

Approximation Algorithms for Robot Tours in Random Fields with Guaranteed Estimation Accuracy

Shamak Dutta, Nils Wilde, Pratap Tokekar, and Stephen L. Smith

Abstract—We study the sample placement and shortest tour problem for robots tasked with mapping environmental phenomena modeled as stationary random fields. The objective is to minimize the resources used (samples or tour length) while guaranteeing estimation accuracy. We give approximation algorithms for both problems in convex environments. These improve previously known results, both in terms of theoretical guarantees and in simulations. In addition, we disprove an existing claim in the literature on a lower bound for a solution to the sample placement problem.

I. INTRODUCTION

In environmental monitoring, robots are tasked with mapping physical phenomena with desired accuracy while minimizing the resources used. For example, farmers want accurate nutrient maps of their fields to direct fertilizer usage which will subsequently maximize their crop yield. Unfortunately, soil sampling is time-consuming and expensive [1]. It is advantageous to minimize the number of samples taken to build a nutrient map. In post-disaster environments, there is a need to map the spatial distribution of radioactivity using unmanned aerial systems in a short period of time [2]. This involves a combination of deciding the measurement locations as well as planning a short path through the environment with the goal of building an accurate map.

We consider the setting where the phenomena is modeled as a stationary random field over a compact two-dimensional environment. The covariance structure of the stationary field is given by the commonly used squared-exponential function. This probabilistic model is popular in spatial statistics [1], [3] since it allows for predictions at unobserved locations along with an estimate on the prediction error. We focus on two problems: sample placement and shortest tours in random fields. Both problems deal with the same constraint: ensure the estimation error at each point in the environment is within a desired tolerance. In sample placement, we seek to find a subset of measurement locations of minimum size that satisfies this constraint. In a similar vein, the shortest tour problem aims to compute the minimum length tour whose vertices satisfy the constraint. Unfortunately, sampling and tour planning in random fields are NP-hard [4], [5] and thus, our objective is to develop approximation algorithms.

This research is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by Nutrien Ltd.

S. Dutta and S. L. Smith are with the Department of Electrical and Computer Engineering, University of Waterloo, Canada {stephen.smith, shamak.dutta}@uwaterloo.ca. N. Wilde is with the Cognitive Robotics Department, Delft University of Technology, Netherlands (N.Wilde@tudelft.nl). P. Tokekar is with the Department of Computer Science, University of Maryland (tokekar@umd.edu).



Fig. 1. An example of a solution to the sample placement and shortest tour problem. The vertex set (white circles) of the tour (black) guarantees accurate predictions in the hexagons (yellow), thereby covering the field.

Our approach, HEXCOVER, to solving the sample placement problem involves tiling the environment with hexagons whose centers serve as measurement locations. The edge length of the hexagons are selected in a way such that the resulting measurement set satisfies the error tolerance. For the shortest tour problem, our proposed method, HEXCOVER-TOUR, uses the measurement set obtained by HEXCOVER to plan an approximately optimal tour. An example of a sample placement and tour are shown in Figure 1.

Related Work: In geostatistics, estimation of random fields is done via the theory of kriging interpolation [3], [6]. With knowledge of the mean and covariance, estimation at unobserved locations is carried out using linear least-squares estimation. The expected error depends on the locations of the measurements and the statistics of the data but not on the outcomes itself. As a result, objectives based on the error can be computed a priori. This is equivalent to Gaussian Process regression in machine learning [7]. We use kriging to make predictions using linear estimators and quantify the expected mean squared error at unobserved locations.

The sample placement and shortest tour problem have been considered previously [8]. The analysis of the approximation algorithms relies on a claim of a lower bound on any solution to the sample placement problem. We provide a counterexample to disprove this claim. However, by using the effective range of the covariance function [1], [9], [10], the provided approximation ratios still hold.

In sensor/sample placement, the objective is to distribute a set of sensors to maximize the sensing quality over a random field. Sensing quality is measured by the entropy and mutual information. Though computing the optimal subset

in this case is NP-hard, greedy algorithms perform well [11] with guarantees stemming from submodularity of the objective [9]. However, guarantees for mutual information do not hold for the estimation error which is the constraint in our formulation. While algorithms have been developed for sensor coverage [12] in deterministic environments, the methods are not applicable in random fields.

