Multi-Robot Gaussian Process Estimation and Coverage: A Deterministic Sequencing Algorithm and Regret Analysis

Lai Wei Andrew McDonald Vaibhav Srivastava

***Abstract*— We study the problem of multi-robot coverage over an unknown, nonuniform sensory field. Modeling the sensory field as a realization of a Gaussian Process and using Bayesian techniques, we devise a policy which aims to balance the tradeoff between *learning* the sensory function and *covering* the environment. We propose an adaptive coverage algorithm called Deterministic Sequencing of Learning and Coverage (DSLC) that schedules learning and coverage epochs such that its emphasis gradually shifts from exploration to exploitation while never fully ceasing to learn. Using a novel definition of coverage regret which characterizes overall coverage performance of a multi-robot team over a time horizon T, we analyze DSLC to provide an upper bound on expected cumulative coverage regret. Finally, we illustrate the empirical performance of the algorithm through simulations of the coverage task over an unknown distribution of wildfires.**

I. Introduction

Autonomous systems must remain robust and resilient in the face of uncertainty, capable of making decisions under the influence of imperfect and incomplete information. Real-world environments are unpredictable, noisy, and stochastic by their nature—various factors including weather, terrain, and human behavior combine with changing mission goals and operating constraints to necessitate adaptive policies. To successfully deal with uncertainty, autonomous systems must strike a balance between exploration and exploitation, simultaneously learning about the environment while executing a task depending on their collective knowledge about it.

The *coverage problem* [1] arises naturally in multi-robot systems when a team of agents wishes to deploy themselves over an environment according to a particular sensory function ϕ, which specifies the degree to which a robot is "needed." Equivalently, the team of agents aims to partition an environment and achieve a configuration which minimizes the coverage cost defined by the sum of the ϕ-weighted distances from every point in the environment to the nearest agent. Example applications of coverage range from search and rescue to wildfire fighting, smart agriculture, ecological surveying, environmental cleanup, and climate monitoring.

Classical approaches to coverage control [1]–[4] assume *a priori* knowledge of ϕ and employ Lloyd's algorithm [5] to guarantee the convergence of agents to a local minimum of the coverage cost. In these algorithms, each agent communicates with the agents in the neighboring partitions at each time and updates its partition. Distributed *gossip-based* coverage algorithms [6] address potential communication bottlenecks in classical approaches by updating partitions pairwise between neighboring agents. While much of the work in coverage considers continuous convex environments, a discrete graph representation of the environment is considered and a corresponding gossip-based coverage algorithm is proposed in [7], which allows for non-convex environments.

Recent works have focused on the problem of *adaptive* coverage, in which agents are not assumed to have knowledge of ϕ *a priori*. Parametric estimation approaches to adaptive coverage [8], [9] model ϕ as a linear combination of basis functions and propose algorithms to learn the weight of each basis function, while non-parametric approaches [10]–[15] model ϕ as the realization of a Gaussian Process and make predictions by conditioning on observations of ϕ sampled over the operating environment. Todescato *et al.* [14] use a Bernoulli random variable for each robot to decide between learning and coverage steps. The probability of exploration decays as the estimation of ϕ becomes more accurate. Benevento *et al.* [15] use a Gaussian process optimization [16] based approach to design an adaptive coverage algorithm and derive an upper bound on the regret with respect to coverage cost. However, they make the strong assumption that Lloyd's algorithm converges to the global minimum of coverage cost. In contrast, the coverage regret in this paper is defined with respect to the local minima and consequently, the assumption is relaxed.

In this paper, we focus on a non-parametric adaptive coverage algorithm with provable regret guarantees. The major contributions of this work are threefold. First, we propose an adaptive coverage algorithm—Deterministic Sequencing of Learning and Coverage (DSLC)—that can balance the exploration-exploitation trade-off. Second, we introduce a novel coverage regret that characterizes the deviation of agent configurations and partitions from a centroidal Voronoi partition, and derive analytic bounds on the expected cumulative regret for DSLC. In particular, we prove that DSLC will achieve sublinear expected cumulative regret under minor assumptions. Third, we illustrate the efficacy of DSLC through extensive simulation and comparison with existing state-of-the-art approaches to adaptive coverage.

The remainder of the paper is organized as follows. The problem setup and preliminaries are presented in Section II. The DSLC algorithm is presented and analyzed in Sections III and IV, respectively. The performance of DSLC is elu-

This work has been supported in part by NSF grant IIS-1734272 and ARO grant W911NF-18-1-0325.

L. Wei and V. Srivastava are with the Department of Electrical and Computer Engineering. Michigan State University, East Lansing, MI 48823 USA. e-mail: {weilai1, vaibhav}@msu.edu

A. McDonald is with the Department of Computer Science and Engineering. Michigan State University, East Lansing, MI 48823 USA. e-mail: mcdon499@msu.edu

cidated through empirical simulations and is compared with the state-of-the-art algorithms in Section V. Conclusions and future directions are discussed in Section VI.

