcal{X} \mid f(\mathbf{z}) = f^*\}$. Note that $\mathbf{0} \in \Gamma \times \mathcal{U}$ and $h(\mathbf{0}; \mathcal{D}, i) \preceq 0$ for all possible $\mathcal{D}$ and i, thus there is always a feasible point.

At step $k \in \mathbb{N}$, $\mathcal{D}_{s,k}$ in line 9 of Algorithm 1 may not yet have converged to $\mathcal{D}$, so we denote the error in the constraint violation by

$$\nu_{s,k} = h_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}\right) - h_+\left(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k}\right). \tag{13}$$

Similar to the definition in (13), we define the error in the gradient of the constraint violation to be

$$\boldsymbol{\delta}_{s,k} = h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k}\right). \tag{14}$$

The vectors $\nu_{s,k}$ and $\boldsymbol{\delta}_{s,k}$ will enable us to quantify the joint effect of the function value error $\nu_{s,k}$ and the subgradient error $\boldsymbol{\delta}_{s,k}$ in studying convergence properties of Algorithm 1.

A. Bounds

The norms of $\nu_{s,k}$ and $\boldsymbol{\delta}_{s,k}$ are bounded, i.e., for some scalars $D, N > 0$, there holds for all $k \in \mathbb{N}$ and $s \in \mathcal{S}$ a.s.: $|\nu_{s,k}| \leq N$ and $\|\boldsymbol{\delta}_{s,k}\| \leq D$. We calculate the exact values of D and N in Corollaries A.2 and A.7, respectively.

The set $\mathcal{X}_0 = \Gamma \times \mathcal{U}$ is convex and compact. Therefore, there must exist a constant $C_{\mathbf{z}} > 0$ such that for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{X}_0$

$$\|\mathbf{z}_1 - \mathbf{z}_2\| \leq C_{\mathbf{z}}. \tag{15}$$

Also, note that the functions $f(\cdot)$ and $h(\mathbf{z}; \mathcal{D}, i)$ for $i \in \mathcal{I}$ are convex (not necessarily differentiable) over some open set that contains $\mathcal{X}_0$. A direct consequence of this and the compactness

of $\mathcal{X}_0$ is that the subdifferentials $\partial f(\mathbf{z})$ and $\partial h_+(\mathbf{z}; \mathcal{D}, i)$ are nonempty over $\mathcal{X}_0$, where note that here we use the symbol ∂ to refer to the subdifferential set and not a partial derivative. It also implies that the subgradients $f'(\mathbf{z}) \in \partial f(\mathbf{z})$ and $h'_+(\mathbf{z}; \mathcal{D}, i) \in \partial h_+(\mathbf{z}; \mathcal{D}, i)$ are uniformly bounded over the set $\mathcal{X}_0$. That is, there is a scalar L_f such that for all $f'(\mathbf{z}) \in \partial f(\mathbf{z})$ and $\mathbf{z} \in \mathcal{X}_0$,

$$\|f'(\mathbf{z})\| \leq L_f, \tag{16a}$$

and for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{X}_0$,

$$|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leq L_f \|\mathbf{z}_1 - \mathbf{z}_2\|. \tag{16b}$$

Note that for the objective function f in (6), we have $L_f = 1$. Also, there is a scalar L_h such that for all $h'_+(\mathbf{z}; \mathcal{D}, i) \in \partial h_+(\mathbf{z}; \mathcal{D}, i)$, $\mathbf{z} \in \mathcal{X}_0$, $i \in \mathcal{I}$,

$$\|h'_+(\mathbf{z}; \mathcal{D}, i)\| \leq L_h. \tag{17}$$

We calculate the exact value of L_h in Corollary A.7. From relation (14) we have for any $i \in \mathcal{I}$, $s \in \mathcal{S}$, $k \in \mathbb{N}$, and $\mathbf{z} \in \mathcal{X}_0$

$$\|h'_+(\mathbf{z}; \mathcal{D}_{s,k}, i)\| = \|h'_+(\mathbf{z}; \mathcal{D}, i) + \boldsymbol{\delta}_{s,k}\| \leq L_h + D, \tag{18a}$$

and for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{X}_0$

$$|h_+(\mathbf{z}_1; \mathcal{D}_{s,k}, i) - h_+(\mathbf{z}_2; \mathcal{D}_{s,k}, i)| \leq (L_h + D)\|\mathbf{z}_1 - \mathbf{z}_2\|. \tag{18b}$$

The compactness of $\mathcal{X}_0$, the boundedness of the data sequences $\mathcal{D}_{s,k}$, and the continuity of the functions $h_+(\mathbf{z}; \mathcal{D}, i)$ for $i \in \mathcal{I}$ imply that there exist constants $C_h > 0$ such that for any $s \in \mathcal{S}$, $k \in \mathbb{N}$, $i \in \mathcal{I}$ and $\mathbf{z} \in \mathcal{X}_0$

$$|h_+(\mathbf{z}; \mathcal{D}, i)| \leq C_h, \tag{19a}$$

$$|h_+(\mathbf{z}; \mathcal{D}_{s,k}, i)| = |h_+(\mathbf{z}; \mathcal{D}, i) + \nu_{s,k}| \leq C_h + N, \tag{19b}$$

where the second relation follows from (13). Furthermore, when $h_+(\mathbf{z}; \mathcal{D}_{s,k}, i) \neq 0$, we have $h'_+(\mathbf{z}; \mathcal{D}_{s,k}, i) \neq \mathbf{0}$. Therefore, there must exist a constant $c_h > 0$ such that

$$\|h'_+(\mathbf{z}; \mathcal{D}_{s,k}, i)\| \geq c_h \tag{20}$$

for all $\mathbf{z} \in \mathcal{X}_0$, $s \in \mathcal{S}$, $k \in \mathbb{N}$, and $i \in \mathcal{I}$. This and relation (19b) imply that

$$\beta_{s,k} = \frac{h_+(\mathbf{z}; \mathcal{D}_{s,k}, \omega_{s,k})}{\|h'_+(\mathbf{z}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2} \leq \frac{C_h + N}{c_h^2}. \tag{21}$$

Note that when $h_+(\mathbf{z}; \mathcal{D}_{s,k}, \omega_{s,k}) = 0$, the above bound holds trivially.

B. Assumptions

In the preceding sections, we have made extensive use of the function $Q \colon \mathbb{R}^q \times \mathbb{R}^p \to \operatorname{Sym}_+(p, \mathbb{R})$. For our algorithm to converge, we require the following to hold true regarding this information model.

Assumption IV.1. *We assume that the information function Q*

(a) is bounded,

(b) is twice subdifferentiable

(c) has bounded subdifferentials up to the second order, and

(d) has relatively few critical points, i.e., the sets of critical points of Q and its partial derivatives up to the second order are measure zero.

Denote the bound on the magnitude of Q by η_0, the bound on its first derivatives by η_1, and the bound on its second derivatives by η_2.

The technical restrictions on the information model Q are not overly restrictive in practice. In particular, items (a), (b), and (c) imply that one cannot obtain infinite information and that, by changing sensor configuration, the information rate (and the rate of change of the rate) cannot change infinitely quickly. Item (d) essentially allows us to distinguish between sensor configurations, i.e., the set of configurations that offer "optimal" information rates are relatively sparse with respect to the space of signal source positions and sensor configurations $\mathbb{R}^q \times \mathbb{R}^p$.

The next two assumptions are related to the random variables $\omega_{s,k}$. At each iteration k of the inner-loop, recall that each sensor s randomly generates $\omega_{s,k} \in \mathcal{I}$. We assume that they are *i.i.d.* samples from some probability distribution on $\mathcal{I}$.

Assumption IV.2. *In the k-th iteration of the inner-loop, each element i of $\mathcal{I}$ is generated with nonzero probability, i.e., for any $s \in \mathcal{S}$ and $k \in \mathbb{N}$ it holds that $\pi_i \triangleq Pr\{\omega_{s,k} = i\} > 0$.*

Assumption IV.2 is easy to satisfy because $\mathcal{I}$ is a finite set.

Assumption IV.3. *For all $s \in \mathcal{S}$ and $k \in \mathbb{N}$, there exists a constant $\kappa > 0$ such that for all $\mathbf{z} \in \mathcal{X}_0$ $\mathrm{dist}^2(\mathbf{z}, \mathcal{X}) \leq \kappa \mathsf{E}\left[h_+^2(\mathbf{z}; \mathcal{D}_{s,k}, \omega_{s,k})\right].$*

The upper bound in Assumption IV.3 is known as *global error bound* and is crucial for the convergence analysis of Algorithm 1. Sufficient conditions for this bound have been shown in [28]–[31], which includes the case when each function $h(\cdot; \cdot, i)$ is linear in $\mathbf{z}$, or when the feasible set $\mathcal{X}$ has a nonempty interior.

Assumption IV.4. *For all $k \in \mathbb{N}$, the weighted graphs $\mathcal{G}_k = (\mathcal{S}, \mathcal{E}_k, W_k)$ satisfy:*
(a) There exists a scalar $\eta \in (0, 1)$ such that $[W_k]_{sj} \geq \eta$ if $j \in \mathcal{N}_{s,k}$. Otherwise, $[W_k]_{sj} = 0$.
(b) The weight matrix W_k is doubly stochastic, i.e., $\sum_{s \in \mathcal{S}}[W_k]_{sj} = 1$ for all $j \in \mathcal{S}$ and $\sum_{j \in \mathcal{S}}[W_k]_{sj} = 1$ for all $s \in \mathcal{S}$.
(c) There exists a scalar $Q > 0$ such that the graph $(\mathcal{S}, \cup_{\ell=0,\dots,Q-1}\mathcal{E}_{t+\ell})$ is strongly connected for any $t \geq 1$.

This assumption ensures a balanced communication between sensors. It also ensures that there exists a path from one sensor to every other sensor infinitely often even if the underlying graph topology is time-varying.

C. Main result

Our main result shows the almost sure convergence of Algorithm 1. Specifically, the result states that the sensors asymptotically reach an agreement to a random point $\mathbf{z}^*$ which is in the optimal set $\mathcal{X}^*$ almost surely (a.s.), as given in the following theorem.

Theorem IV.5 (Asymptotic Convergence Under Noise). *Let Assumptions IV.1 - IV.4 hold. Let also the stepsizes be such that $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then, the iterates $\{\mathbf{z}_{s,k}\}$ generated by each agent $s \in \mathcal{S}$ via Algorithm 1 converge almost surely to the same point in the optimal set $\mathcal{X}^*$, i.e., for a random point $\mathbf{z}^* \in \mathcal{X}^*$ $\lim_{k \to \infty} \mathbf{z}_{s,k} = \mathbf{z}^*, \quad \forall s \in \mathcal{S} \quad a.s.$*

V. CONVERGENCE OF ALGORITHM 1

A. Preliminary Results

Lemma V.1 (Non-expansiveness [32]). *Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a nonempty closed convex set. The function $\Pi_{\mathcal{X}} : \mathbb{R}^d \to \mathcal{X}$ is nonexpansive, i.e., $\|\Pi_{\mathcal{X}}[\mathbf{a}] - \Pi_{\mathcal{X}}[\mathbf{b}]\| \leq \|\mathbf{a} - \mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$.*

Lemma V.2 ([33, Lemma 10-11, p. 49-50]). *Let $\mathcal{F}_k \triangleq \{v_\ell, u_\ell, a_\ell, b_\ell\}_{\ell=0}^k$ denote a collection of nonnegative real random variables for $k \in \mathbb{N} \cup \{\infty\}$ such that $\mathsf{E}[v_{k+1}|\mathcal{F}_k] \leq (1+a_k)v_k - u_k + b_k$ for all $k \in \{0\} \cup \mathbb{N}$ a.s. Assume further that $\{a_k\}$ and $\{b_k\}$ are a.s. summable. Then, we have a.s. that (i) $\{u_k\}$ is summable and (ii) there exists a nonnegative random variable v such that $\{v_k\} \to v$.*

In the next lemma, we relate the two iterates $\mathbf{p}_{s,k}$ and $\mathbf{z}_{s,k}$ in Line 6 and 13 of Algorithm 1. In particular, we show a relation of $\mathbf{p}_{s,k}$ and $\mathbf{z}_{s,k-1}$ associated with any convex function g which will be often used in the analysis. For example, $g(\mathbf{z}) = \|\mathbf{z} - \mathbf{a}\|^2$ for some $\mathbf{a} \in \mathbb{R}^n$ or $g(\mathbf{z}) = \mathrm{dist}^2(\mathbf{z}, \mathcal{X})$.

Lemma V.3 (Convexity and Double Stochasticity). *For any convex function $g : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\sum_{s \in \mathcal{S}} g(\mathbf{p}_{s,k}) \leq \sum_{s \in \mathcal{S}} g(\mathbf{z}_{s,k-1})$$

Proof. The double stochasticity of the weights plays a crucial role in this lemma. From the definition of $\mathbf{p}_{s,k}$ in Line 6 of Algorithm 1, we obtain

$$\sum_{s \in \mathcal{S}} g(\mathbf{p}_{s,k}) \leq \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{S}} [W_k]_{sj} g(\mathbf{z}_{j,k-1})$$
$$= \sum_{j \in \mathcal{S}} \left(\sum_{s \in \mathcal{S}} [W_k]_{sj}\right) g(\mathbf{z}_{j,k-1}) = \sum_{j \in \mathcal{S}} g(\mathbf{z}_{j,k-1}).$$
$\square$

Lastly, for the convergence proof of our algorithm, we use a result from [34], which ensures successful consensus in the presence of a well behaved disturbance sequence.