In informative path planning, robots aim to maximize the information gained about the environment subject to constraints on path length. One approach to tackle continuous environments is by discretization where optimal solutions have been computed via branch-and-bound [13] and mixed-integer programming [14]. When the objective is submodular, recursive greedy approaches [15], [16], [17] provide near-optimal solutions. However, the estimation error is not a supermodular set function [4], [18]. A related kriging variance minimization problem is considered in [19] but the approach does not provide approximation guarantees. Methods that operate in continuous space have also been developed [20], [21] with guarantees on classification error [22]. Certain scenarios require adaptive approaches for informative planning to adjust for outliers [23] or to the non-uniform spatial correlations [24], [25] or account for robot pose uncertainty [26]. Unfortunately, none of these algorithms provide guarantees for the estimation error constraint considered in this paper.

Contributions: The contributions of this work are twofold. First, we disprove a claim in the literature [8] on a necessary condition for the sample placement problem. Second, we give approximation algorithms for the sample placement and shortest tour problems in convex environments, improving previous theoretical guarantees. In simulations, we demonstrate the effectiveness of our approach over prior work.

II. PRELIMINARIES

Consider a metric space (\mathcal{X}, ρ) where $\rho : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ is a metric. Let $A \subset \mathcal{X}$ be a compact set and let $B(x, r) := \{y \in \mathcal{X} : \rho(x, y) \leq r\}$ denote a closed ball.

A. Covering & Packing Number

Definition 1 (*r*-packing). The set $\{x_1, \dots, x_n\} \subset A$ is an *r*-packing of *A* if the sets $\{B(x_i, r/2) : i \in [n]\}$ are pairwise disjoint i.e., for any $i \neq j$, $\rho(x_i, x_j) > r$.

Definition 2 (*r*-covering). The set $\{x_1, \dots, x_n\} \subset A$ is an *r*-covering of *A* if $A \subset \bigcup_{i=1}^n B(x_i, r)$ i.e., $\forall x \in A, \exists i$ such that $\rho(x_i, x) \leq r$.

B. The Metric Traveling Salesman Problem

The input to the metric traveling salesman problem (TSP) is a complete graph $G = (V, E)$, where the vertex set *V* lies in a metric space (\mathcal{X}, ρ) . A cycle is a sequence of distinct vertices $\langle v_1, \dots, v_n, v_1 \rangle$ and a tour *T* is a cycle that visits each vertex in the graph exactly once. The cost of a tour, denoted by $\text{len}(T)$, is the sum of the costs of its associated edge set i.e., $\text{len}(T) := \sum_{i=1}^{n-1} \rho(v_i, v_{i+1}) + \rho(v_n, v_1)$. The objective of the TSP is then to find a tour of minimum cost. With slight abuse of notation, we let *T* refer to the tour as well the set of vertices visited on the tour.

III. PROBLEM FORMULATION

We follow a similar problem setup to [18], [14], [8]. We provide a summarized description here. Consider a zero-mean Gaussian random field $\{Z(x) : x \in D\}$ indexed on a convex and compact environment $D \subset \mathbb{R}^2$ with a non-empty interior. Let $\rho : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ be the Euclidean metric and denote the squared exponential covariance function associated with the field by $\phi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ i.e., for any $x, y \in D$,

$$\text{Cov}(Z(x), Z(y)) = \phi(\rho(x, y)) = \sigma_0^2 e^{-\frac{\rho^2(x, y)}{2L^2}}, \quad (1)$$

where $\sigma_0 > 0, L > 0$ are assumed to be known hyperparameters, typically determined by a pilot study.

It is common to assume that points that are sufficiently distant are uncorrelated since the covariance decays exponentially with the distance [9], [10]. In spatial statistics, this is known as the *effective range* of the covariance function [1] and is usually taken to be the distance at which $\phi(\cdot)$ equals 5% of the a priori field variance σ_0^2 .