II. Problem Formulation

We consider a team of N agents tasked with providing coverage to a finite set of points in an environment represented by an undirected graph. The team is required to navigate within the graph to learn an unknown sensory function while maintaining a near optimal configuration. In this section, we present the preliminaries of the estimation and coverage problem.

A. Graph Representation of Environment

We consider a discrete environment modeled by an undirected graph $G = (V, E)$, where the vertex set V contains the finite set of points to be covered and the edge set $E \subseteq V \times V$ is the collection of physically adjacent pairs of vertices that can be reached from each other without passing through other vertices. Let the weight map $w : E \to \mathbb{R}_{>0}$ indicate the distance between connected vertices. We assume G is connected. Following standard definition of weighted undirected graph, a path in G is an ordered sequence of vertices where there exists an edge between consecutive vertices. The distance between vertices v_i and v_j in G, denoted by $d_G(v_i, v_j)$, is defined by the minimum of the sums of the weights in the paths between v_i and v_j.

Suppose there exists an unknown sensory function $\phi : V \to \mathbb{R}_{>0}$ that assigns a nonnegative weight to each vertex in G. Intuitively, $\phi(v_i)$ could represent the intensity of signal of interest such as brightness or sound. A robot at vertex v_i is capable of measuring $\phi(v_i)$ by collecting a sample $y = \phi(v_i) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is an additive zero mean Gaussian noise.

B. Nonparametric Estimation

Let $\boldsymbol{\phi}$ be a vector with the i-th entry $\phi(v_i)$, $i \in \{1, \ldots, |V|\}$, where $|\cdot|$ denotes set cardinality. We assume a multivariate Gaussian prior for $\boldsymbol{\phi}$ such that $\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0^{-1})$, where $\boldsymbol{\mu}_0$ is the mean vector and $\boldsymbol{\Lambda}_0$ is the inverse covariance matrix. Let $n_i(t)$ be the number of samples and $s_i(t)$ be the summation of sampling results from v_i until time t. Then, the posterior distribution of $\boldsymbol{\phi}$ at time t is $\mathcal{N}\big(\boldsymbol{\mu}(t), \boldsymbol{\Lambda}^{-1}(t)\big)$ [17, Chapter 10], where

$$\begin{aligned} \boldsymbol{\Lambda}(t) &= \boldsymbol{\Lambda}_0 + \sum_{i=1}^{|V|} \frac{n_i(t)}{\sigma^2} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}} \\ \boldsymbol{\mu}(t) &= \boldsymbol{\Lambda}^{-1}(t) \left(\boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0 + \sum_{i=1}^{|V|} \boldsymbol{e}_i \frac{s_i(t)}{\sigma^2} \right). \end{aligned} \tag{1}$$

Here, $\boldsymbol{e}_i$ is the standard unit vector with i-th entry to be 1.

C. Voronoi Partition and Coverage Problem

We define the N-partition of a graph G as a collection $P = \{P_i\}_{i=1}^N$ of N nonempty subsets of V such that $\cup_{i=1}^N P_i = V$ and $P_i \cap P_j = \emptyset$ for any $i \neq j$. P is said to be connected if the subgraph induced by P_i, denoted $G[P_i]$, is connected for each $i \in N$. $G[P_i]$ being an induced subgraph means its vertex set is P_i and its edge set includes all edges in G for which both end vertices are included in P_i.

The configuration of the robot team is a vector of N vertices $\boldsymbol{\eta} \in V^N$ occupied by the robot team, where the i-th entry η_i corresponds to position of the i-th robot. The i-th robot is tasked to cover vertices in P_i. The coverage cost corresponding to configuration $\boldsymbol{\eta}$ and connected N-partition P can be defined as

$$\mathcal{H}(\boldsymbol{\eta}, P) = \sum_{i=1}^{N} \sum_{v' \in P_i} d_{G[P_i]}(\eta_i, v') \phi(v'). \tag{2}$$

In a coverage problem, the objective is to minimize this coverage cost by selecting appropriate configuration $\boldsymbol{\eta}$ and connected N-partition P. However, how to efficiently find the optimal configuration-partition pair in a large graph with arbitrary sensory function ϕ remains an open problem. There are two intermediate results about the optimal selection of configuration or partition when the other is fixed [7].

1) Optimal Partition with Fixed Configuration: For a fixed configuration $\boldsymbol{\eta}$, an optimal connected N-partition P minimizing coverage cost is a Voronoi partition, denoted by $\mathcal{V}(\boldsymbol{\eta})$. Formally, for each $P_i \in \mathcal{V}(\boldsymbol{\eta})$ and any $v' \in P_i$, $\forall j \in \{1, \ldots, N\} : d_G(v', \eta_i) \leq d_G(v', \eta_j)$.

2) Optimal Configuration with Fixed Partition: For a fixed, connected N-partition P, the centroid of the i-th partition $P_i \in P$ is defined by

$$c_i := \operatorname*{arg\,min}_{v \in P_i} \sum_{v' \in P_i} d_{G[P_i]}(v, v') \phi(v'),$$

and the optimal configuration is to place a robot at the centroid of every $P_i \in P$. We denote the vector of centroids of P by $\boldsymbol{c}(P)$ with c_i as its i-th element.