Lemma V.4 (Perturbed Consensus). *Let Assumption IV.4 hold. Consider the iterates generated by*

$$\boldsymbol{\theta}_{s,k} = \sum_{v \in \mathcal{S}} [W_k]_{sv} \boldsymbol{\theta}_{v,k-1} + \mathbf{e}_{s,k}, \ \forall s \in \mathcal{S}, \tag{22}$$

Suppose there exists a nonnegative nonincreasing scalar sequence $\{\alpha_k\}$ such that $\sum_{k=1}^{\infty} \alpha_k \|\mathbf{e}_{s,k}\| < \infty$ for all $s \in \mathcal{S}$. Then, for all $s, v \in \mathcal{S}$, $\sum_{k=1}^{\infty} \alpha_k \|\boldsymbol{\theta}_{s,k} - \boldsymbol{\theta}_{v,k}\| < \infty$.

In addition to the well-known results of Lemmas V.1-V.4, we need the following intermediate results before the main result in Theorem IV.5 can be proven. The proofs of all of these Lemmas are deferred to Appendices A and B. In the first

two of them, we posit that the two error sequences $\{\nu_{s,k}\}$ and $\{\boldsymbol{\delta}_{s,k}\}$ are summable.

Lemma V.5 (Summable Constraint Errors). *For almost every bounded sequence* $\{\mathbf{z}_{s,k}\}$, *the error in the constraint* $\{|\nu_{s,k}|\}_{k \in \mathbb{N}}$ *is summable a.s., i.e.,* $\sum_{k=0}^{\infty} |\nu_{s,k}| < \infty$.

Lemma V.6 (Summable Constraint Violation Errors). *For almost every bounded sequence* $\{\mathbf{z}_{s,k}\}$, *the error in the gradient of the constraint* $\{\|\boldsymbol{\delta}_{s,k}\|\}_{k \in \mathbb{N}}$ *is summable a.s., i.e.,* $\sum_{k=0}^{\infty} \|\boldsymbol{\delta}_{s,k}\| < \infty$.

In the next lemma, we relate the two iterates $\mathbf{p}_{s,k}$ and $\mathbf{z}_{s,k}$ in Line 6 and 13 of Algorithm 1.

Lemma V.7 (Basic Iterate Relation). *Consider the sequences* $\{\mathbf{z}_{s,k}\}_{k=0}^{\infty}$ *and* $\{\mathbf{p}_{s,k}\}_{k=0}^{\infty}$ *for* $s \in \mathcal{S}$ *generated by Algorithm 1. Then, for any* $\mathbf{z}, \check{\mathbf{z}} \in \mathcal{X}$, $s \in \mathcal{S}$ *and* $k \in \mathbb{N}$, *we have a.s.:*

$$\|\mathbf{z}_{s,k} - \mathbf{z}\|^2 \leq \|\mathbf{p}_{s,k} - \mathbf{z}\|^2 - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z}))$$
$$+ 2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\|\boldsymbol{\delta}_{s,k}\| + 2\frac{C_h + N}{c_h^2}|\nu_{s,k}|$$
$$+ \frac{1}{4\eta}\|\mathbf{p}_{s,k} - \check{\mathbf{z}}\|^2 - \frac{\tau - 1}{\tau(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$+ A_{\eta,\tau}\alpha_k^2 L_f^2,$$

where $A_{\eta,\tau} = 1 + 4\eta + \tau$ *and* $\eta, \tau > 0$ *are arbitrary.*

Since we use a random approximate projection, we cannot guarantee the feasibility of the sequences $\mathbf{p}_{s,k}$ and $\mathbf{z}_{s,k}$ for every $k \in \mathbb{N}$ and $s \in \mathcal{S}$. In the next lemma, we prove that $\mathbf{p}_{s,k}$ for all $s \in \mathcal{S}$ asymptotically achieve feasibility even under the effect of the disturbances $\{\nu_{s,k}\}$ and $\{\boldsymbol{\delta}_{s,k}\}$.

Lemma V.8 (Asymptotic Feasibility Under Noise). *Let Assumption IV.2 and IV.3 hold. Let* α_k *be square summable Consider the sequences* $\{\mathbf{p}_{s,k}\}$ *for* $s \in \mathcal{S}$ *generated by Algorithm 1. Then, for any* $s \in \mathcal{S}$, *the sequence* $\{\text{dist}(\mathbf{p}_{s,k}, \mathcal{X})\}_{k \in \mathbb{N}}$ *is a.s. square summable.*

In the final Lemma, we show that $\|\mathbf{z}_{s,k} - \mathbf{p}_{s,k}\|$ eventually converges to zero for all $s \in \mathcal{S}$. This result combined with Lemma V.8 implies that the sequences $\{\mathbf{z}_{s,k}\}_{k \in \mathbb{N}}$ also achieve asymptotic feasibility.

Lemma V.9 (Network Disagreement Under Noise). *Let Assumptions IV.2-IV.4 hold. Let the sequence* $\{\alpha_k\}$ *is such that* $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. *Define* $\mathbf{e}_{s,k} = \mathbf{z}_{s,k} - \mathbf{p}_{s,k}$ *for all* $s \in \mathcal{S}$ *and* $k \geq 1$. *Then, we have a.s.:*
(a) $\sum_{k=1}^{\infty} \|\mathbf{e}_{s,k}\|^2 < \infty$ *for all* $s \in \mathcal{S}$,
(b) $\sum_{k=1}^{\infty} \alpha_k \|\check{\mathbf{z}}_{s,k} - \bar{\check{\mathbf{z}}}_k\| < \infty$ *for all* $s \in \mathcal{S}$,
where $\check{\mathbf{z}}_{s,k} = \Pi_{\mathcal{X}}[\mathbf{p}_{s,k}]$ *and* $\bar{\check{\mathbf{z}}}_k = \frac{1}{n}\sum_{s \in \mathcal{S}} \check{\mathbf{z}}_{s,k}$.

Note that the sequences $\{\mathbf{z}_{s,k}\}_{k \in \mathbb{N}}$ for $s \in \mathcal{S}$ generated by Algorithm 1 can be represented as relation (22). That is, $\mathbf{z}_{s,k} = \sum_{s \in \mathcal{S}}[W_k]_{sj}\mathbf{z}_{j,k-1} + \mathbf{e}_{s,k}$ for all $s \in \mathcal{S}$ and $\mathbf{e}_{s,k} = \mathbf{z}_{s,k} - \mathbf{p}_{s,k}$. Therefore, from Lemma V.9(a), we have $\sum_{k=1}^{\infty} \alpha_k \|\mathbf{e}_{s,k}\| = \frac{1}{2}\sum_{k=1}^{\infty} \alpha_k^2 + \frac{1}{2}\sum_{k=1}^{\infty} \|\mathbf{e}_{s,k}\|^2 < \infty$. Invoking Lemma V.4, we see that there is consensus among $\mathbf{z}_{s,k}$ for $s \in \mathcal{S}$.

B. Proof of Theorem IV.5

We invoke Lemma V.7 with $\check{\mathbf{z}} = \check{\mathbf{z}}_{s,k}$ (Note that $\check{\mathbf{z}}_{s,k}$ is defined as $\Pi_{\mathcal{X}}[\mathbf{p}_{s,k}]$ in Lemma V.9), $\tau = 4$ and $\eta = \kappa(L_h +$

$D)^2$. We also let $\mathbf{z} = \mathbf{z}^*$ for an arbitrary $\mathbf{z}^* \in \mathcal{X}^*$. Therefore, for any $\mathbf{z}^* \in \mathcal{X}^*$, $s \in \mathcal{S}$ and $k \in \mathbb{N}$, we have *a.s.*:

$$\|\mathbf{z}_{s,k} - \mathbf{z}^*\|^2 \leq \|\mathbf{p}_{s,k} - \mathbf{z}^*\|^2 - 2\alpha_k(f(\check{\mathbf{z}}_{s,k}) - f(\mathbf{z}^*))$$
$$+ 2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\|\boldsymbol{\delta}_{s,k}\| + 2\frac{C_h + N}{c_h^2}|\nu_{s,k}|$$
$$+ \frac{1}{4\kappa(L_h + D)^2}\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A\alpha_k^2 L_f^2$$
$$- \frac{3}{4(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}), \tag{23}$$

where $A = 5 + 4\kappa(L_h + D)^2$. From Assumption IV.3, we know that $\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) \leq \kappa \mathsf{E}\left[h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \mid \mathcal{F}_{k-1}\right]$.

Denote by $\mathcal{F}_k$ the sigma-field induced by the history of the algorithm up to time k, i.e., $\mathcal{F}_k = \{\mathbf{z}_{s,0}, (\omega_{s,\ell}, 1 \leq \ell \leq k), s \in \mathcal{S}\}$ for $k \in \mathbb{N}$, and $\mathcal{F}_0 = \{\mathbf{z}_{s,0}, s \in \mathcal{S}\}$. Taking the expectation conditioned on $\mathcal{F}_{k-1}$ in relation (23), summing this over $s \in \mathcal{S}$, and using the above relation, we obtain

$$\sum_{s \in \mathcal{S}} \mathsf{E}\left[\|\mathbf{z}_{s,k} - \mathbf{z}^*\|^2 \mid \mathcal{F}_{k-1}\right] \tag{24}$$
$$\leq \sum_{s \in \mathcal{S}} \|\mathbf{z}_{s,k-1} - \mathbf{z}^*\|^2 - 2\alpha_k \sum_{s \in \mathcal{S}}(f(\check{\mathbf{z}}_{s,k}) - f(\mathbf{z}^*))$$
$$+ 2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\sum_{s \in \mathcal{S}} \mathsf{E}[\|\boldsymbol{\delta}_{s,k}\| \mid \mathcal{F}_{k-1}]$$
$$+ 2\frac{C_h + N}{c_h^2}\sum_{s \in \mathcal{S}} \mathsf{E}[|\nu_{s,k}| \mid \mathcal{F}_{k-1}]$$
$$- \frac{1}{2\kappa_2(L_h + D)^2}\sum_{s \in \mathcal{S}} \text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A\alpha_k^2 n L_f^2,$$

where we used Lemma V.3 for the first term on the right-hand side. Recall that $f(\mathbf{z}) = \sum_{s \in \mathcal{S}} f(\mathbf{z})$ and $\bar{\check{\mathbf{z}}}_k \triangleq \frac{1}{n}\sum_{\ell \in \mathcal{S}} \check{\mathbf{z}}_{\ell,k}$. Using $\bar{\check{\mathbf{z}}}_k$ and f, we can rewrite the term $\sum_{s \in \mathcal{S}}(f(\check{\mathbf{z}}_{s,k}) - f(\mathbf{z}^*))$ as follows:

$$\sum_{s \in \mathcal{S}}(f(\check{\mathbf{z}}_{s,k}) - f(\mathbf{z}^*))$$
$$= \sum_{s \in \mathcal{S}}(f(\check{\mathbf{z}}_{s,k}) - f(\bar{\check{\mathbf{z}}}_k)) + (f(\bar{\check{\mathbf{z}}}_k) - f^*) \tag{25a}$$
$$\leq \sum_{s \in \mathcal{S}}\langle f'(\bar{\check{\mathbf{z}}}_k), \bar{\check{\mathbf{z}}}_k - \check{\mathbf{z}}_{s,k}\rangle + (f(\bar{\check{\mathbf{z}}}_k) - f^*) \tag{25b}$$
$$\leq \sum_{s \in \mathcal{S}}\|f'(\bar{\check{\mathbf{z}}}_k)\| \|\check{\mathbf{z}}_{s,k} - \bar{\check{\mathbf{z}}}_k\| + (f(\bar{\check{\mathbf{z}}}_k) - f^*) \tag{25c}$$
$$\leq L_f \sum_{s \in \mathcal{S}}\|\check{\mathbf{z}}_{s,k} - \bar{\check{\mathbf{z}}}_k\| + (f(\bar{\check{\mathbf{z}}}_k) - f^*), \tag{25d}$$

where (25a) follows from adding and subtracting $f(\bar{\check{\mathbf{z}}}_k)$; (25b) follows from the convexity of the function f; (25c) follows from the Schwarz inequality; and (25d) follows from relation (16a) and $\bar{\check{\mathbf{z}}}_k \in \mathcal{X} \subseteq \mathcal{X}_0$. Combining (25d) with (24), we obtain

$$\sum_{s \in \mathcal{S}} \mathsf{E}\left[\|\mathbf{z}_{s,k} - \mathbf{z}^*\|^2 \mid \mathcal{F}_{k-1}\right]$$
$$\leq \sum_{s \in \mathcal{S}}\|\mathbf{z}_{s,k-1} - \mathbf{z}^*\|^2 - 2\alpha_k(f(\bar{\check{\mathbf{z}}}_k) - f^*)$$
$$+ 2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\sum_{s \in \mathcal{S}} \mathsf{E}[\|\boldsymbol{\delta}_{s,k}\| \mid \mathcal{F}_{k-1}]$$

$$+ 2\frac{C_h + N}{\check{c}_h^2} \sum_{s \in \mathcal{S}} \mathsf{E}[|\nu_{s,k}| \mid \mathcal{F}_{k-1}]$$
$$+ L_f \sum_{s \in \mathcal{S}} \alpha_k \|\check{\mathbf{z}}_{s,k} - \bar{\bar{\mathbf{z}}}_k\| + A\alpha_k^2 n L_f^2,$$

where we omitted the negative term.