Assumption 1 (Effective Range, [1]). The distance r_{\max} beyond which any two points are assumed to be uncorrelated is given by

$$\phi(r_{\max}) = 0.05 \times \sigma_0^2 \implies r_{\max} = \sqrt{6}L. \quad (2)$$

The measurements of the random field are corrupted with zero-mean Gaussian noise with variance $\sigma^2 > 0$. Given a measurement set $S = \{x_1, \dots, x_n\} \subset D$, the mean squared estimation error resulting from the optimal linear estimator of $Z(x)$, which is equal to the predictive variance in Gaussian Process regression [3], [27], is given by:

$$f_x(S) := \phi(0) - \mathbf{b}'_{x,S} C_S^{-1} \mathbf{b}_{x,S}, \quad (3)$$

where the following notation is used:

$$\begin{aligned} \mathbf{b}_{x,S} &:= [\phi(\rho(x, x_1)), \dots, \phi(\rho(x, x_n))] \in \mathbb{R}^n \\ C_S &:= \mathbb{E} \left[\mathbf{Z}_S \mathbf{Z}_S^T \right] + \sigma^2 I_n \in \mathbb{R}^{n \times n} \\ &= \begin{bmatrix} \phi(0) & \dots & \phi(\rho(x, x_1)) \\ \vdots & \ddots & \vdots \\ \phi(\rho(x_n, x_1)) & \dots & \phi(0) \end{bmatrix} + \sigma^2 I_n \end{aligned} \quad (4)$$

We are ready to state the two problems we wish to solve.

Problem 1 (Sample Placement). Given an environment $D \subset \mathbb{R}^2$ and $\Delta < \sigma_0^2$, find a measurement set $S \subset D$ of minimum cardinality such that the resulting estimation error for each $x \in D$ is within Δ i.e.,

$$\begin{aligned} &\underset{S \subset D}{\text{minimize}} && |S| \\ &\text{subject to} && f_x(S) \leq \Delta, \forall x \in D. \end{aligned} \quad (5)$$

Problem 2 (Shortest Tour). Given an environment $D \subset \mathbb{R}^2$ and $\Delta < \sigma_0^2$, find a tour *T* of minimum length such that the vertex set of the tour achieves an estimation error within Δ for each point $x \in D$ i.e.,

$$\begin{aligned} &\underset{T}{\text{minimize}} && \text{len}(T) \\ &\text{subject to} && f_x(T) \leq \Delta, \forall x \in D. \end{aligned} \quad (6)$$

Throughout this paper, we will denote the optimal solutions to the sample placement and shortest tour problem by S^* and T^* respectively.

IV. SOLUTION APPROACH

We begin with a discussion of a previous approach [8] to solving the sample placement problem. In Section IV-A, we disprove a claimed lower bound on the optimal solution. In light of this, we also observed the proposed algorithm in [8] is sample inefficient. This motivates the development of new algorithms: HEXCOVER for the sample placement problem (Section IV-B) and HEXCOVERTOUR for the shortest tour problem (Section IV-C).

A. Disproving a Necessary Condition

The approach for sample placement proposed in [8] is called DISKCOVER and uses two steps. First, it covers the environment with large circles of a certain radius r_{\max} . Second, it then covers these circles with smaller circles to obtain a feasible solution which yields an approximation algorithm. The analysis does not use Assumption 1. Instead, it involves a claim on a necessary condition for the problem. We provide a counterexample to disprove this claim.

It would be useful to obtain a distance r_{\max} such that if a prediction location does not have a sample within r_{\max} , then the constraint in (5) cannot be satisfied. Then, a packing of circles of radius r_{\max} provides a lower bound for the sample placement problem. This is the idea behind the necessary condition in [8]. We state the lemma below.

Lemma 1 (Lemma 1, [8]). For any test location x , if the nearest measurement location is at a distance r_{\max} away, and

$$r_{\max} > L \sqrt{-\log \left(1 - \frac{\Delta}{\sigma_0^2} \right)}, \quad (7)$$

then it is not possible to bring down the MSE (estimation error) below Δ at x .