Building upon the above two properties, the classic Lloyd algorithm [5] iteratively places each robot at the centroid of the current Voronoi partition and computes the new Voronoi partition using the updated configuration. Under this algorithm, it is known that the robot team will eventually converge to a centroidal Voronoi partition, defined below.

Definition 1 (Centroidal Voronoi Partition, [18]): An N-partition P is a centroidal Voronoi partition of G if P is a Voronoi partition generated by some configuration with distinct entries $\boldsymbol{\eta}$ such that $P = \mathcal{V}(\boldsymbol{\eta})$, and $\boldsymbol{c}\big(\mathcal{V}(\boldsymbol{\eta})\big) = \boldsymbol{\eta}$.

It should be noted that an optimal partition and configuration pair minimizing the coverage cost $\mathcal{H}(\boldsymbol{\eta}, P)$ is of the form $(\boldsymbol{\eta}^*, \mathcal{V}(\boldsymbol{\eta}^*))$, where $\boldsymbol{\eta}^*$ has distinct entries and $\mathcal{V}(\boldsymbol{\eta}^*)$ is a centroidal Voronoi partition. A configuration-partition pair $(\boldsymbol{\eta}', \mathcal{V}(\boldsymbol{\eta}'))$ is considered to be an efficient solution to the coverage problem if $\mathcal{V}(\boldsymbol{\eta}')$ is a centroidal Voronoi partition, even though it is possibly suboptimal [18].

D. Performance Evaluation

To evaluate the performance of online estimation and coverage, we introduce a notion of coverage regret.

Definition 2 (Coverage Regret): At each time t, let the team configuration be $\boldsymbol{\eta}_t$ and the connected N-partition be P_t. The coverage regret until time T is defined by

Algorithm 1: DSLC

Input : Environment graph G, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0$;
Set : $\alpha \in (0,1)$ and $\beta > 1$;

for *epoch* $j = 1, 2, \ldots$ **do**
  Exploration phase:
  The robot team sample at vertices in V to make
$$\max_{i \in \{1,\ldots,|V|\}} \sigma_i^2(t) \leq \alpha^j \sigma_0^2.$$
  Information propagation phase:
  Each robot agent propagates its sampling result to the team.
  Each robot updates estimated sensory function $\hat{\phi}$.
  Coverage phase:
  for $t_j = 1, 2, \ldots, \lceil \beta^j \rceil$ **do**
    Based on $\hat{\phi}$, follow pairwise partitioning rule to update robot team configuration and partition.

$\sum_{t=1}^{T} R_t(\phi)$, where $R_t(\phi)$ is the instantaneous coverage regret with respect to sensory function ϕ, and is defined by

$$R_t(\phi) = 2\mathcal{H}(\boldsymbol{\eta}_t, P_t) - \mathcal{H}(\boldsymbol{c}(P_t), P_t) - \mathcal{H}(\boldsymbol{\eta}_t, \mathcal{V}(\boldsymbol{\eta}_t)),$$

which is the sum of two terms $\mathcal{H}(\boldsymbol{\eta}_t, P_t) - \mathcal{H}(\boldsymbol{c}(P_t), P_t)$ and $\mathcal{H}(\boldsymbol{\eta}_t, P_t) - \mathcal{H}(\boldsymbol{\eta}_t, \mathcal{V}(\boldsymbol{\eta}_t))$. The former (resp., latter) term is the regret induced by the deviation of the current configuration (resp., partition) from the optimal configuration (resp., partition) for the current partition (resp., configuration). Accordingly, no regret is incurred at time t if and only if P_t is a centroidal Voronoi N-partition and $\boldsymbol{\eta}_t = \boldsymbol{c}(P_t)$. Thus, there are two sources contributing to coverage regret: (i) the estimation error in the sensory function ϕ, and (ii) the deviation from a centroidal Voronoi partition inevitable when agents sample the environment to learn ϕ.

III. Deterministic Sequencing of Learning and Coverage (DSLC) Algorithm

In this section, we describe the DSLC algorithm (Algorithm 1). It operates with an epoch-wise structure, where each epoch consists of an exploration (learning) phase and an exploitation (coverage) phase. The exploration phase comprises two sub-phases: estimation and information propagation.

A. Estimation Phase

Let $\sigma_i^2(t)$ be the marginal posterior variance of $\phi(v_i)$ at time t, i.e., the i-th diagonal entry of $\boldsymbol{\Lambda}^{-1}(t)$. Suppose the marginal prior variance $\sigma_i^2(0) \leq \sigma_0^2$, for each i. Within each epoch j, agents first determine the points to be sampled in order to reduce $\max_{i\in\{1,\ldots,|V|\}} \sigma_i^2(t)$ below a threshold $\alpha^j \sigma_0^2$, where $\alpha \in (0,1)$ is a prespecified parameter.