Since $\bar{\bar{\mathbf{z}}}_k \in \mathcal{X}$, we have $f(\bar{\bar{\mathbf{z}}}_k) - f^* \geq 0$. Thus, under the assumption $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ and Lemma V.9(b), the above relation satisfies all the conditions of Lemma V.2. Using this, we have the following results.

Result 1: For some $\mathbf{z}^* \in \mathcal{X}^*$ and any $s \in \mathcal{S}$, the sequence $\{\|\mathbf{z}_{s,k} - \mathbf{z}^*\|\}$ is convergent *a.s.*

Result 2: $\sum_{k=1}^{\infty} \alpha_k(f(\bar{\bar{\mathbf{z}}}_k) - f^*) < \infty$ *a.s.*

As a direct consequence of *Result 1*, we know that the sequences $\{\|\mathbf{p}_{s,k} - \mathbf{z}^*\|\}$ and $\{\|\check{\mathbf{z}}_{s,k} - \mathbf{z}^*\|\}$ are also convergent *a.s.* (This is straightforward from Line 6 of Algorithm 1 and Lemma V.8 with the knowledge that $\mathrm{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) = \|\mathbf{p}_{s,k} - \check{\mathbf{z}}_{s,k}\|^2$.) Since $\|\bar{\bar{\mathbf{z}}}_k - \mathbf{z}^*\| \leq \frac{1}{N} \sum_{i \in \mathcal{S}} \|\check{\mathbf{z}}_{s,k} - \mathbf{z}^*\|$, we further know that the sequence $\{\|\bar{\bar{\mathbf{z}}}_k - \mathbf{z}^*\|\}$ is also convergent *a.s.*

As a direct consequence of *Result 2*, since α_k is not summable, $\liminf_{k \to \infty}(f(\bar{\bar{\mathbf{z}}}_k) - f^*) = 0$ *a.s.* From this relation and the continuity of f, it follows that the sequence $\{\bar{\bar{\mathbf{z}}}_k\}$ must have one accumulation point in the set $\mathcal{X}^*$ *a.s.* This and the fact that $\{\|\bar{\bar{\mathbf{z}}}_k - \mathbf{z}^*\|\}$ is convergent *a.s.* for every $\mathbf{z}^* \in \mathcal{X}^*$ imply that

$$\lim_{k \to \infty} \bar{\bar{\mathbf{z}}}_k = \mathbf{z}^* \quad a.s. \tag{26}$$

Also, by Lemma V.9(b), we have

$$\liminf_{k \to \infty} \|\check{\mathbf{z}}_{s,k} - \bar{\bar{\mathbf{z}}}_k\| = 0 \quad \text{for all } s \in \mathcal{S} \quad a.s. \tag{27}$$

The fact that $\{\|\check{\mathbf{z}}_{s,k} - \mathbf{z}^*\|\}$ is convergent *a.s.* for all $s \in \mathcal{S}$, together with (26) and (27) implies that

$$\lim_{k \to \infty} \check{\mathbf{z}}_{s,k} = \mathbf{z}^* \quad \text{for all } s \in \mathcal{S} \quad a.s. \tag{28}$$

Since $\|\mathbf{z}_{s,k} - \check{\mathbf{z}}_{s,k}\| \leq \|\mathbf{z}_{s,k} - \mathbf{p}_{s,k}\| + \|\mathbf{p}_{s,k} - \check{\mathbf{z}}_{s,k}\|$, by invoking Lemma V.8 and V.9(a), we have $\lim_{k \to \infty} \|\mathbf{z}_{s,k} - \check{\mathbf{z}}_{s,k}\| = 0$ for all $s \in \mathcal{S}$ *a.s.* This and relation (28) give us the desired result, which is $\lim_{k \to \infty} \mathbf{z}_{s,k} = \mathbf{z}^* $ for all $s \in \mathcal{S}$ *a.s.*

VI. NUMERICAL EXPERIMENTS

In this section, we illustrate Algorithm 1 on a network of n mobile sensors that collaboratively localize a collection of m landmarks. We use a distributed connectivity maintenance controller [35] to ensure that the sensor network can share information and collaborate for all time. The controller operates by signaling for a strong attractive force whenever a link is going to be deleted that can break connectivity. The controller also deletes or replaces weak links as long as connectivity is preserved.

A. Sensor Model

We assume that the landmarks live in $\mathbb{R}^2$ and denote the configuration of landmark i by $\mathbf{x}_i$. We also assume that the configuration space of the robot is $\mathbb{R}^2$.

Generic models for the measurement covariance matrix for sparse landmark localization have been proposed [1], [2], [7].
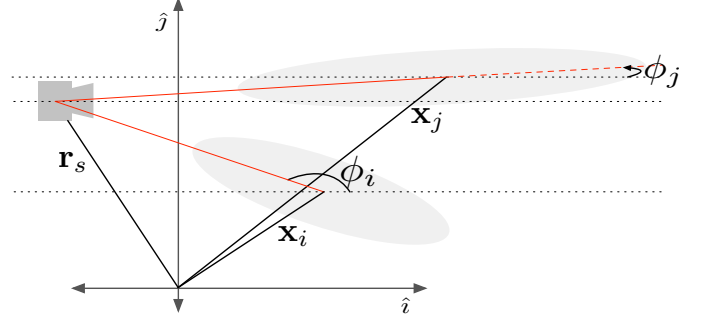

Figure 1. A sensor at $\mathbf{r}_s$ observes targets it believes to be at $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j$. Possible confidence regions are shown in grey. Dotted lines are parallel to the global $\hat{\imath}$ axis. Red lines guide the eye from the sensor to targets

These models have two important characteristics that apply to a wide variety of localization sensors: (*i*) Measurement quality is inversely proportional to viewing distance, and (*ii*) The direction of maximum uncertainty is the viewing direction, i.e., the sensors are more proficient at sensing bearing than range. The general idea is to use the vector $\hat{\mathbf{x}}_i(t) - \mathbf{r}_s(t)$ to construct a diagonal matrix in a coordinate frame relative to the sensor, then rotate the matrix to a global coordinate frame so that it can be compared to other observations.

Following the accepted generic sensor models, we express $Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)$ by its eigenvalue decomposition $Q = R_{\phi_i} \Lambda_i R_{\phi_i}^\top$, where the argument $(\mathbf{r}_s, \hat{\mathbf{x}}_i)$ of Q, Λ_i, and R_{ϕ_i} is implicit. The angle $\phi_i \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is defined so that the first column of R_{ϕ_i}, which by convention is $[\cos \phi_i \ \sin \phi_i]^\top$, is parallel to the subspace spanned by the vector $\hat{\mathbf{x}}_i - \mathbf{r}_s$. This angle is given by $\phi_i = \tan^{-1}(([\hat{\mathbf{x}}_i]_2 - [\mathbf{r}_s]_2)/([\hat{\mathbf{x}}_i]_1 - [\mathbf{r}_s]_1))$, where the subscripts outside the brackets refer to the first and second coordinate of the vector representing the location of the landmark and sensor located at $\hat{\mathbf{x}}_i$ and $\mathbf{r}_s$. The eigenvalue matrix is given by $\Lambda_i = \mathrm{diag}(\lambda_{1,i}, \lambda_{2,i})$, where $\lambda_{1,i} = c_0(/1 + c_1\|\hat{\mathbf{x}}_i - \mathbf{r}_s\|_2)$ and $\lambda_{2,i} = \rho\lambda_{1,i}$. In the body frame of sensor s, the eigenvalues $\lambda_{1,i}$ and $\lambda_{2,i}$ control the shape of the confidence ellipses for individual measurements of the i-th target. The parameter $c_0 > 0$ represents the overall sensor quality and scales the whole region equally, $c_1 > 0$ controls the sensitivity to depth, and $\rho > 1$ controls eccentricity of the confidence regions associated with the measurements. Figure 1 illustrates the measurement model for one sensor and two targets.

Since $\tan^{-1}$ is not continuous at zero, we need to impose the following restriction so that Assumption IV.1 holds. Essentially, we assume that, for the selected $\delta > 0$, which recall has an affect on the robot's ability to translate in $\mathbb{R}^2$, we choose a τ_i such that $\tau_i < c_0/(1 + c_1\delta)$. Thus, if the robot is within δ of a target i, the constraint $h(\mathbf{z}, \mathcal{D}, i)$ is trivially satisfied, and therefore we will never evaluate Q or any of its derivatives in this region.

B. Results

Figure 2 shows an example of the resulting active sensing trajectories for a network of four sensors that cooperatively localize a uniformly random distribution of 100 landmarks in a square $100\,\mathrm{m}^2$ workspace to a desired accuracy of $1\,\mathrm{m}$. This
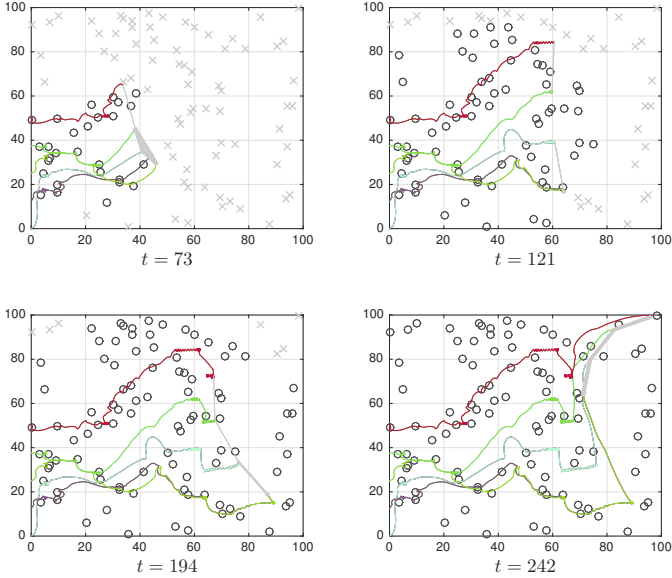
Figure 2. Four sensors localize 100 landmarks. Gray ×'s indicate landmarks with uncertainty above the threshold, and black ∘'s indicate localized landmarks. Gray lines connecting the agents indicate the communication links.

corresponds to an eigenvalue tolerance of $\tau_1 = \cdots = \tau_m = 9\,\mathrm{m}^{-2}$. The prior distribution on the landmarks was uninformative, i.e., the initial mean estimate and information matrix are given by the initial observations $\{\mathbf{y}_{i,s}(0) \mid i \in \mathcal{I}, s \in \mathcal{S}\}$ and the corresponding information matrices were set to $\{Q(\mathbf{r}_s(0), \mathbf{y}_{i,s}(0))\}_{i\in\mathcal{I}, s\in\mathcal{S}}$. The parameters of the sensing model Q were set to $c_0 = \frac{1}{2}, c_1 = 10,$ and $\rho = 30$. The feasible control set $\mathcal{U}$ was set to $\mathcal{U} = \{\mathbf{u}_s \in \mathbb{R}^2 \mid \|\mathbf{u}_s\| \leq 1\,\mathrm{m}\}$. In the figure, note that between $t = 194$ and $t = 242$, the agent in the bottom right moves north in order to maintain connectivity while the other agents move to finish the localization task. It can be seen from the four snapshots of the trajectories in Fig. 2, that the robots effectively "divide and conquer" the task based on proximity to unlocalized targets.

Fig. 3 shows the evolution of the minimum eigenvalues of the information matrices $S_i(t)$ from the simulation shown in Fig. 2. Note that, even after the minimum eigenvalue associated with a landmark has surpassed τ, more information is still collected, but it does not affect the control signal. Note also that the maximal uncertainty in the bottom panel does not go below the desired threshold, even though the minimal eigenvalue does go above the threshold in the top panel. This is expected because we use the 95% confidence interval to generate the threshold τ, thus we expect 5% of landmarks to have errors larger than the desired threshold.

In this case study, we selected the termination criterion for Algorithm 1 to be at most 30 iterations. When testing the algorithm with $\alpha_k = 1/k$, we observed that the relative change in the iterates $\mathbf{z}_{s,k}$ and the data $\mathcal{D}_{s,k}$ would often reach levels at or below 10% inside of thirty iterations, so we allowed the message passing to terminate early if this criterion was met.