The proof constructs a bound for the estimation error by placing all samples at the nearest location to the test location x . This is shown in the left plot in Figure 2. While this seems intuitive, this does not give a valid lower bound on the error. We construct a measurement set, as shown in the right plot in Figure 2, that obtains a lower estimation error.

Example 1. Consider the following environment setup: $\sigma = 1$, $\sigma_0 = 1$, $L = 1$, $x = (0, 0)$, $\Delta = 0.5$. Then, the RHS of (7) is 0.83255461. Let $r_{\max} = 0.93255461$ and consider the following measurement set:

$$S := \{(r_{\max}, 0), (-r_{\max}, 0), (0, r_{\max}), (0, -r_{\max})\}. \quad (8)$$

The estimation error at x using measurement set S is

$$f_x(S) = 0.443771 < \Delta, \quad (9)$$

which shows the contradiction. •

The counterexample uses four points to invalidate the lower bound on r_{\max} . However, by placing a large number

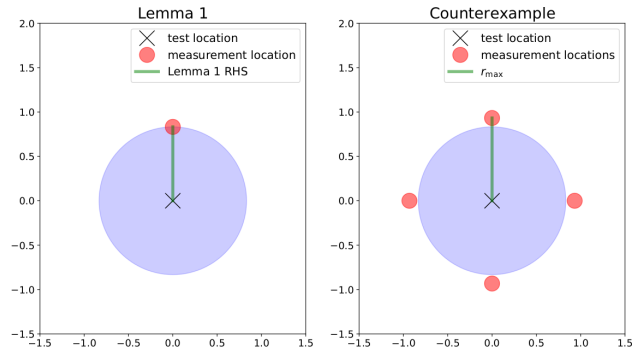


Fig. 2. Visual depiction of the counterexample. Left: Lemma 1 in [8], constructs a lower bound by placing samples at the red circle. Right: The set of four measurement locations obtains a lower estimation error.

of samples outside r_{\max} , the error can be much lower, thereby underestimating the lower bound on r_{\max} by a large margin. By using Assumption 1, which is a common practical assumption in geostatistics [1], the approximation ratios in [8] still hold (though Lemmas 1 and 3 in [8] do not hold). Since the proposed algorithm [8] covers the environments with circles of radius r_{\max} (which can be large), and covers it with smaller circles, it uses an excessive number of samples. This motivates us to develop an alternative algorithm.

B. HEXCOVER

Our approach to solving the sample placement problem is simple: cover the environment with circles of a specific radius r_{\min} . Then, one solution is the set of circle centers. We shall see that an appropriate selection of r_{\min} guarantees feasibility of this approach.

Lemma 2 (Sufficient Condition). Let $S \subset D$ be a r_{\min} -covering of the environment D , where

$$r_{\min} := L \sqrt{-\log \left((\sigma_0^2 - \Delta) \frac{\sigma_0^2 + \sigma^2}{\sigma_0^4} \right)}, \quad (10)$$

then S is feasible for the sample placement problem.

Proof. We will show that for any point $x \in D$ in the environment, the estimation error $f_x(S) \leq \Delta$. Since S is a r_{\min} -covering, there exists a point $x_i \in S$ such that $\rho(x, x_i) \leq r_{\min}$. In addition, since $f_x(S)$ is a monotonically decreasing set function [28], it suffices to show $f_x(\{x_i\}) \leq \Delta$. Then,

$$\begin{aligned} \rho(x, x_i) \leq r_{\min} &\implies e^{-\frac{\rho(x, x_i)^2}{L^2}} \geq (\sigma_0^2 - \Delta) \frac{\sigma_0^2 + \sigma^2}{\sigma_0^4} \\ &\implies f_x(\{x_i\}) \leq \Delta. \end{aligned}$$

□

The last step is to generate the r_{\min} -covering. Since the objective is to minimize the number of samples, one idea is to compute the minimum r_{\min} -covering. Unfortunately, this is difficult to compute in general. Instead, we compute a cover from a hexagonal tiling, which is the densest way to pack circles in the Euclidean plane [29], [30]. The steps are described in Algorithm 1. The main step is HEXAGONAL TILING (Line

2), which returns the set of centers in the hexagonal tiling of edge length r_{\min} . The centers are arranged in staggered columns where the vertical distance between two samples is $\sqrt{3}r_{\min}$ and the horizontal spacing between the columns is $1.5r_{\min}$. An example of a hexagonal tiling is shown in Figure 1. By construction, the circles centered at these points will circumscribe the hexagon, thereby covering the environment. Note that HEXCOVER runs in polynomial time since it scales linearly with the area of the environment i.e., the time complexity is $O(\text{area}(D))$.