Note that the posterior covariance computed in (1) depends only the number of samples at each vertex, and does not require actual sampling results. Therefore, the sequence of sampling locations can be computed before physically visiting the locations. Leveraging this deterministic evolution of the covariance, we take a greedy sampling policy that repeatedly selects the point v_{i_t} with maximum marginal posterior variance, i.e.,

$$i_t = \operatorname*{arg\,max}_{i \in \{1,\ldots,|V|\}} \sigma_i(t), \tag{3}$$

for $t \in \{\underline{t}_j, \ldots, \bar{t}_j\}$, where $\underline{t}_j$ and $\bar{t}_j$ are the starting and ending time of estimation phase in the j-th epoch. It has been shown that the greedy sampling policy is near-optimal in terms of maximizing the mutual information of the sampling results and sensory function ϕ [19].

Let the set of points to be sampled during epoch j be X^j and let $X_r^j = X^j \cap P_{\underline{t}_j, r}$ be the set of sampling points that belong to $P_{\underline{t}_j, r}$, the partition assigned to agent r at time $\underline{t}_j$. Each agent r computes a path through the sampling points in X_r^j and collects noisy measurements from those points. The traveling path can be optimized by solving a Traveling Salesperson Problem (TSP).

B. Information Propagation Phase

After the estimation phase, sampling results from each agent must be passed to all other agents. There are several mechanisms to accomplish this in a finite number of steps. For example, agents can communicate with their neighboring agents and use flooding algorithms [20] to relay their sampling results to every agent. Alternatively, the agents may be able to send their sampling results to a cloud and receive global estimates after a finite delay. Another possibility for the agents is to use finite time consensus protocols [21] in the distributed inference algorithm discussed in [22].

For any of the above mechanisms, the sampling results from the entire robot team can be propagated to each robot agent in finite time. Then, each agent has an identical posterior distribution $\mathcal{N}\big(\boldsymbol{\mu}(t), \boldsymbol{\Lambda}^{-1}(t)\big)$ of $\boldsymbol{\phi}$, and $\hat{\phi} := \boldsymbol{\mu}(t)$ will be used as the estimate of the sensory function.

C. Coverage Phase

After the estimation and information propagation phases, agents have the same estimate of the sensory function $\hat{\phi}$. The coverage phase involves no environmental sampling and its length is designed to grow exponentially with epochs, i.e., the number of time steps in the coverage phase of the j-th epoch is $\lceil \beta^j \rceil$, for some $\beta > 1$. We use pairwise partitioning, a distributed gossip-based coverage algorithm proposed in [7], with the estimated sensory function $\hat{\phi}$.

In a connected N-partition P, P_i and P_j are said to be adjacent if there exists a vertex pair $v \in P_i$ and $v' \in P_j$ and an edge in E connecting v and v'. At each time, a random pair of agents (i, j), with P_i and P_j adjacent, compute an optimal pair of vertices (a^*, b^*) within $P_i \cup P_j$ that minimize

$$\sum_{v' \in P_i \cup P_j} \hat{\phi}(v') \min\Big(d_{G[P_i \cup P_j]}(a, v'), d_{G[P_i \cup P_j]}(b, v')\Big).$$

Then, agents i and j move to a^* and b^*. Subsequently, P_i and P_j are updated to

$$P_i \leftarrow \{v \in P_i \cup P_j \mid d_{G[P_i \cup P_j]}(\eta_i, v) \leq d_{G[P_i \cup P_j]}(\eta_j, v)\}$$
$$P_j \leftarrow \{v \in P_i \cup P_j \mid d_{G[P_i \cup P_j]}(\eta_i, v) > d_{G[P_i \cup P_j]}(\eta_j, v)\}.$$

IV. Analysis of DSLC Algorithm

In this section, we analyze DSLC to provide a performance guarantee on the expected cumulative coverage regret. To this end, we leverage the information gain from the estimation phase to analyze the convergence rate of uncertainty. Then, we recall convergence properties of the pairwise partitioning algorithm used in DSLC. Based on these results, we establish the main result of this paper: an upper bound on the expected cumulative coverage regret.

A. Mutual Information and Uncertainty Reduction

Let $X^g = (v_{i_1}, \ldots, v_{i_n})$ be a sequence of n vertices selected by the greedy policy and $Y_{X^g} = (y_1, \ldots, y_n)$ be observed sampling results corresponding to X^g. With a slight abuse of notation, we denote the marginal posterior variance of $\phi(v_i)$ after sampling at $v_{i_1} \ldots v_{i_k}$ by $\sigma_i^2(k)$. With greedy sampling policy, $i_k = \arg\max_{i \in \{1,\ldots,|V|\}} \sigma_i^2(k-1)$. Then, the mutual information of Y^g and $\boldsymbol{\phi}$ is

$$\begin{aligned} I\left(Y_{X^g}; \boldsymbol{\phi}\right) &= H\left(Y_{X^g}\right) - H\left(Y_{X^g} \mid \boldsymbol{\phi}\right) \\ &= \frac{1}{2}\sum_{k=1}^{n} \log\left(1 + \sigma^{-2}\sigma_{i_k}^2\left(k-1\right)\right), \end{aligned} \tag{4}$$

where $H(Y_{X^g})$ and $H(Y_{X^g} \mid \boldsymbol{\phi})$ denote the entropy and conditional entropy respectively. Let $\gamma_n := \max_{X \in V^n} I\left(Y_X; \boldsymbol{\phi}\right)$ be the maximal mutual information gain that can be achieved with n samples. It is shown in [23] that $I\left(Y_{X^g}; \boldsymbol{\phi}\right)$ achieved by the greedy sampling policy is near optimal, i.e.,

$$\left(1 - 1/e\right)\gamma_n \le I\left(Y_{X_g}; \boldsymbol{\phi}\right) \le \gamma_n. \tag{5}$$