Figure 4 measures the amount of constraint violation of the original nonlinear constraints for the solution obtained from the linearized version that we observed in a simulation with twenty landmarks. For the $\mathcal{U}$ and the parameters for the function Q used in these simulations, the constraint violation was very small, and driven to zero typically for $t > 10$.
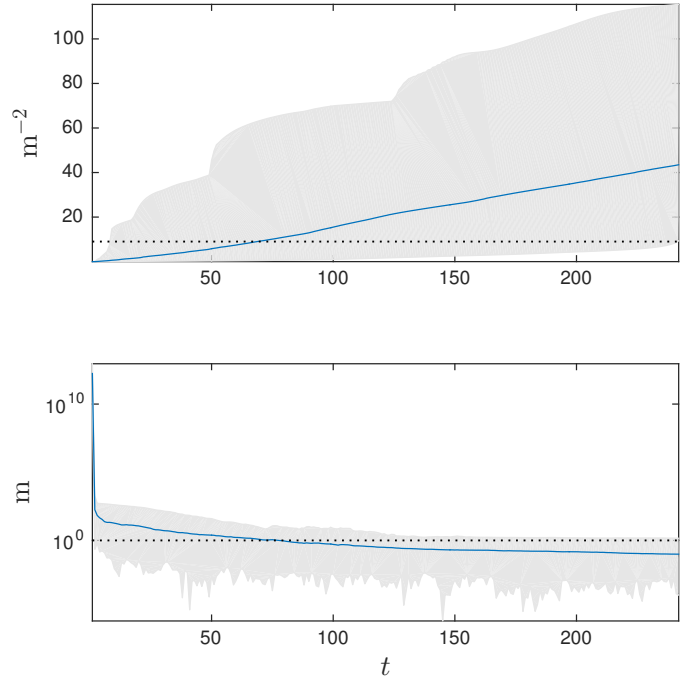

Figure 3. Top panel: Evolution of the minimum eigenvalues. The center curve is the mean among all 100 targets, with the grey block extending to the lowest and largest eigenvalues at each time instant. The dotted line is at the tolerance $\tau = 9\,\mathrm{m}^{-2}$. Bottom panel: Evolution of localization error. The center curve is the mean among all 100 targets, with the grey block extending to the minimum and maximum errors at each time instant. The dotted line is at the desired accuracy of $1\,\mathrm{m}$. Note that the scale on vertical axis is large because the initial uncertainty is infinite.
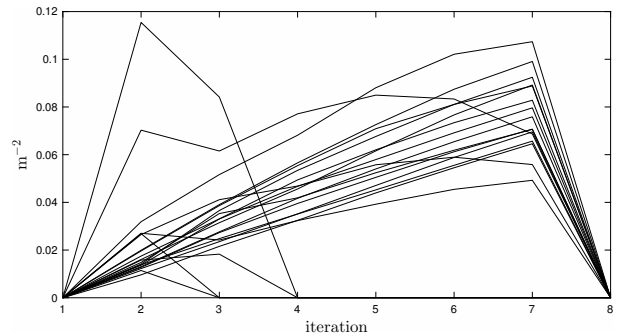

Figure 4. Plotting the amount of constraint violation of the original nonlinear constraint that will occur because we solve a linearized version. Twenty lines are drawn, one for each landmark in a simulation with $m = 20$. The horizontal axis represents the outer loop iterations t. Lines are drawn to guide the eye

We tested larger feasible sets $\mathcal{U}$ and (predictably) observed a general increase in the constraint violation, although we still observed reasonable robot behaviors.

VII. CONCLUSION

In this paper, we addressed the problem of controlling a network of mobile sensors so that a set of hidden states are estimated up to a user-specified accuracy. We proposed a new algorithm that combines the ICF with control. The first major contribution is the capability of controlling worst-case error in the hidden states. The second major contribution is the novel numerical technique, which is robust toward noisy gradients, computationally light, can handle large systems of coupled LMIs. We showed that our method converges to the optimal solution and presented simulations for sparse

landmark localization, where the team was able to achieve the desired estimation tolerances while exhibiting interesting emergent behaviors, notably "dividing and conquering" the distributed estimation task, observing landmarks from diverse viewing angles.

APPENDIX A
PROOF OF LEMMAS V.5 AND V.6

To prove Lemmas V.5 and V.6, consider the following.

Proposition A.1 (Summable Disagreement in Constraints). *Let Assumptions IV.1 - IV.4 hold. For any bounded sequence* $\{\mathbf{z}_{s,k}\}$ *and any* $i \in \mathcal{I}$,

$$\sum_{k=0}^{\infty} \|h\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}\right) - h\left(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k}\right)\|_F < \infty.$$

Proof. In this proof, we will essentially be comparing the function h for the unconverged local data $\mathcal{D}_{s,k}$ and the consensus data $\mathcal{D}$. Note that at some iteration $k \in \mathbb{N}$, the ICF is actually computing the information summary $(\boldsymbol{\xi}_{1,s,k}, \ldots, \boldsymbol{\xi}_{m,s,k}, \Xi_{1,s,k}, \ldots, \Xi_{m,s,k})$, which is used to compute $\mathcal{D}_{s,k}$ in line 9 of Algorithm 1. In particular, we will need to refer to the unconverged mean and information matrix, which is

$$\mathcal{D}_{s,k} \triangleq \Big(\big((n\Xi_{1,s,k})^{-1}\boldsymbol{\xi}_{1,s,k}, \ldots, (n\Xi_{m,s,k})^{-1}\boldsymbol{\xi}_{m,s,k} \big),$$
$$(n\Xi_{1,s,k}, \ldots, n\Xi_{m,s,k}) \Big).$$

In the proof, we will drop the reference to the sensor s, as we only consider one sensor at a time. In other words, the unconverged data will be written as

$$\mathcal{D}_k = \Bigg(\Big(\underbrace{(n\Xi_{1,k})^{-1}\boldsymbol{\xi}_{1,k}}_{\hat{\mathbf{x}}_{1,k}}, \ldots, \underbrace{(n\Xi_{m,k})^{-1}\boldsymbol{\xi}_{m,k}}_{\hat{\mathbf{x}}_{m,k}} \Big), \Big(\underbrace{n\Xi_{1,k}}_{S_{1,k}}, \ldots, \underbrace{n\Xi_{1,m}}_{S_{m,k}} \Big) \Bigg)$$

for $k \in \mathbb{N}$, i.e., this is implicitly the data local to sensor s in the proof. For the consensus data $\mathcal{D}_\infty = ((\hat{\mathbf{x}}_{1,\infty}, \ldots, \hat{\mathbf{x}}_{m,\infty}), (S_{1,\infty}, \ldots, S_{m,\infty}))$, we also drop the k subscript, i.e., we write the consensus data as $\mathcal{D} = (\hat{\mathbf{x}}, S) = ((\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_m), (S_1, \ldots, S_m))$.

Let us fix an arbitrary $i \in \mathcal{I}$, and prove the proposition for the sequence $\omega_k = \{i, i, i, \ldots\}$. If the proof holds for this sequence, it thus holds for the random sequences $\omega_k \subset \mathcal{I}$ because

$$\sum_{k=0}^{\infty} \|h\left(\mathbf{z}_k; \mathcal{D}_k, \omega_k\right) - h\left(\mathbf{z}_k; \mathcal{D}, \omega_k\right)\|_F$$
$$\leq \sum_{i \in \mathcal{I}} \sum_{k=0}^{\infty} \|h\left(\mathbf{z}_k; \mathcal{D}_k, i\right) - h\left(\mathbf{z}_k; \mathcal{D}, i\right)\|_F,$$

and if each sequence $\|h\left(\mathbf{z}_k; \mathcal{D}_k, i\right) - h\left(\mathbf{z}_k; \mathcal{D}, i\right)\|_F$ is summable, so is their sum.

Recall the definition of h from (5). We have that

$$h(\mathbf{z}_k; \mathcal{D}_k, i) - h\left(\mathbf{z}_k; \mathcal{D}, i\right) = S_{i,k} - S_i \tag{29a}$$
$$+ \sum_{v \in \mathcal{S}} Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_{i,k}\right) - Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_i\right) \tag{29b}$$
$$+ \sum_{j=1}^{q} \left(\nabla_j Q(\mathbf{r}_v, \hat{\mathbf{x}}_{i,k}) - \nabla_j Q(\mathbf{r}_v, \hat{\mathbf{x}}_i)\right) [\mathbf{u}_v]_j, \tag{29c}$$

where S_i and $\hat{\mathbf{x}}_i$ are the eventual consensus variables that compose the consensus data $\mathcal{D}$, i.e., $S_i = S_{i,\infty}$ and $\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{i,\infty}$. The remainder of the proof will show that each of the constituent terms, specifically, (29a), (29b), and (29c), are summable, and the proof will follow from that immediately.

(29a) For this term, recall from line 9 of Algorithm 1 that $S_{i,k} \leftarrow n\Xi_{i,k}$, where n is the number of agents in the network. It is well known that for the consensus sequence $\{\Xi_{i,k}\}_{k \in \mathbb{N}}$ the errors $\{\Xi_{i,k} - \Xi_{i,\infty}\}_{k \in \mathbb{N}}$ are summable; cf., e.g., [36]. Noting that $S_i = n\Xi_{i,\infty}$ and $S_{i,k} = n\Xi_{i,k}$, we have the result.

(29b) By Assumption IV.1, $\|\nabla_{q+j}Q(\mathbf{r}_v, \mathbf{x}_i)\|_F \leq \eta_1$ for all $\mathbf{r}_v$, $\mathbf{x}_i$ and j. Then, for each $v \in \mathcal{S}$ and $k \in \mathbb{N}$, using the taylor expansion of $Q(\mathbf{r}_v, \cdot)$ about the point $\hat{\mathbf{x}}_i$, we have that

$$\|Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_{i,k}\right) - Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_i\right)\|_F =$$
$$= \left\| \sum_{j=1}^{p} \nabla_{q+j}Q(\mathbf{r}_v, \hat{\mathbf{x}}_i)[\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i]_j + o\left(\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|\right) \right\|_F$$
$$\leq \eta_1 \|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|_1 + o\left(\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|\right),$$

which is summable because $\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}\|$ is the error sequence of a consensus filter, which is summable.

(29c) First, since $\mathbf{u}_v < \delta$ for all $v \in \mathcal{S}$, for any $j = 1, \ldots, p$, we have that

$$\|(\nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_{i,k}) - \nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)) [\mathbf{u}_s]_j\|_F \tag{30}$$
$$\leq \delta \|\nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_{i,k}) - \nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)\|_F$$

Also due to Assumption IV.1, η_2 is the upper bound for the norm of the second derivative of Q. In particular, let $\nabla_{q+\ell}\nabla_j Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_i\right)$ denote the mixed second partial derivative of Q with respect to the ℓ-th component of the $\mathbf{x}_i$ coordinates and the j-th component of the $\mathbf{r}_v$ coordinates of the function Q, evaluated at the point $(\mathbf{r}_v, \hat{\mathbf{x}}_i)$. This notation is useful for the following inequality, which uses the first order Taylor expansion with respect to the $\mathbf{x}_i$ coordinates of $\nabla_j Q$. In particular, the norm on the right hand side of (30) is bounded above by

$$\left\| \sum_{\ell=1}^{p} \nabla_{q+\ell}\nabla_j Q\left(\mathbf{r}_v, \hat{\mathbf{x}}_i\right) [\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i]_\ell + o\left(\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|\right) \right\|_F$$
$$\leq \left(\eta_2 \|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|_1 + o\left(\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|\right) \right).$$

The sequence above is summable again because $\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|$ is the error sequence of a consensus filter, which is summable.

Since (29a), (29b), and (29c) are each summable, their sum is also a summable sequence, and the proof follows. $\square$

Proof of Lemma V.5. Suppose $\omega_{s,k} = \{i, i, i, \ldots\}$. If the proof holds for this arbitrary i, then it holds for arbitrary sequences $\omega_{s,k} \subset \mathcal{I}$ for the same reasons as in Proposition A.1. Let $X_k = h\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right)$ and $Y_k = h\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)$. From the definition of $\nu_{s,k}$,

$$|\nu_{s,k}| = |h_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) - h_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)| \tag{31a}$$
$$= |\|\Pi_+ X_k\|_F - \|\Pi_+ Y_k\|_F| \tag{31b}$$
$$\leq \|\Pi_+ X_k - \Pi_+ Y_k\|_F \tag{31c}$$
$$\leq \|X_k - Y_k\|_F, \tag{31d}$$

which is summable by Proposition A.1. In the above manipulations, (31b) is less than or equal to (31c) by the reverse triangle inequality and (31c) is less than or equal to (31d)

because $\mathrm{Sym}_+(p,\mathbb{R})$ is a convex set. Thus, $|\nu_{s,k}|$ is element-wise less than or equal to a summable sequence, and the proof follows. $\qquad\square$

Corollary A.2 (Bound for $|\nu_{s,k}|$). *The constant N, which is the bound for $|\nu_{s,k}|$, is given by $N = \frac{1}{n}\max_{v\in\mathcal{S}}\|(\Xi_{i,0})_v\|_F + m(\eta_0 + 2\eta_1 q\delta)$.*

Proof. From the proof of Lemma V.5, each term $|\nu_{s,k}|$ is bounded above by $\|X_k - Y_k\|_F$, which, in view of (29), is the sum of three terms. The corollary follows by bounding each term, then taking the sum of the bounds. We have that

$$|\nu_{s,k}| = |h_+(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k})|$$

$$\leq \|X_k - Y_k\|_F$$

$$\leq \|S_{i,k} - S_i\|_F \tag{32a}$$

$$+ \sum_{v\in\mathcal{S}} Q(\mathbf{r}_v, \hat{\mathbf{x}}_{i,k}) - Q(\mathbf{r}_v, \hat{\mathbf{x}}_i) \tag{32b}$$

$$+ \sum_{j=1}^{q} (\nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_{i,k}) - \nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i))[\mathbf{u}_s]_j. \tag{32c}$$

The first term (32a) is the error sequence of a consensus filter, which is bounded by the largest initial condition in the network $\frac{1}{n}\max_{v\in\mathcal{S}}\|(\Xi_{i,0})_v\|_F$. For the second term (32b), recall Assumption IV.1 bounded $\|Q\|_F$ by η_0. Given that Q is positive semidefinite, (32b) is bounded by $m\eta_0$. Similarly, the third term is bounded by $2\eta_1 q\delta$, and, by taking the sum, the proof follows. $\qquad\square$

The proof of Lemma V.6 requires the help of several intermediate results. First, we present two basic results in classical analysis. Both are direct results of the Implicit Function Theorem; see, e.g., [37].