Algorithm 1: HEXCOVER

Input: Environment $D \subset \mathbb{R}^2$, Error Threshold: Δ ,
Hyperparameters: L, σ_0^2 , Noise Variance σ^2

Output: Measurement Set: $S \subset D$

- 1 Compute r_{\min} according to Equation (10).
 - 2 $S_{\text{HC}} = \text{HEXAGONALTILING}(D, r_{\min})$
 - 3 **return** S_{HC}
-

Assumption 2 (Boundary Conditions). For some environments, there exist regions close to the boundary that are not covered by the hexagonal tiling. This is solved by taking a few more measurements to ensure feasibility. However, for the purposes of algorithm analysis, we will assume that the hexagonal tiling (Line 2) covers the environment completely.

To analyse the performance of this approach, the first step is to compute an upper bound on the solution $|S_{\text{HC}}|$ returned by HEXCOVER. To do this, we require the definition of a Minkowski sum and a lemma bounding the maximum area of the Minkowski sum of a convex set and a unit circle. Our proof of this result is based on the lecture notes [31] and is included here for completeness.

Let $B := \{(x, y) : x^2 + y^2 \leq 1\}$ be the unit circle and for any convex set Θ and $c > 0$, define $c\Theta := \{cx : x \in \Theta\}$.

Definition 3 (Minkowski Sum). For any two sets A, B , the Minkowski sum is $A + B := \{x + y : x \in A, y \in B\}$.

Lemma 3. Given a convex set $\Theta \subset \mathbb{R}^2$, the unit circle B , and $r > 0$ with $rB \subset \Theta$, then

$$\text{area}\left(\Theta + \frac{r}{2}B\right) \leq \frac{9}{4}\text{area}(\Theta).$$

Proof. Take any $a \in \Theta + \frac{r}{2}B$. Then, it can be expressed as $a = x + y$, where $x \in \Theta, y \in \frac{r}{2}B$. Since $rB \subset \Theta$, then $\frac{r}{2}B \subset \frac{1}{2}\Theta$. Thus, $a \in \Theta + \frac{1}{2}\Theta$. Now, a can be expressed as $a = s + \frac{t}{2}$, where $s, t \in \Theta$. Then, $\frac{2}{3}a = \frac{2}{3}s + \frac{t}{3}$. Since Θ is convex, we have $\frac{2}{3}a \in \Theta$ which implies $a \in \frac{3}{2}\Theta$. Thus, $\Theta + \frac{r}{2}B \subset \frac{3}{2}\Theta$ from which the final result follows. \square

Based on Lemma 3 we now establish an upper bound on the number of measurements $|S_{\text{HC}}|$.

Lemma 4 (Upper Bound). Let $\delta > 0$ be arbitrarily small and let $(\sqrt{3} - \delta)r_{\min}B \subset D$, where B is the unit circle. Then for some arbitrarily small $\epsilon_\delta > 0$,

$$|S_{\text{HC}}| \leq (3 + \epsilon_\delta) \frac{\text{area}(D)}{\pi r_{\min}^2}.$$