We now present a bound on the maximal posterior variance after sampling at vertices within X^g. The following lemma and proof techniques are adapted from our previous work [24] to incorporate the discrete environment.

Lemma 1 (Uncertainty reduction): Under the greedy sampling policy, the maximum posterior variance after n sampling rounds satisfies

$$\max_{i \in \{1,\ldots,|V|\}} \sigma_i^2(n) \le \frac{2\sigma_0^2}{\log\left(1 + \sigma^{-2}\sigma_0^2\right)} \frac{\gamma_n}{n}.$$

Proof: For any $i \in \{1, \ldots, |V|\}$, $\sigma_i^2(k)$ is monotonically non-increasing in k. So, we get

$$\begin{aligned} \sigma_{i_{k+1}}^2(k) &\le \sigma_{i_{k+1}}^2(k-1) \\ &\le \max_{i \in \{1,\ldots,|V|\}} \sigma_i^2(k-1) = \sigma_{i_k}^2(k-1), \end{aligned} \tag{6}$$

which indicates that $\sigma_{i_{k+1}}^2(k)$ is monotonically non-increasing. Hence, from (4) and (5), $\log\Big(1 + \sigma^{-2}\sigma_{i_n}^2\left(n-1\right)\Big) \le 2\gamma_n/n$. Since $x^2/\log\left(1+x^2\right)$ is an increasing function on $[0, \infty)$,

$$\sigma_{i_n}^2\left(n-1\right) \le \frac{\sigma_0^2}{\log\left(1 + \sigma^{-2}\sigma_0^2\right)} \log\Big(1 + \sigma^{-2}\sigma_{i_n}^2\left(n-1\right)\Big).$$

Substituting (6) into the above equation, we conclude that

$$\sigma_{i_n}^2\left(n-1\right) \le \frac{2\sigma_0^2}{\log\left(1 + \sigma^{-2}\sigma_0^2\right)} \frac{\gamma_n}{n},$$

which establishes the lemma. ■

Typically, it is hard to characterize γ_n for a general Gaussian random vector $\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0^{-1})$. Therefore, we make the following assumption.

Assumption 1: Vertices in V lie in a convex and compact set $D \in \mathbb{R}^2$ and the covariance of any pair $\phi(v_i)$ and $\phi(v_j)$ is determined by an exponential kernel function

$$k(\phi(v_i), \phi(v_j)) = \sigma_v^2 \exp\left(-\frac{d_{\text{eu}}^2(v_i, v_j)}{2l^2}\right), \tag{7}$$

where $d_{\text{eu}}(v_i, v_j)$ is the Euclidean distance between v_i and v_j, l is the length scale, and σ_v^2 is the variability parameter.

We now recall an upper bound on γ_n from [16].

Lemma 2 (Information gain for squared exp. kernel): With Assumption 1, the maximum mutual information satisfies $\gamma_n \in O((\log(|V|\, n))^3)$.

B. Convergence within Coverage Phase

Since the sampling results of each agent are relayed to the entire team before each coverage phase, the team have a consensus estimate of the sensory function $\hat{\phi}$. It has been shown in [7] that using the pairwise partitioning algorithm, the N-partition P for the team converges almost surely to a class of near optimal partitions defined below.

Definition 3 (Pairwise-optimal Partition): A connected N-partition P is pairwise-optimal if for each pair of adjacent regions P_i and P_j,

$$\begin{aligned} &\sum_{v' \in P_i} d_G(c(P_i), v')\phi(v') + \sum_{v' \in P_j} d_G(c(P_j), v')\phi(v') \\ &= \min_{a,b \in P_i \cup P_j} \sum_{v' \in P_i \cup P_j} \phi(v') \min\left(d_G(a, v'), d(b, v')\right). \end{aligned}$$

This means that, within the induced subgraph generated by any pair of adjacent regions, the 2-partition is optimal. It is proved in [7] that if a connected N-partition P is pairwise-optimal then it is also a centroidal Voronoi partition. The following result on the convergence time of pairwise partitioning algorithm is established in [7].

Lemma 3 (Expected Convergence Time): Using the pairwise partitioning algorithm, the expected time to converge to a pairwise-optimal N-partition is finite.

Lemma 3 implies that the expected time for the $R_t(\hat{\phi})$ to converge to 0 in each coverage phase is finite.

C. An Upper Bound on Expected Coverage Regret

We now present the main result for this paper.