Corollary A.3. *Let $p\colon \mathbb{R}^n \to \mathbb{R}$ be a nonzero polynomial function. Then, the set of roots of p has measure zero in $\mathbb{R}^n$.*

Corollary A.4. *Let $f\colon \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a smooth function, f' surjective, and $A \subset \mathbb{R}^m$ a measure zero set. Then, the preimage of A under f is measure zero in $\mathbb{R}^{n+m}$.*

Proposition A.5 (Set of Reachable Information Matrices that Are Singular or Have Nonsimple Eigenvalues is Measure Zero). *Let*

$$\mathcal{Z}_1 = \{X \in \mathrm{Sym}(p,\mathbb{R}) \mid \lambda_i(X) = \lambda_j(X), i\neq j\}, \tag{33a}$$

$$\mathcal{Z}_2 = \{X \in \mathrm{Sym}(p,\mathbb{R}) \mid \det(X) = 0\}. \tag{33b}$$

Now, recall the function h defined in (5) and fix arbitrary $i\in\mathcal{I}$. The set

$$\mathcal{Z} = \{(\mathbf{z},\mathcal{D}) \mid h(\mathbf{z};\mathcal{D},i) \in \mathcal{Z}_1 \cup \mathcal{Z}_2\} \tag{34}$$

is measure zero.

Proof. Both the determinant and the discriminant of the characteristic polynomial are nonvanishing polynomials on $\mathrm{Sym}(p,\mathbb{R})$. Thus, $\mathcal{Z}_1$ and $\mathcal{Z}_2$ represent the zero locii of their respective polynomials, and they are measure zero sets by Corollary A.3.

To complete the proof, we need only to check surjectivity of h' and apply Theorem A.4. Recall the first definition of h with respect to the constant $\mathbf{r}$ and the variables $\gamma, S, \hat{\mathbf{x}}$, and $\mathbf{u}$, which were later shortened to the inputs $\mathbf{z}$ and the data $\mathcal{D}$.

Omitting unnecessary subscripts from the original definition in (5), slightly abusing notation, we have

$$h(\gamma, S, \hat{\mathbf{x}}_i, \mathbf{u}) = \gamma I - S - \sum_{s\in\mathcal{V}} Q(\mathbf{r}_s, \hat{\mathbf{x}}_i) - \sum_{j=1}^{q} \nabla_j Q(\mathbf{r}_s, \hat{\mathbf{x}}_i)[\mathbf{u}_s]_j.$$

Note that, in this notation, $h : \mathbb{R} \times \mathrm{Sym}_{++}(p,\mathbb{R}) \times \mathbb{R}^{p+qn} \to \mathrm{Sym}(p,\mathbb{R})$. Since $\mathrm{Sym}_{++}(p,\mathbb{R})$ is homeomorphic to $\mathbb{R}^{p(p+1)/2}$, without loss of generality, we may write h as a map from $\mathbb{R}^{a+b}$ to $\mathbb{R}^b$ with $a = 1 + p(p+1)/2 + p + qn$ and $b = p(p+1)/2$. If we write the derivative $h' : \mathbb{R}^{a+b} \to \mathbb{R}^b$ as a $b \times a$ matrix $\nabla h(\gamma, S, \hat{\mathbf{x}}_i, \mathbf{u})$, the partial derivatives corresponding to the vectorized version of S, i.e., $\nabla_S h(\gamma, S, \hat{\mathbf{x}}_i, \mathbf{u})$ will be I_b, so that $\nabla h(\gamma, S, \hat{\mathbf{x}}_i, \mathbf{u})$ will be easily be rank b, which is the required condition for surjectivity of h'. $\qquad\square$

Proposition A.6 (Bounded Entries of ∇h_+). *For any input $\mathbf{z}$ and data $\mathcal{D}$, the function $\psi_{ij}(\cdot, \mathbf{z}, \mathcal{D})$, defined in (11), representing the j-th entry of the vector $\nabla h_+(\mathbf{z}; \mathcal{D}, i)$, is bounded for all $i \in \mathcal{I}$ and $j \in \{1,\dots,m+qn\}$.*

Proof. Recall the definition of $\psi_{ij}(\cdot, \mathbf{z}, \mathcal{D})\colon \mathrm{Sym}(p,\mathbb{R}) \to \mathbb{R}$, repeated here for convenience,

$$\psi_{ij}(X, \mathbf{z}, \mathcal{D}) \triangleq \begin{cases} \frac{\mathrm{Tr}(\nabla_j h(\mathbf{z},\mathcal{D},i)X)}{\|X\|_F} & \text{if } \|X\|_F > 0 \\ d & \text{else} \end{cases}$$

Note that $\nabla_j h(\mathbf{z}, \mathcal{D}, i)$ is a constant matrix for any j because h is a linear function of $\mathbf{z}$. To simplify notation, call this matrix A_{ij}. Let $\boldsymbol{\lambda} = [\lambda_1 \ \cdots \ \lambda_p]^\top$ denote the vector of eigenvalues of X and $\mathbf{v}_j$ denote the eigenvector corresponding to λ_j. We can write $X = \sum_{\alpha=1}^{p} \lambda_\alpha \mathbf{v}_\alpha \mathbf{v}_\alpha^\top$. If $X \neq \mathbf{0}$, then $\|\boldsymbol{\lambda}\| \neq 0$ and

$$\psi_{ij} = \frac{\mathrm{Tr} A_{ij} X}{\sqrt{\mathrm{Tr} X^2}} \tag{35a}$$

$$= \frac{\mathrm{Tr} A_{ij}\left(\sum_{\alpha=1}^{p} \lambda_\alpha \mathbf{v}_\alpha \mathbf{v}_\alpha^\top\right)}{\sqrt{\sum_{\alpha=1}^{p} \lambda_\alpha^2}} = \frac{\sum_{\alpha=1}^{p} \lambda_\alpha \mathbf{v}_\alpha^\top A_{ij} \mathbf{v}_\alpha}{\|\boldsymbol{\lambda}\|}. \tag{35b}$$

Let $B_{ij} \triangleq \mathrm{diag}(v_1^\top A_{ij} v_1, \dots, v_p^\top A_{ij} v_p)$, and note that $\psi_{ij} = \mathbf{1}^\top B_{ij}\boldsymbol{\lambda}/\|\boldsymbol{\lambda}\|$. Note that the limit of $\psi_{ij}(X, \mathbf{z}, \mathcal{D})$ therefore exists for sequences $\{X_k\} \to \mathbf{0}$. With respect to bounding the magnitude of ψ_{ij}, we have that

$$|\psi_{ij}| = \frac{|\mathbf{1}^\top B_{ij}\boldsymbol{\lambda}|}{\|\boldsymbol{\lambda}\|} \leq \|B_{ij}\mathbf{1}\| = \sqrt{\sum_{\alpha=1}^{p} (\mathbf{v}_\alpha^\top A_{ij}\mathbf{v}_\alpha)^2}.$$

Therefore, ψ_{ij} is bounded. Additionally, $|\psi_{ij}|$ is maximized if $\{\mathbf{v}_\alpha\}$ represent a set of orthogonal eigenvectors of A_{ij}, that is, if X has the same eigenvectors as A_{ij}. In this case, the bound is tight and equal to $\|\boldsymbol{\mu}_{ij}\|$, where $\boldsymbol{\mu}_{ij}$ is the vector of eigenvalues of the matrix $A_{ij} \triangleq \nabla_j h(\mathbf{z}, \mathcal{D}, i)$.

It is left to show that A_{ij} is bounded. This obvious when we write out the individual terms of h from (5). It can be seen that if $j \leq m$, then $A_{ij} = I$. Otherwise, it is the matrix of partial derivatives of Q, which are bounded by Assumption IV.1. $\qquad\square$

Corollary A.7 (Bound for $\|\boldsymbol{\delta}_{s,k}\|$). *Let $A_{ij} \triangleq \nabla_j h(\mathbf{z}, \mathcal{D}, i)$. The constant L_h and the constant D, which are bounds for $\|h'_+(\mathbf{z}, \mathcal{D}, i)\|$ and $\|\boldsymbol{\delta}_{s,k}\|$, respectively, are given by $L_h = \sqrt{\bar{\mu}_1^2 + \cdots + \bar{\mu}_{m+qn}^2}$, and $D = 2L_h$, where $\bar{\mu}_j \triangleq \max_{i\in\mathcal{I}}\{\|\boldsymbol{\mu}_{ij}\| \mid \boldsymbol{\mu}_{ij}\}$ is the vector of eigenvalues of A_{ij}.*

Proof. The j-th entry of h'_+ is a restriction of ψ_{ij} to a subset of its domain. Thus, bounds for ψ_{ij} also hold for the j-th entry of h'_+. In the proof of Proposition A.6, we saw that ψ_{ij} is bounded by the 2-norm of the eigenvalues of $A_{ij} \triangleq \nabla_j h(\mathbf{z}, \mathcal{D}, i)$. Therefore, $L_h = \|\boldsymbol{\mu}_{ij}\|_2$. Taking the norm of both sides of (14), we have that

$$\|\boldsymbol{\delta}_{s,k}\| = \left\| h'_+(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h'_+(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k}) \right\|.$$

Thus, the j-th entry of $\boldsymbol{\delta}_{s,k}$ is bounded by $2\bar{\mu}_j$ and $\|\boldsymbol{\delta}_{s,k}\|$. Taking the norm, we retrieve the result of the Corollary. $\square$

Proposition A.8 (Bounded Derivatives of ψ_{ij}). *Assume $X \in \mathrm{Sym}_+(p, \mathbb{R})$. Then, the magnitudes of the partial derivatives $\|\nabla_{\mathcal{D}} \psi_{ij}(X, \mathbf{z}, \mathcal{D})\|$ and $\|\nabla_X \psi_{ij}(X, \mathbf{z}, \mathcal{D})\|$ are pointwise bounded for all $i \in \mathcal{I}$ and $j \in \{1, \ldots, m + qn\}$.*

Proof. If $\|X\| = 0$, then ψ_{ij} is constant, and thus the magnitudes of its partial derivatives are zero. The rest of the proof deals with $\|X\| > 0$.

To simplify notation, we refer to $\nabla_j h(\mathbf{z}, \mathcal{D}, i)$ by the shorthand A_{ij}. Choose $\epsilon \in \mathbb{R}$ such that $\|X\| > \epsilon > 0$. Fix arbitrary i. Since X and A_{ij} are symmetric, for any j we have that $\nabla_X \psi_{ij}(X, \mathbf{z}, \mathcal{D}) = \frac{1}{\|X\|} A_{ij} - \frac{\mathrm{Tr} A_{ij} X}{\|X\|^3} X$ and

$$\|\nabla_X \psi_{ij}(X, \mathbf{z}, \mathcal{D})\|^2 = \|X\|^{-4} (\|A_{ij}\|^2 \|X\|^2 - (\mathrm{Tr} A_{ij} X)^2), \tag{36}$$

which is strictly less than 1 for all $\|X\| > \|A_{ij}\|$. For $\|X\| \leq \|A_{ij}\|$, $(\|A_{ij}\|/\epsilon)^4$ is an upper bound for (36).

For the term $\nabla_{\mathcal{D}} \psi_{ij}(X, \mathbf{z}, \mathcal{D})$, note that the only part of ψ_{ij} that depends on $\mathcal{D}$ is A_{ij}. Thus by the chain rule,

$$\nabla_{\mathcal{D}} \psi_{ij}(X, \mathbf{z}, \mathcal{D}) = \left(\nabla_{A_{ij}} \psi_{ij}(X, \mathbf{z}, \mathcal{D}) \right) \left(\nabla_{\mathcal{D}} A_{ij} \right).$$

The first term is given by $\nabla_{A_{ij}} \psi_{ij}(X, \mathbf{z}, \mathcal{D}) = \frac{1}{\|X\|_F} I \prec \frac{1}{\epsilon} I$. For the second term, recall that $\mathcal{D} = (\hat{\mathbf{x}}, S)$. Split the partial derivative into two parts $\nabla_{\mathcal{D}} = (\nabla_{\hat{\mathbf{x}}}, \nabla_S)$ so that

$$\nabla_{\mathcal{D}} \nabla_j h(\mathbf{z}, \mathcal{D}, i) = (\nabla_{\hat{\mathbf{x}}} \nabla_j h(\mathbf{z}, \mathcal{D}, i), \nabla_S \nabla_j h(\mathbf{z}, \mathcal{D}, i)).$$

For any j, $\nabla_S \nabla_j h(\mathbf{z}, \mathcal{D}, i) = I_p$, and obviously $\|I_p\| \leq p$. Similarly, $\nabla_j h(\mathbf{z}, \mathcal{D}, i) = I_p$ for $j \leq m$. If $j > m$, then $\nabla_{\hat{\mathbf{x}}} \nabla_j h(\mathbf{z}, \mathcal{D}, i)$ results in second mixed partial derivatives of Q, which we assumed to be bounded in Assumption IV.1. $\square$

Proposition A.9 (Local Differentiability of Constraint Gradient). *Consider an arbitrary $\mathbf{z} \in \mathbb{R}^{m+qn}$ and $\mathcal{D} = (\hat{\mathbf{x}}, S)$ with $\hat{\mathbf{x}}_i \in \mathbb{R}^p$ and $S_i \in \mathrm{Sym}_+(p, \mathbb{R})$. Suppose $h(\mathbf{z}; \mathcal{D}, i)$ is nonsigular with distinct eigenvalues. Then, $\exists$ an open neighborhood Ω containing $h(\mathbf{z}; \mathcal{D}, i)$ such that $\psi_{ij}(\Pi_+(\cdot), \mathbf{z}, \mathcal{D})\big|_\Omega$ is differentiable.*

Proof. **Notation** Note that $f \in C^\infty(\mathcal{A}, \mathcal{B})$ means that the n-th derivative of f exists at all points in $\mathcal{A}$ for all $n \in \mathbb{N}$. We also use the notation $f|_\Omega$ to denote the restriction of f to Ω.