Proof. Denote the solution $S_{\text{HC}} := \{x_1, \dots, x_k\}$. Then, under Assumption 2, S_{HC} is a $(\sqrt{3} - \delta)r_{\min}$ -packing for the environment D . This is because the minimum distance between any two points in the hexagonal cover is $\sqrt{3}r_{\min}$ i.e., $\min_{i \neq j} \rho(x_i, x_j) = \sqrt{3}r_{\min} > (\sqrt{3} - \delta)r_{\min}$, which ensures it is a $(\sqrt{3} - \delta)r_{\min}$ -packing. Since the circles in the packing are disjoint, we have $\bigcup_i^k B(x_i, \frac{(\sqrt{3}-\delta)}{2}r_{\min}) \subset D + \frac{(\sqrt{3}-\delta)}{2}r_{\min}B$. Then, letting $c = \sqrt{3} - \delta$ and taking the area on both sides gives

$$\begin{aligned} \text{area}\left(\bigcup_i^{|S_{\text{HC}}|} B(x_i, cr_{\min}/2)\right) &= |S_{\text{HC}}| \frac{\pi c^2 r_{\min}^2}{4} \\ &\leq \text{area}\left(D + \frac{cr_{\min}}{2}B\right) \leq \frac{9}{4}\text{area}(D) \\ &\implies |S_{\text{HC}}| \leq \left(\frac{3}{c}\right)^2 \frac{\text{area}(D)}{\pi r_{\min}^2} \\ &= (3 + \epsilon_\delta) \frac{\text{area}(D)}{\pi r_{\min}^2}, \end{aligned} \tag{11}$$

where the second inequality follows from Lemma 3 and

$$\epsilon_\delta = \left(\frac{3}{\sqrt{3} - \delta}\right)^2 - 3$$

is arbitrarily small. \square

To obtain an approximation factor for Algorithm 1, we characterize an optimal solution S^* . First, we establish a property on any feasible solution which we will then use to get a lower bound on $|S^*|$.

Lemma 5 (Feasible Sample Placement). Any feasible solution S is a r_{\max} -covering of the environment D .

Proof. Suppose not. Then, there exists $x \in D$ such that for all $y \in S$, $\rho(x, y) > r_{\max}$. Using Assumption 1, the estimation error at x using S is $f_x(S) = \sigma_0^2 > \Delta$, contradicting feasibility of S for the sample placement problem. \square

Lemma 6 (Lower Bound). A lower bound on the optimal value $|S^*|$ for the sample placement problem is

$$|S^*| \geq \frac{\text{area}(D)}{\pi r_{\max}^2}.$$

Proof. Using Lemma 5, S^* is a r_{\max} -covering of the environment D i.e., $\bigcup_{x \in S^*} B(x, r_{\max}) \supset D$. Taking area on both sides gives

$$\begin{aligned} \text{area}(D) &\leq \text{area}\left(\bigcup_{x \in S^*} B(x, r_{\max})\right) \\ &\leq \sum_{x \in S^*} \text{area}(B(x, r_{\max})) = |S^*| \pi r_{\max}^2, \end{aligned}$$

where the second inequality follows from the fact that area is a sub-additive function. \square

Theorem 1. For an arbitrarily small $\epsilon > 0$, HEXCOVER is an α -approximation algorithm for the sample placement problem where

$$\alpha := 3 \frac{r_{\max}^2}{r_{\min}^2} + \epsilon.$$

Proof. Using Lemmas 4 and 6, for some arbitrarily small $\epsilon_\delta > 0$, $|S_{\text{HC}}| \leq (3 + \epsilon_\delta) \frac{\text{area}(D)}{\pi r_{\min}^2} \leq (3 + \epsilon_\delta) \frac{r_{\max}^2}{r_{\min}^2} |S^*| = (3 \frac{r_{\max}^2}{r_{\min}^2} + \epsilon) |S^*|$, where $\epsilon = \epsilon_\delta \frac{r_{\max}^2}{r_{\min}^2}$ is arbitrarily small. \square

Theorem 1 improves the previous approximation ratio of $18 \frac{r_{\max}^2}{r_{\min}^2}$ [8]. Our approximation factor is independent of the environment D but does depend on the error threshold Δ through r_{\min} . As we seek more accurate predictions (decreasing Δ), the radius of accurate estimation r_{\min} reduces, and the number of measurements required will increase.