Theorem 4: For any time horizon T, if Assumption 1 holds and $\alpha = \beta^{-2/3}$, then the expected cumulative coverage regret for DSLC with respect to sensory function ϕ satisfies

$$\mathbb{E}\left[\sum_{t=1}^{T} R_t(\phi)\right] \in O\big(T^{2/3}(\log(T))^3\big).$$

Proof: We establish the theorem in four steps.

Step 1 (Regret from estimation phases): Let the total number of sampling steps before the end of the j-th epoch be s_j. By applying Lemma 1 and 2, we get $s_j \in O((\log(T))^3/\alpha^j)$. Thus, the coverage regret in the estimation phases until the end of the j-th epoch belongs to $O((\log(T))^3/\alpha^j)$.

Step 2 (Regret from information propagation phases): The sampling information by each robot propagate to all the team members in finite time. Thus, before the end of the j-th epoch, the coverage regret from information propagation phases can be bounded by $c_1 j$ for some constant $c_1 > 0$.

Step 3 (Regret from coverage phases): According to Lemma 3, in each coverage phase, the expected time before converging to a pairwise-optimal partition is finite. Thus, before the end of the j-th epoch, the expected coverage regret from converging steps can be upper bounded by $c_2 j$ for some constant $c_2 > 0$.

Note that the robot team converges to a pairwise optimal partition with respect to the estimated sensory function $\hat{\phi}$, which may differ from the actual ϕ. The instantaneous coverage regret $R_t(\phi)$ due to the estimation error is

$$2\mathcal{H}(\boldsymbol{\eta}_t, P_t) - \mathcal{H}(\boldsymbol{c}(P_t), P_t) - \mathcal{H}(\boldsymbol{\eta}_t \mathcal{V}(\boldsymbol{\eta}_t)) := A_t^{\mathrm{T}} \boldsymbol{\phi},$$

for some $A_t \in \mathbb{R}^{|V|}$. Moreover, the posterior distribution of $R_t(\phi)$ can be written as $\mathcal{N}(A_t^{\mathrm{T}} \boldsymbol{\mu}(t), A_t^{\mathrm{T}} \boldsymbol{\Sigma}(t) A_t)$, where $\boldsymbol{\Sigma}(t) = \boldsymbol{\Lambda}^{-1}(t)$ is the posterior covariance matrix. Since a pairwise-optimal partition P is also a centroidal Voronoi partition and $\hat{\phi} = \boldsymbol{\mu}(t)$, $R_t(\hat{\phi}) = 0$ indicates $A_t^{\mathrm{T}} \boldsymbol{\mu}(t) = 0$. Now, we get $R_t(\phi) \sim \mathcal{N}(0, A_t^{\mathrm{T}} \boldsymbol{\Sigma}(t) A_t)$ and

$$\mathbb{E}[R_t(\phi)] \leq \mathbb{E}\left[\left|R_t(\phi)\right|\right] = \sqrt{\frac{2}{\pi} A_t^{\mathrm{T}} \boldsymbol{\Sigma}(t) A_t}.$$

Note that $A_t^{\mathrm{T}} \boldsymbol{\Sigma}(t) A_t$ is a weighed summation of the eigenvalues of $\boldsymbol{\Sigma}(t)$. At any time t in the coverage phase of the k-th epoch, $\max_{i \in \{1,\ldots,|V|\}} \sigma_i^2(t) \leq \alpha^k \sigma_0^2$, and its follows that the summation of eigenvalues of $\boldsymbol{\Sigma}(t)$ equals trace$(\boldsymbol{\Sigma}(t)) \leq |V| \, \alpha^k \sigma_0^2$. Thus, we get

$$\mathbb{E}\left[\sum_{t \in \mathcal{T}_k^{\mathrm{cov}}} R_t(\phi)\right] \leq c_3 (\beta \sqrt{\alpha})^k,$$

for some constant $c_3 > 0$, where $\mathcal{T}_k^{\mathrm{cov}}$ are the time slots in the coverage phase of the k-th epoch excluding those associated with the transient of the pairwise partitioning algorithm, and where we have used the fact that $|\mathcal{T}_k^{\mathrm{cov}}| \leq \lceil \beta^k \rceil$.

Step 4 (Summary): Summing up the expected coverage regret from above steps, the expected cumulative coverage regret at the end of the j-th epoch T_j satisfies

$$\mathbb{E}\left[\sum_{t=1}^{T_j} R_t(\phi)\right] \leq C_1 j + C_2 s_j + \sum_{k=1}^{j} c_3 (\beta \sqrt{\alpha})^k, \quad (8)$$

where $C_1, C_2 > 0$ are some constants. The theorem statement follows by plugging in $\alpha = \beta^{-2/3}$, using $j \in O(\log T)$ and some simple calculations. ■

V. Simulation Results

Here, we present simulations which illustrate the empirical promise of DSLC. In particular, we highlight DSLC's ability to achieve sublinear regret, and compare the performance of DSLC to algorithms presented in [1] and [14].