Roughly speaking, the proof uses a well known result [38] to show that, if $h(\mathbf{z}; \mathcal{D}, i)$ has distinct eigenvalues, then for $\alpha = 1, \ldots, p$ we can find an open neighborhood Ω in which there are differentiable maps $\lambda_\alpha \in C^\infty(\Omega_\alpha, \mathbb{R})$ and $\mathbf{v}_\alpha \in C^\infty(\Omega_\alpha, \mathbb{R}^p)$ that represent the α-th eigenvalue-eigenvector pair of each $X \in \Omega$ and agree with the eigenvalues and eigenvectors of $h(\mathbf{z}; \mathcal{D}, i)$. Since $\psi_{ij}(\Pi_+(\cdot), \mathbf{z}, \Sigma)$ is well

behaved in an open neighborhood, it is not difficult to arrive at the desired result. The details of the proof are as follows, and generalize the line of reasoning in [38].

Let $(\bar{\lambda}_\alpha, \bar{\mathbf{v}}_\alpha)$ be an eigenvalue-eigenvector pair of $h(\mathbf{z}; \mathcal{D}, i)$. Since $h(\mathbf{z}; \mathcal{D}, i) \in \mathrm{Sym}(p, \mathbb{R})$, we know that $\bar{\lambda}_\alpha$ is real valued and distinct. Consider the mapping $\phi \colon \mathbb{R} \times \mathbb{R}^p \times \mathrm{Sym}(p, \mathbb{R}) \to \mathbb{R}^{p+1}$ given by

$$\phi(\lambda, \mathbf{u}, X) \triangleq \begin{bmatrix} (\lambda I - X)\mathbf{v} \\ \mathbf{v}^\top \mathbf{v} - 1 \end{bmatrix}. \tag{37}$$

By definition, $\phi(\lambda, \mathbf{u}, X) = \mathbf{0}$ if and only if $(\lambda, \mathbf{v})$ are a (normalized) eigenvalue-eigenvector pair of X. It is also a well known properties of eigenvalues and eigenvectors that

$$\nabla_{(\lambda, \mathbf{u})} \phi \left(\bar{\lambda}_\alpha, \bar{\mathbf{v}}_\alpha, h(\mathbf{z}; \mathcal{D}, i) \right)$$

is full rank if and only if $\bar{\lambda}_\alpha$ is simple, which is true by assumption. By the implicit function theorem, there is an open neighborhood Ω_α of $h(\mathbf{z}; \mathcal{D}, i)$ in which we can find differentiable functions $\lambda_\alpha \in C^\infty(\Omega_\alpha, \mathbb{R})$ and $\mathbf{v}_\alpha \in C^\infty(\Omega_\alpha, \mathbb{R}^p)$ such that $\lambda_\alpha(h(\mathbf{z}; \mathcal{D}, i)) = \bar{\lambda}_\alpha$ and $\mathbf{v}_\alpha(h(\mathbf{z}; \mathcal{D}, i)) = \bar{\mathbf{v}}_\alpha$ Let

$$\Omega_\alpha^+ = \{ X \in \Omega_\alpha \mid \lambda_\alpha(X) > 0 \}, \quad \text{and} \tag{38a}$$
$$\Omega_\alpha^- = \{ X \in \Omega_\alpha \mid \lambda_\alpha(X) < 0 \}. \tag{38b}$$

If $\bar{\lambda}_\alpha$ is positive for all $\alpha = 1, \ldots, p$, every Ω_α^+ is an open set that contains $h(\mathbf{z}; \mathcal{D}, i)$. Let

$$\Omega = \cap_\alpha \Omega_\alpha^+, \tag{39}$$

which is nonempty and open. When restricted to Ω, the mapping Π_+ is the identity map, and thus (at least for this case) the proof follows if $\psi_{ij}(\cdot, \mathbf{z}, \mathcal{D})\big|_\Omega$ is differentiable. To verify that it is, consider arbitrary $X \in \Omega$. Recall again the definition of $\psi_{ij}(\cdot, \mathbf{z}, \mathcal{D}) \colon \mathrm{Sym}(p, \mathbb{R}) \to \mathbb{R}$ from (11). Note that $\nabla_j h(\mathbf{z}, \mathcal{D}, i)$ is a constant matrix because h is a linear function of $\mathbf{z}$. Call this matrix A_{ij}. We have that

$$\psi_{ij}(X, \mathbf{z}, \mathcal{D}) = \frac{\mathrm{Tr}(A_{ij} X)}{\|X\|_F} = \frac{\sum_\alpha \lambda_\alpha(X) \mathbf{v}_\alpha^\top(X) A_{ij} \mathbf{v}_\alpha(X)}{\sqrt{\sum_\alpha \lambda_\alpha^2(X)}},$$

which uses the same algebraic manipulation as the derivation of (35b). By inspection, $\psi_{ij}(\cdot, \mathbf{z}, \mathcal{D})$ is differentiable at all $X \in \Omega$, and the result follows.

If $\bar{\lambda}_\alpha$ is negative for all $\alpha = 1, \ldots, p$, every Ω_α^- is an open set that contains $h(\mathbf{z}; \mathcal{D}, i)$. Let

$$\Omega = \cap_\alpha \Omega_\alpha^-, \tag{40}$$

which is again open and nonempty. Note that Π_+ projects every matrix $X \in \Omega$ to the matrix of zeros. The map $\psi_{ij}(\mathbf{0}, \mathbf{z}, \mathcal{D})$ is thus constant w.r.t. $X \in \Omega$ and is thus differentiable.

If $h(\mathbf{z}; \mathcal{D}, i)$ has some positive and negative eigenvalues, then we can split them into two groups. Let α index the positive eigenvalues and β denote the negative eigenvalues. Similar to before, let

$$\Omega = \left(\cap_\alpha \Omega_\alpha^+ \right) \cup \left(\cap_\beta \Omega_\beta^- \right), \tag{41}$$

which is again nonempty, open, and contains $h(\mathbf{z}; \mathcal{D}, i)$. Note that Π_+ does not change λ_α or $\mathbf{v}_\alpha$ for matrices $X \in \Omega$, i.e.,

$$\lambda_\alpha(X) = \lambda_\alpha(\Pi_+(X)), \ \forall X \in \Omega \tag{42a}$$
$$\mathbf{v}_\alpha(X) = \mathbf{v}_\alpha(\Pi_+(X)), \ \forall X \in \Omega. \tag{42b}$$

The eigenvalues corresponding to β of $\Pi_+(X)$, on the other hand, are always projected to zero. Thus we write that $\lambda_\beta(\Pi_+(X)) \equiv 0$, and we have, omitting some elementary algebra, that

$$\psi_{ij}\left(\Pi_+(X), \mathbf{z}, \mathcal{D}\right) = \frac{\operatorname{Tr}\left(A_{ij}\Pi_+(X)\right)}{\|\Pi_+(X)\|_F} = \mathbf{v}_1^\top(X)A_{ij}\mathbf{v}_1(X).$$

By inspection, $\psi_{ij}\left(\Pi_+(\cdot), \mathbf{z}, \mathcal{D}\right)$ is differentiable at X. Since X was an arbitrary matrix from the set Ω, the result follows.

Since $h(\mathbf{z}; \mathcal{D}, i)$ was assumed to be nonsingular, all possibilities of combinations eigenvalues have been covered, and, for each case, we found an open set Ω in which the desired result holds. $\square$

Proof of Lemma V.6. Recall the definition

$$\boldsymbol{\delta}_{s,k} \triangleq h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, \omega_{s,k}\right),$$

where $\{\omega_{s,k}\}_{k\in\mathbb{N}} \subset \mathcal{I}$. Consider sequences of the form $\{i, i, i, \dots\}$. We will use the fact that

$$\sum_{k\in\mathbb{N}} \|\boldsymbol{\delta}_{s,k}\| \leq \sum_{k\in\mathbb{N}}\sum_{i\in\mathcal{I}} \left\|h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)\right\|$$

to show that that $\boldsymbol{\delta}_{s,k}$ is summable, provided that each sequence $\left\|h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)\right\|$ is summable.

To show that $h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)$ is summable, we will show that the sequence of the entries for an arbitrary entry $j \in \{1, \dots, m+qn\}$ is summable. If this is true, the norm of the whole vector will be summable as well because $\|\cdot\|_a \geq \|\cdot\|_b$. if $a \leq b$. Fix $j \in \{1, \dots, m+qn\}$.
Case 1 If $j \leq m$, then this entry corresponds to the variable from the original problem that we denoted as γ_j. Note that, for this choice of j, we have that

$$\nabla_j h(\mathbf{z}, \mathcal{D}, i) = \begin{cases} I_p & \text{if } i = j \\ \mathbf{0} & \text{else} \end{cases} \tag{43}$$

for all possible variables $\mathbf{z}$ or data $\mathcal{D}$. Therefore, the j-th entry of $h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) - h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)$, which is the difference of two functions

$$\nabla_j h_+(\mathbf{z}_{s,k}, \mathcal{D}_{s,k}, i) - \nabla_j h_+(\mathbf{z}_{s,k}, \mathcal{D}, i) = 0 \tag{44}$$

because these functions are actually the same constant.
Case 2 If $j > m$, then we are looking at an entry of the gradient with respect to the control variables $\mathbf{u}_s$, which are non trivial. In this case, recall from the definition in (11) that $\nabla_j h_+(\mathbf{z}, \mathcal{D}, i) = \psi_{ij}\left(\Pi_+\left(h\left(\mathbf{z}; \mathcal{D}, i\right)\right), \mathbf{z}, \mathcal{D}\right)$. In other words, we can look at $\nabla_j h_+$ as essentially a composition of three functions: $\psi_{ij}, \Pi_+,$ and h. In the preceding lemmas and propositions, we have established useful facts about each of these constituent functions that will enable us to complete the proof.

Let $X_k = h\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right)$ and $Y_k = h\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)$, and note that $X_k \to Y_k$ because h is continuous and $\mathcal{D}_{s,k} \to \mathcal{D}$. By Proposition A.5, both sequences $\{\mathbf{z}_{s,k}, \mathcal{D}_{s,k}\}_{k\in\mathbb{N}}$ and $\{\mathbf{z}_{s,k}, \mathcal{D}\}_{k\in\mathbb{N}}$ almost surely lie in the preimage of the set of nonsingular matrices that have distinct eigenvalues with respect to the mapping $h(\cdot; \cdot, i)$. Let Ω_k be an open neighborhood of Y_k such that $\psi_{ij}\left(\Pi_+(\cdot), \mathbf{z}, \mathcal{D}\right)\big|_{\Omega_k}$ is differentiable.

Such a neighborhood is guaranteed to be available by Proposition A.9 as long as Y_k converges, and the limit point is nonsingular and does not have a simple spectrum. We know that Y_k converges by construction of the decreasing step sizes in Algorithm 1, and w.p.1. this limit point will be nonsingular and have a nonsimple spectrum by Proposition A.5. Then, we can find $K \in \mathbb{N}$ such that $\forall \ell \geq K$, $X_\ell \in \Omega_\ell$.

By Proposition A.8, the derivatives of $\psi_{ij}\left(\Pi_+(X), \mathbf{z}, \mathcal{D}\right)$ with respect to both X and $\mathcal{D}$ are bounded. Denote the larger bound by μ_ℓ. Proposition A.6 bounds ψ_{ij} globally almost surely. We have that

$$
\begin{aligned}
h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}_{s,k}, i\right) &- h'_+\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right) = \\
&= \psi_{ij}\left(\Pi_+\left(h\left(\mathbf{z}_{s,\ell}; \mathcal{D}_{s,\ell}, i\right)\right), \mathbf{z}_{s,\ell}, \mathcal{D}_{s,\ell}\right) \\
&\qquad - \psi_{ij}\left(\Pi_+\left(h\left(\mathbf{z}_{s,\ell}; \mathcal{D}, i\right)\right), \mathbf{z}_{s,\ell}, \mathcal{D}\right) \\
&= \psi_{ij}\left(\Pi_+(X_\ell), \mathbf{z}_{s,\ell}, \mathcal{D}_{s,\ell}\right) - \psi_{ij}\left(\Pi_+(Y_\ell), \mathbf{z}_{s,\ell}, \mathcal{D}\right) \\
&= \langle \nabla_X(\psi_{ij}\circ\Pi_+)(Y_\ell, \mathbf{z}_{s,\ell}, \mathcal{D}), X_\ell - Y_\ell \rangle_{\operatorname{Sym}(p,\mathbb{R})} \\
&\qquad + o(\|X^\ell - Y^\ell\|) \\
&\leq \|\nabla_X(\psi_{ij}\circ\Pi_+)(Y_\ell, \mathbf{z}_{s,\ell}, \mathcal{D})\|\|X_\ell - Y_\ell\| + o\left(\|X_\ell - Y_\ell\|\right) \\
&\leq \mu_\ell\|X_\ell - Y_\ell\| + o\left(\|X_\ell - Y_\ell\|\right),
\end{aligned}
$$

where $\langle\cdot,\cdot\rangle_{\operatorname{Sym}(p,\mathbb{R})}$ denotes the matrix inner product. The final term is summable if and only if $\|X_\ell - Y_\ell\|$ is summable. To verify that it is, note that

$$\|X_\ell - Y_\ell\| = \|h\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right) - h\left(\mathbf{z}_{s,k}; \mathcal{D}, i\right)\| \tag{45}$$

which is summable by Proposition A.1. $\square$

APPENDIX B
PROOF OF LEMMAS V.7-V.9

Proof of Lemma V.7. From Line 13 of Algorithm 1, the following chain of relations hold for any feasible point $\mathbf{z} \in \mathcal{X} \subseteq \mathcal{X}_0$ a.s.:

$$
\begin{aligned}
\|\mathbf{z}_{s,k} - \mathbf{z}\|^2 &\leq \|\mathbf{v}_{s,k} - \mathbf{z} - \beta_{s,k}h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2 & \text{(46a)} \\
&\leq \|\mathbf{v}_{s,k} - \mathbf{z}\|^2 - 2\beta_{s,k}\langle h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}), \mathbf{v}_{s,k} - \mathbf{z}\rangle \\
&\quad + \beta_{s,k}^2\|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2 & \text{(46b)} \\
&\leq \|\mathbf{v}_{s,k} - \mathbf{z}\|^2 - 2\beta_{s,k}\langle h'_+(\mathbf{v}_{s,k}; \mathcal{D}, \omega_{s,k}), \mathbf{v}_{s,k} - \mathbf{z}\rangle & \text{(46c)} \\
&\quad + 2\beta_{s,k}\|\boldsymbol{\delta}_{s,k}\|\|\mathbf{v}_{s,k} - \mathbf{z}\| + \beta_{s,k}^2\|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2 \\
&\leq \|\mathbf{v}_{s,k} - \mathbf{z}\|^2 - 2\beta_{s,k}h_+(\mathbf{v}_{s,k}; \mathcal{D}, \omega_{s,k}) & \text{(46d)} \\
&\quad + 2\beta_{s,k}\|\boldsymbol{\delta}_{s,k}\|\|\mathbf{v}_{s,k} - \mathbf{z}\| + \beta_{s,k}^2\|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2,
\end{aligned}
$$

where (46a) follows from the nonexpansiveness of the projection operator; (46b) follows from the expansion of $\|\cdot\|^2$; (46c) follows from relation (14) and the Schwarz inequality; and (46d) follows from the convexity of the function $h_+(\cdot; \cdot, \omega_{s,k})$, and $h_+(\mathbf{z}; \mathcal{D}, \omega_{s,k}) = 0$ due to the feasibility of the point $\mathbf{z}$,

i.e., $h_+(\mathbf{v}_{s,k}; \mathcal{D}, \omega_{s,k}) \le \langle h'_+(\mathbf{v}_{s,k}; \mathcal{D}, \omega_{s,k}), \mathbf{v}_{s,k} - \mathbf{z} \rangle$. From relation (13) and the definition of $\beta_{s,k}$ in (12), we further have

$$-2\beta_{s,k}h_+(\mathbf{v}_{s,k}; \mathcal{D}, \omega_{s,k}) + \beta_{s,k}^2 \|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2$$
$$= -2\beta_{s,k}h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) + 2\beta_{s,k}|\nu_{s,k}|$$
$$+ \beta_{s,k}^2 \|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2$$
$$= 2\beta_{s,k}|\nu_{s,k}| - \frac{h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{\|h'_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\|^2}$$
$$\le 2\beta_{s,k}|\nu_{s,k}| - \frac{h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{(L_h + D)^2},$$

where the last inequality follows from relation (18a). Combining this with inequality (46d), we obtain for any $\mathbf{z} \in \mathcal{X}$ a.s.:

$$\|\mathbf{z}_{s,k} - \mathbf{z}\|^2 \le \|\mathbf{v}_{s,k} - \mathbf{z}\|^2 + 2\beta_{s,k}\|\boldsymbol{\delta}_{s,k}\|\|\mathbf{v}_{s,k} - \mathbf{z}\| \quad (47)$$
$$+ 2\beta_{s,k}|\nu_{s,k}| - \frac{h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{(L_h + D)^2}.$$

For the last term on the right-hand side, we can rewrite $h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) = h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) + h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$. Therefore,

$$h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \ge 2h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\times (h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}))$$
$$+ h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}). \quad (48)$$

Regarding the first term on the right-hand side of (48), the following chain of relations holds:

$$2h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\times (h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}))$$
$$\ge -2h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\times |h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})| \quad (49a)$$
$$\ge -2(L_h + D)\|\mathbf{v}_{s,k} - \mathbf{p}_{s,k}\|h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \quad (49b)$$
$$\ge -2\alpha_k(L_h + D)L_f h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \quad (49c)$$
$$\ge -\tau\alpha_k^2(L_h + D)^2 L_f^2 - \frac{1}{\tau}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}), \quad (49d)$$

where (49a) follows from the Schwarz inequality; (49b) follows from relation (18b); (49c) follows from Line 10 of Algorithm 1; and (49d) follows from using $2|a||b| \le \tau a^2 + \frac{1}{\tau}b^2$ with $a = \alpha_k(L_h + D)L_f$ and $b = h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$, and $\tau > 0$ is arbitrary. Using relations (48)-(49) in (47), we obtain for all $\mathbf{z} \in \mathcal{X}$ a.s.:

$$\|\mathbf{z}_{s,k} - \mathbf{z}\|^2 \le \|\mathbf{v}_{s,k} - \mathbf{z}\|^2 + 2\beta_{s,k}\|\boldsymbol{\delta}_{s,k}\|\|\mathbf{v}_{s,k} - \mathbf{z}\| \quad (50)$$
$$+ 2\beta_{s,k}|\nu_{s,k}| - \frac{\tau - 1}{\tau(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) + \tau\alpha_k^2 L_f^2.$$

Similarly to (46a)-(46d), from Line 10 of Algorithm 1, the following chain of relations hold for any $\mathbf{z} \in \mathcal{X} \subseteq \mathcal{X}_0$ a.s.:

$$\|\mathbf{v}_{s,k} - \mathbf{z}\|^2 \le \|\mathbf{p}_{s,k} - \mathbf{z} - \alpha_k f'(\mathbf{p}_{s,k})\|^2 \quad (51a)$$
$$\le \|\mathbf{p}_{s,k} - \mathbf{z}\|^2 - 2\alpha_k\langle f'(\mathbf{p}_{s,k}), \mathbf{p}_{s,k} - \mathbf{z}\rangle + \alpha_k^2\|f'(\mathbf{p}_{s,k})\|^2 \quad (51b)$$
$$\le \|\mathbf{p}_{s,k} - \mathbf{z}\|^2 - 2\alpha_k(f(\mathbf{p}_{s,k}) - f(\mathbf{z})) + \alpha_k^2 L_f^2, \quad (51c)$$

where (51a) follows from the nonexpansive projection; (51b) follows from the expansion of $\|\cdot\|^2$; (51c) follows from the convexity of the function f and relation (16a). For the second term on the right-hand side, we further have for any $\check{\mathbf{z}} \in \mathcal{X}$:

$$-2\alpha_k(f(\mathbf{p}_{s,k}) - f(\mathbf{z}))$$
$$= -2\alpha_k(f(\mathbf{p}_{s,k}) - f(\check{\mathbf{z}}) + f(\check{\mathbf{z}}) - f(\mathbf{z})) \quad (52a)$$
$$\le -2\alpha_k\langle f'(\check{\mathbf{z}}), \mathbf{p}_{s,k} - \check{\mathbf{z}}\rangle - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z})) \quad (52b)$$
$$\le 2\alpha_k L_f\|\mathbf{p}_{s,k} - \check{\mathbf{z}}\| - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z})) \quad (52c)$$
$$\le 4\eta\alpha_k^2 L_f^2 + \frac{1}{4\eta}\|\mathbf{p}_{s,k} - \check{\mathbf{z}}\|^2 - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z})), \quad (52d)$$

where (52a) follows from adding and subtracting $f(\check{\mathbf{z}})$; (52b) follows from the convexity of the function f; (52c) follows from the Schwarz inequality; and (52d) follows from using $2|a||b| \le 4\eta a^2 + \frac{1}{4\eta}b^2$ with $a = \alpha_k L_f$ and $b = \|\mathbf{p}_{s,k} - \check{\mathbf{z}}\|$. Combining (52d) in (51c), we have for any $\mathbf{z}, \check{\mathbf{z}} \in \mathcal{X} \subseteq \mathcal{X}_0$ a.s.:

$$\|\mathbf{v}_{s,k} - \mathbf{z}\|^2 \le \|\mathbf{p}_{s,k} - \mathbf{z}\|^2 - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z}))$$
$$+ \frac{1}{4\eta}\|\mathbf{p}_{s,k} - \check{\mathbf{z}}\|^2 + (1 + 4\eta)\alpha_k^2 L_f^2.$$

Substituting this inequality in relation (50), we obtain

$$\|\mathbf{z}_{s,k} - \mathbf{z}\|^2 \le \|\mathbf{p}_{s,k} - \mathbf{z}\|^2 - 2\alpha_k(f(\check{\mathbf{z}}) - f(\mathbf{z}))$$
$$+ 2\beta_{s,k}\|\boldsymbol{\delta}_{s,k}\|\|\mathbf{v}_{s,k} - \mathbf{z}\| + 2\beta_{s,k}|\nu_{s,k}|$$
$$+ \frac{1}{4\eta}\|\mathbf{p}_{s,k} - \check{\mathbf{z}}\|^2 + A_{\eta,\tau}\alpha_k^2 L_f^2$$
$$- \frac{\tau - 1}{\tau(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}),$$

where $A_{\eta,\tau} = 1 + 4\eta + \tau$. From relation (15), we have $\|\mathbf{v}_{s,k} - \mathbf{z}\| \le C_\mathbf{z}$. Using this and the upper estimate for $\beta_{s,k}$ in (21) for bounding the three error terms completes the proof. $\square$

Proof of Lemma V.8. We use Lemma V.7 with $\check{\mathbf{z}} = \mathbf{z} = \Pi_\mathcal{X}[\mathbf{p}_{s,k}]$. Therefore, for any $s \in \mathcal{S}$, and $k \ge 1$, we obtain a.s.:

$$\|\mathbf{z}_{s,k} - \Pi_\mathcal{X}[\mathbf{p}_{s,k}]\|^2 \le \|\mathbf{p}_{s,k} - \Pi_\mathcal{X}[\mathbf{p}_{s,k}]\|^2 \quad (53)$$
$$+ 2\frac{C_\mathbf{z}(C_h + N)}{c_h^2}\|\boldsymbol{\delta}_{s,k}\| + 2\frac{C_h + N}{c_h^2}|\nu_{s,k}|$$
$$+ \frac{1}{4\eta}\|\mathbf{p}_{s,k} - \Pi_\mathcal{X}[\mathbf{p}_{s,k}]\|^2 + A_{\eta,\tau}\alpha_k^2 L_f^2$$
$$- \frac{\tau - 1}{\tau(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}),$$

where $A_{\eta,\tau} = 1 + 4\eta + \tau$ and $\eta, \tau > 0$ are arbitrary. By the definition of the projection, we have $\text{dist}(\mathbf{p}_{s,k}, \mathcal{X}) = \|\mathbf{p}_{s,k} - \Pi_\mathcal{X}[\mathbf{p}_{s,k}]\|$ and $\text{dist}(\mathbf{z}_{s,k}, \mathcal{X}) = \|\mathbf{z}_{s,k} - \Pi_\mathcal{X}[\mathbf{z}_{s,k}]\| \le \|\mathbf{z}_{s,k} - \Pi_\mathcal{X}[\mathbf{p}_{s,k}]\|$. Upon substituting these estimates in relation (53), we obtain

$$\text{dist}^2(\mathbf{z}_{s,k}, \mathcal{X}) \le \text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) \quad (54)$$
$$+ 2\frac{C_\mathbf{z}(C_h + N)}{c_h^2}\|\boldsymbol{\delta}_{s,k}\| + 2\frac{C_h + N}{c_h^2}|\nu_{s,k}|$$
$$+ \frac{1}{4\eta}\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A_{\eta,\tau}\alpha_k^2 L_f^2$$
$$- \frac{\tau - 1}{\tau(L_h + D)^2}h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}).$$

Taking the expectation conditioned on $\mathcal{F}_{k-1}$ and noting that $\mathbf{p}_{s,k}$ is fully determined by $\mathcal{F}_{k-1}$, we have for any $s \in \mathcal{S}$ and $k \geq 1$ *a.s.*:

$$\mathsf{E}\left[\text{dist}^2(\mathbf{z}_{s,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \leq \text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) \tag{55}$$
$$+2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\mathsf{E}[\|\boldsymbol{\delta}_{s,k}\| \mid \mathcal{F}_{k-1}] + 2\frac{C_h + N}{c_h^2}\mathsf{E}[|\nu_{s,k}| \mid \mathcal{F}_{k-1}]$$
$$+\frac{1}{4\eta}\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A_{\eta,\tau}\alpha_k^2 L_f^2$$
$$-\frac{\tau - 1}{\tau(L_h + D)^2}\mathsf{E}\left[h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \mid \mathcal{F}_{k-1}\right].$$