Motivated by environmental applications, we construct the sensory function ϕ over a discrete 21×21 point grid in $[0, 1]^2$ by performing kernel density estimation on a subset of the geospatial distribution of Australian wildfires observed by NASA in 2019 [25]. Intuitively, ϕ represents the

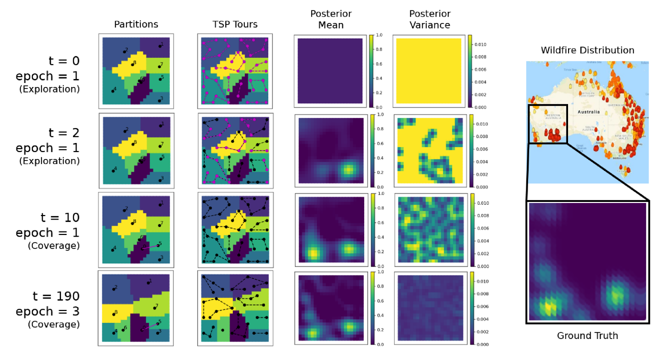

Fig. 1: Distributed implementation of DSLC in the unit square with $j = 3$ epochs of length 16, 46, and 128. From left to right: agent positions $\boldsymbol{\eta}_t$ and partitions P_t; TSP sampling tours; posterior mean and variance of $\hat{\phi}$; ground truth sensory function ϕ based on data from [25]. Pairwise partition updates between gossiping agents are denoted by magenta lines in the leftmost column of panels. Points along TSP tours in second-from-leftmost column of panels are plotted in magenta prior to sampling, and in black after sampling. Video is available online.[1]

probability that a wildfire may occur at a particular point of the unit square, and may be used to model the demand for an autonomous sensing agent at that point. In each simulation, nine agents are placed uniformly at random over the grid and execute three epochs of length 16, 46, and 128 to achieve adaptive coverage of the environment. Partitions are initialized by iterating over the grid and assigning each point to the nearest agent. During the exploration phase of each epoch, partitions are fixed; during the exploitation phase of each epoch, partitions are updated according to the protocol established in [7], where gossip-based repartitioning occurs between randomly selected neighbors. Coverage cost, regret and maximum variance are computed throughout using (2), Definition 2, and the maximum diagonal entry of $\boldsymbol{\Lambda}^{-1}(t)$ from (1), respectively.

The sensory function ϕ is normalized in the range $[0, 1]$ and sampled by agents with Gaussian noise parameterized by mean and standard deviation $\mu = 0$, $\sigma = 0.1$. A global Gaussian Process model is assumed to simplify estimation of $\hat{\phi}$ throughout the simulation, though in practice estimation of $\hat{\phi}$ could be implemented in a fully-distributed manner by assuming each agent maintains its own model of $\hat{\phi}$ and employing an information propagation phase described in Section III. Setting the parameter $\alpha = 0.5$ to reduce uncertainty by half within each epoch, $\beta = \alpha^{-3/2}$ is fully determined by Theorem 4. Fig. 1 visualizes the simulation of DSLC. A video of the simulation is available online.[1]

Fig. 2 compares the evolution of regret and cost incurred by DSLC with algorithms proposed in [1] and [14], denoted `Cortes` and `Todescato`, respectively. Agents in `Cortes` are assumed to have perfect knowledge of ϕ and simply go to the centroid of their cell at each iteration, while agents in `Todescato` follow a stochastic sampling approach with probability of exploration proportional to posterior variance in the estimate $\hat{\phi}$ at each iteration. As expected, `Cortes`

[1] `https://youtu.be/nalwrZC6GiI`

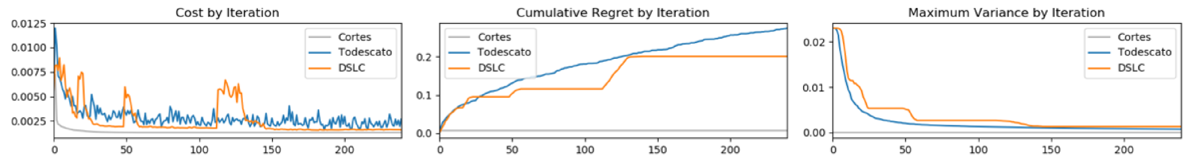

Fig. 2: Cost, regret and maximum posterior variance of $\hat{\phi}$ for DSLC, `Todescato` and `Cortes` averaged over 16 simulations of 190 iterations each. Note that DSLC empirically achieves sublinear regret. Spikes in regret occur during the exploration phase of each epoch, before agents converge to a pairwise-optimal coverage configuration with respect to $\hat{\phi}$ during the exploitation phase.

achieves minimal cost and coverage regret over the course of execution given complete access to ϕ, and serves as a baseline. On the other hand, DSLC achieves lower average cost per iteration and cumulative regret than `Todescato`, even after accounting for spikes in cost and regret incurred by DSLC during the exploration phase of each epoch. This suggests the deterministic, near-optimal greedy sampling policy of DSLC is both theoretically and empirically sound. All results are averaged over 16 simulations of 190 iterations, aligned with the implementation of DSLC comprising three epochs of lengths 16, 46, and 128.