We now choose $\tau = 4$, $\eta = \kappa(L_h + D)^2$ and use Assumption IV.3 to yield

$$\mathsf{E}\left[\text{dist}^2(\mathbf{z}_{s,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \leq \text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X})$$
$$+2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\mathsf{E}[\|\boldsymbol{\delta}_{s,k}\| \mid \mathcal{F}_{k-1}] + 2\frac{C_h + N}{c_h^2}\mathsf{E}[|\nu_{s,k}| \mid \mathcal{F}_{k-1}]$$
$$-\frac{1}{2\kappa(L_h + D)^2}\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A\alpha_k^2 L_f^2,$$

where $A = 5 + 4\kappa(L_h + D)^2$. Finally, by summing over all $s \in \mathcal{S}$ and using Lemma V.3 with $h(x) = \text{dist}^2(x, \mathcal{X})$, we arrive at the following relation:

$$\sum_{s \in \mathcal{S}} \mathsf{E}\left[\text{dist}^2(\mathbf{z}_{s,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \leq \sum_{s \in \mathcal{S}} \text{dist}^2(\mathbf{z}_{s,k-1}, \mathcal{X}) \tag{56}$$
$$+2\frac{C_{\mathbf{z}}(C_h + N)}{c_h^2}\sum_{s \in \mathcal{S}}\mathsf{E}[\|\boldsymbol{\delta}_{s,k}\| \mid \mathcal{F}_{k-1}]$$
$$+2\frac{C_h + N}{c_h^2}\sum_{s \in \mathcal{S}}\mathsf{E}[|\nu_{s,k}| \mid \mathcal{F}_{k-1}]$$
$$-\frac{1}{2\kappa(L_h + D)^2}\sum_{s \in \mathcal{S}}\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X}) + A\alpha_k^2 n L_f^2.$$

Therefore, from $\sum_{k \geq 0} \alpha_k^2 < \infty$ and Lemmas V.5 and V.6, all the conditions in the convergence theorem (Theorem V.2) are satisfied and the desired result follows. $\square$

Proof of Lemma V.9(a). From Line 6-10 and 13 of Algorithm 1, we define $\mathbf{e}_{s,k} \triangleq \mathbf{z}_{s,k} - \mathbf{p}_{s,k}$ for $s \in \mathcal{S}$ and $k \geq 0$. Hence, $\mathbf{e}_{s,k}$ can be viewed as the perturbation that we make on $\mathbf{p}_{s,k}$ after the network consensus step in Line 6 of Algorithm 1. Consider $\|\mathbf{e}_{s,k}\|$, for which we can write

$$\|\mathbf{e}_{s,k}\| \leq \|\mathbf{z}_{s,k} - \mathbf{v}_{s,k}\| + \|\mathbf{v}_{s,k} - \mathbf{p}_{s,k}\|$$
$$= \left\|\Pi_{\mathcal{X}_0}\left[\mathbf{v}_{s,k} - \beta_{s,k}h_+'(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})\right] - \mathbf{v}_{s,k}\right\|$$
$$+ \left\|\Pi_{\mathcal{X}_0}[\mathbf{p}_{s,k} - \alpha_k f'(\mathbf{p}_{s,k})] - \mathbf{p}_{s,k}\right\|$$
$$\leq \frac{h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{c_h} + \alpha_k L_f, \tag{57}$$

where the second inequality follows from fact that $\mathbf{p}_{s,k}, \mathbf{v}_{s,k} \in \mathcal{X}_0$, the nonexpansiveness of the projection operator, relations (16a) and (20).

Applying $(a + b)^2 \leq 2a^2 + 2b^2$ in inequality (57), we have for all $s \in \mathcal{S}$ and $k \geq 1$,

$$\|\mathbf{e}_{s,k}\|^2 \leq 2\frac{h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{c_h^2} + 2\alpha_k^2 L_f^2 \tag{58}$$

For the term $h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$ in (58), we have the following chain of relations:

$$h_+^2(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$= (h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}))^2 \tag{59}$$
$$+ h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$+ 2h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\times (h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) - h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}))$$
$$\leq (L_h + D)^2\|\mathbf{v}_{s,k} - \mathbf{p}_{s,k}\|^2 + h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \tag{60}$$
$$+ 2(L_h + D)\|\mathbf{v}_{s,k} - \mathbf{p}_{s,k}\|h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\leq \alpha_k^2(L_h + D)^2 L_f^2 + h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \tag{61}$$
$$+ 2\alpha_k(L_h + D)L_f h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$$
$$\leq 2\alpha_k^2(L_h + D)^2 L_f^2 + h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}), \tag{62}$$

where (59) follows from adding and subtracting $h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$; (60) follows from relation (18b); (61) follows from Line 10 of Algorithm 1; (62) follows from using the relation $2|a||b| \leq a^2 + b^2$ with $a = \alpha_k(L_h + D)L_f$ and $b = h_+(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})$. By combining (58) and (62), we obtain

$$\|\mathbf{e}_{s,k}\|^2 \leq 2\alpha_k^2 L_f^2\left(1 + \frac{2(L_h + D)^2}{c_h^2}\right) + \frac{2h_+^2(\mathbf{p}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{c_h^2}$$
$$\leq 2\alpha_k^2 L_f^2\left(1 + \frac{2(L_h + D)^2}{c_h^2}\right) + \frac{2L_h^2\text{dist}^2(\mathbf{p}_{s,k}, \mathcal{X})}{c_h^2},$$

where the last inequality is from (17). Therefore, from Lemma V.8 and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, we conclude that $\sum_{k=1}^{\infty} \|\mathbf{e}_{s,k}\|^2 < \infty$ for all $s \in \mathcal{S}$ *a.s.* $\square$

Proof of Lemma V.9(b). The proof of this part of Lemma will coincide with [26, Lemma 6(b)]. $\square$

REFERENCES

[1] T. H. Chung, J. W. Burdick, and R. M. Murray, "A decentralized motion coordination strategy for dynamic target tracking," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2006, pp. 2416 –2422.

[2] P. Jalalkamali and R. Olfati-Saber, "Information-driven self-deployment and dynamic sensor coverage for mobile sensor networks," in *IEEE American Control Conf. (ACC)*, 2012, pp. 4933–4938.

[3] C. Freundlich, P. Mordohai, and M. M. Zavlanos, "Hybrid control for mobile target localization with stereo vision," in *IEEE Conf. on Decision and Control (CDC)*, 2013, pp. 2635–2640.

[4] ——, "Optimal path planning and resource allocation for active target localization," in *IEEE American Control Conf. (ACC)*, 2015, pp. 3088–3093.

[5] ——, "A hybrid control approach to the next-best-view problem using stereo vision," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 4478–4483.

[6] J. Vander Hook, P. Tokekar, and V. Isler, "Algorithms for cooperative active localization of static targets with mobile bearing sensors under communication constraints," *IEEE Trans. on Robotics*, vol. 31, no. 4, pp. 864–876, 2015.

[7] A. Simonetto and T. Keviczky, "Distributed multi-target tracking via mobile robotic networks: a localized non-iterative SDP approach," in *IEEE Conf. on Decision and Control (CDC)*, 2011, pp. 4226–4231.

[8] J. Derenick, J. Spletzer, and A. Hsieh, "An optimal approach to collaborative target tracking with performance guarantees," *J. Intelligent and Robotic Systems*, vol. 56, no. 1-2, pp. 47–67, 2009.

[9] A. T. Kamal, J. Farrell, and A. K. Roy-Chowdhury, "Information weighted consensus," in *IEEE Conf. on Decision and Control (CDC)*, 2012, pp. 2732–2737.

[10] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[11] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *J. Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[12] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed sparsity-aware sensor selection," *IEEE Trans. on Signal Processing*, vol. 63, no. 22, pp. 5951–5964, 2015.

[13] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," vol. 168, no. 1, pp. 172–197, 2016.

[14] E. D. Nerurkar, K. J. Wu, and S. I. Roumeliotis, "C-KLAM: Constrained keyframe-based localization and mapping," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2014, pp. 3638–3643.

[15] G. Huang, K. Zhou, N. Trawny, and S. I. Roumeliotis, "A bank of maximum a posteriori (MAP) estimators for target tracking," *IEEE Trans. on Robotics*, vol. 31, no. 1, pp. 85–103, 2015.

[16] B. J. Julian, M. Angermann, M. Schwager, and D. Rus, "Distributed robotic sensor networks: An information-theoretic approach," *Int. J. Robotics Research*, vol. 31, no. 10, pp. 1134–1154, 2012.

[17] M. Patan, *Optimal sensor networks scheduling in identification of distributed parameter systems*. Springer Science & Business Media, 2012, vol. 425.

[18] R. Khodayi-mehr, W. Aquino, and M. M. Zavlanos, "Model-based sparse source identification," in *IEEE American Control Conf. (ACC)*.

[19] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[20] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. on Signal Processing*, vol. 62, no. 5, pp. 1135–1146, 2014.

[21] S. P. Chepuri and G. Leus, "Compression schemes for time-varying sparse signals," in *48th Asilomar Conf. on Signals, Systems and Computers*. IEEE, 2014, pp. 1647–1651.

[22] M. Patan, "Decentralized mobile sensor routing for parameter estimation of distributed systems," *IFAC Proceedings Volumes*, vol. 42, no. 20, pp. 210–215, 2009.

[23] M. T. Spaan, T. S. Veiga, and P. U. Lima, "Decision-theoretic planning under uncertainty with information rewards for active cooperative perception," *Int. Found. for Autonomous Agents and Multiagent Sys. (IFAAMAS)*, pp. 1–29, 2014.

[24] S. Omidshafiei, A.-a. Agha-mohammadi, C. Amato, and J. P. How, "Decentralized control of partially observable markov decision processes using belief space macro-actions," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2015.

[25] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE J. Selected Topics in Signal Processing*, vol. 7, pp. 221–229, April 2013.

[26] S. Lee and M. Zavlanos, "Approximate projections for decentralized optimization with SDP constraints," 2015, http://arxiv.org/abs/1509.08007.

[27] B. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14 – 29, 1969.

[28] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2003.

[29] A. Lewis and J.-S. Pang, "Error bounds for convex inequality systems," in *Generalized Convexity*. Kluwer Academic Publishers, 1996, pp. 75–110.

[30] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, Sep. 1996.

[31] L. Gubin, B. Polyak, and E. Raik, "The method of projections for finding the common point of convex sets," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 6, pp. 1 – 24, 1967.

[32] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex analysis and optimization*. Athena Scientific, 2003.

[33] B. Polyak, *Intro. to optimization*. Optimization software, Inc., Publications division, New York, 1987.

[34] S. S. Ram, A. Nedić, and V. V. Veeravalli, "A new class of distributed optimization algorithms: application to regression of distributed data," *Optimization Methods and Software*, vol. 27, no. 1, pp. 71–88, 2012.

[35] B. M. M. Zavlanos, M. B. Egerstedt, and G. J. Pappas, "Graph-theoretic connectivity control of mobile robot networks," *Proceedings of the IEEE*, vol. 99, no. 9, pp. 1525–1540, 2011.

[36] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proceedings of 47th IEEE Conference on Decision and Control*, December 2008, pp. 4177 –4184.

[37] J. E. Marsden and M. J. Hoffman, *Elementary classical analysis*. Macmillan, 1993.

[38] J. R. Magnus, "On differentiating eigenvalues and eigenvectors," *Econometric Theory*, vol. 1, no. 2, pp. 179–191, 1985.

Charles Freundlich (S'13) received a B.A. in physics from Middlebury College, Middlebury, VT in 2010, an M.Eng. in mechanical engineering from Stevens Institute of Technology, Hoboken, NJ, in 2013, and a PhD in mechanical engineering from Duke University, Durham, NC, in 2017. As a doctorate student, he studied control theory, AI, and computer vision. Currently, Dr. Freundlich works on Supply Chain Automation Tesla, Inc. in Fremont, CA. His projects include cooperative SLAM and large-scale estimation.

Soomin Lee (S'07–M'13) is a Research Engineer at Georgia Institute of Technology. She was formerly a Postdoctoral Associate in Mechanical Engineering and Materials Science at Duke University. She received her Ph.D. in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign (2013). She received two master's degrees from the Korea Advanced Institute of Science and Technology in Electrical Engineering, and from the University of Illinois at Urbana-Champaign in Computer Science. In 2009, she was an assistant research officer at the Advanced Digital Science Center (ADSC) in Singapore. Her interests include theoretical optimization (convex, non-convex, online and stochastic), distributed control and optimization of dynamic networks, human-robot interactions and machine learning.

Michael Zavlanos (S'05–M'09) received the Diploma in mechanical engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2002, and the M.S.E. and Ph.D. degrees in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, in 2005 and 2008, respectively.

He is currently an Assistant Professor in the Department of Mechanical Engineering and Materials Science at Duke University, Durham, NC. He also holds a secondary appointment in the Department of Electrical and Computer Engineering and the Department of Computer Science. Prior to joining Duke University, Dr. Zavlanos was an Assistant Professor in the Department of Mechanical Engineering at Stevens Institute of Technology, Hoboken, NJ, and a Postdoctoral Researcher in the GRASP Lab, University of Pennsylvania, Philadelphia, PA. His research interests include a wide range of topics in the emerging discipline of networked systems, with applications in robotic, sensor, and communication networks. He is particularly interested in hybrid solution techniques, on the interface of control theory, distributed optimization, estimation, and networking.

Dr. Zavlanos is a recipient of various awards including the 2014 Office of Naval Research Young Investigator Program (YIP) Award and the 2011 National Science Foundation Faculty Early Career Development (CAREER) Award.