VI. Conclusions

In this paper, we study the multi-robot coverage problem over an unknown nonuniform sensory field. DSLC, a novel adaptive coverage algorithm, is designed to drive agents to simultaneously learn the sensory function and provide satisfactory coverage. Defining a novel characterization of coverage regret, we analyze DSLC to get a sublinear upper bound on the cumulative expected coverage regret. We illustrate the empirical promise of DSLC through simulations in which a team of aerial robots is tasked with coverage of an unknown geospatial distribution of wildfires.

In future works, we hope to extend our approach to settings in which agents are assumed to have heterogeneous sensing and motion capabilities. We also see potential for extension to nonstationary settings in which a sensing field evolves with time. Such settings more accurately reflect the challenges presented by real-world implementation of multi-robot control algorithms, and offer promising avenues to broader impacts in multi-robot systems research.

References

[1] J. Cortés, S. Martínez, T. Karataş, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[2] J. Cortés and F. Bullo, "Coordination and Geometric Optimization via Distributed Dynamical Systems," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1543–1574, 2005.

[3] F. Lekien and N. E. Leonard, "Nonuniform coverage and cartograms," *Proceedings of the IEEE Conference on Decision and Control*, pp. 5518–5523, 2010.

[4] I. I. Hussein and D. M. Stipanovic, "Effective coverage control for mobile sensor networks with guaranteed collision avoidance," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 4, pp. 642–657, 2007.

[5] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[6] F. Bullo, R. Carli, and P. Frasca, "Gossip coverage control for robotic networks: Dynamical systems on the space of partitions," *SIAM Journal on Control and Optimization*, vol. 50, no. 1, pp. 419–447, 2012.

[7] J. W. Durham, R. Carli, P. Frasca, and F. Bullo, "Discrete partitioning and coverage control for gossiping robots," *IEEE Transactions on Robotics*, vol. 28, no. 2, pp. 364–378, 2012.

[8] M. Schwager, D. Rus, and J. J. Slotine, "Decentralized, adaptive coverage control for networked robots," *International Journal of Robotics Research*, vol. 28, no. 3, pp. 357–375, 2009.

[9] M. Schwager, M. P. Vitus, S. Powers, D. Rus, and C. J. Tomlin, "Robust adaptive coverage control for robotic sensor networks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 462–476, 2017.

[10] J. Choi, J. Lee, and S. Oh, "Swarm intelligence for achieving the global maximum using spatio-temporal Gaussian processes," *Proceedings of the American Control Conference*, pp. 135–140, 2008.

[11] Y. Xu and J. Choi, "Adaptive sampling for learning Gaussian processes using mobile sensor networks," *Sensors*, vol. 11, no. 3, pp. 3051–3066, 2011.

[12] W. Luo and K. Sycara, "Adaptive Sampling and Online Learning in Multi-Robot Sensor Coverage with Mixture of Gaussian Processes," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 6359–6364, 2018.

[13] W. Luo, C. Nam, G. Kantor, and K. Sycara, "Distributed environmental modeling and adaptive sampling for multi-robot sensor coverage," in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, 2019, pp. 1488–1496.

[14] M. Todescato, A. Carron, R. Carli, G. Pillonetto, and L. Schenato, "Multi-robots Gaussian estimation and coverage control: From client–server to peer-to-peer architectures," *Automatica*, vol. 80, pp. 284–294, 2017.

[15] A. Benevento, M. Santos, G. Notarstefano, K. Paynabar, M. Bloch, and M. Egerstedt, "Multi-robot coordination for estimation and coverage of unknown spatial fields," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 7740–7746.

[16] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[17] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I : Estimation Theory*. Prentice Hall, 1993.

[18] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*, ser. Applied Mathematics Series. Princeton University Press, 2009, electronically available at http://coordinationbook.info.

[19] A. Krause and C. E. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Conf. on Uncertainty in Artificial Intelligence*, Edinburgh, Scotland, Jul. 2005, pp. 324–331.

[20] H. Lim and C. Kim, "Flooding in wireless ad hoc networks," *Computer Communications*, vol. 24, no. 3-4, pp. 353–363, 2001.

[21] L. Wang and F. Xiao, "Finite-time consensus problems for networks of dynamic agents," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 950–955, 2010.

[22] P. Landgren, V. Srivastava, and N. E. Leonard, "Distributed cooperative decision-making in multiarmed bandits: Frequentist and Bayesian algorithms," in *IEEE Conf. on Decision and Control*, Las Vegas, NV, Dec. 2016, pp. 167–172.

[23] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical programming*, vol. 14, no. 1, pp. 265–294, 1978.

[24] L. Wei, X. Tan, and V. Srivastava, "Expedited multi-target search with guaranteed performance via multi-fidelity Gaussian processes," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV (Virtual), Oct. 2020, pp. 7095–7100.

[25] C. Paradis, "Fires from Space: Australia," 2019. [Online]. Available: https://www.kaggle.com/carlosparadis/fires-from-space-australia-and-new-zeland