C. HEXCOVERTOUR

Our approach to solving the shortest tour problem builds on HEXCOVER. This is because the constraints are the same in both problems. Since we have already identified a feasible vertex set via HEXCOVER, our method uses Christofides' algorithm which finds a tour of length no more than $3/2$ times the optimal [32]. The steps are outlined in Algorithm 2. The time complexity is dominated by Christofides' algorithm (Line 2) which runs in $O(n^3)$ time, where n is the number of vertices in the graph. Since the number of vertices is $O(\text{area}(D))$, HEXCOVERTOUR runs in $O((\text{area}(D))^3)$ time.

Algorithm 2: HEXCOVERTOUR

Input: Environment $D \subset \mathbb{R}^2$, Error Threshold: Δ ,
Hyperparameters: L, σ_0^2 , Noise Variance σ^2

Output: Tour T with $T \subset D$.

- 1 $S_{\text{HC}} = \text{HEXCOVER}(D, \Delta, L, \sigma_0^2, \sigma^2)$
 - 2 $T_{\text{HC}} = \text{CHRISTOFIDESTOUR}(S_{\text{HC}})$
 - 3 **return** T_{HC}
-

We begin by characterizing an upper bound on the length of the tour T_{HC} produced by HEXCOVERTOUR.

Lemma 7 (Upper Bound). For an arbitrarily small $\epsilon > 0$,

$$\text{len}(T_{\text{HC}}) \leq 15.6 \frac{\text{area}(D)}{\pi r_{\min}} + \epsilon.$$

Proof. Denote the vertex set of the tour by $V_{\text{HC}} := \{x_1, \dots, x_k\}$ and let T_k denote the optimal TSP tour on the k vertices of the tour T_{HC} . Note that in the hexagonal covering, the minimum distance between two points is $\sqrt{3}r_{\min}$. Now, we can bound T_k by constructing a sub-optimal tour as follows. Take any two points $x_i, x_j \in V_{\text{HC}}$ such that $\rho(x_i, x_j) \leq \sqrt{3}r_{\min}$ and find the shortest tour in $V_{\text{HC}} \setminus x_i$. Then, we can create a tour by adding the edge to connect x_j to x_i and back. This gives an upper bound:

$$\text{len}(T_k) \leq \text{len}(T_{k-1}) + 2\sqrt{3}r_{\min}.$$

Using the fact that $\text{len}(T_1) = 0$ and Lemma 4, we get

$$\text{len}(T_k) \leq 2\sqrt{3}r_{\min}|V_{\text{HC}}| \leq 2\sqrt{3}(3 + \epsilon_\delta) \frac{\text{area}(D)}{\pi r_{\min}}.$$

Since we are using Christofides' algorithm, we have

$$\begin{aligned} \text{len}(T_{\text{HC}}) &\leq 1.5\text{len}(T_k) \leq 3\sqrt{3}(3 + \epsilon_\delta) \frac{\text{area}(D)}{\pi r_{\min}} \\ &= 15.6 \frac{\text{area}(D)}{\pi r_{\min}} + \epsilon, \end{aligned}$$

where $\epsilon = 3\sqrt{3} \frac{\text{area}(D)}{\pi r_{\min}} \epsilon_\delta$ is an arbitrarily small number. \square

In the following lemma, we compute a lower bound on the length of the optimal tour T^* . First, we need a property of a maximal packing of a r_{\max} -covering.

Lemma 8. Let $S := \{x_1, \dots, x_n\}$ be a r_{\max} -covering of the environment D . Then, any maximal $2r_{\max}$ -packing $P \subset S$ of S is a $3r_{\max}$ -covering of D .

Proof. Using the triangle inequality and the fact that S is a r_{\max} -covering of D , it suffices to show that for any $x_i \in S$, there exists a point in the packing $y \in P$ such that $\rho(x_i, y) \leq 2r_{\max}$. Suppose not. Then, there exists $x_i \in S$ such that for all $y \in P$, $\rho(x_i, y) > 2r_{\max}$. But, we could add x_i to P and increase its size, contradicting the fact it is maximal. \square

Lemma 9 (Lower Bound). A lower bound on the length of the optimal tour is

$$\text{len}(T^*) \geq \frac{2}{9} \frac{\text{area}(D)}{\pi r_{\max}}.$$

Proof. We compute a maximal disjoint set $T' \subset T^*$ from the set of circles of radius r_{\max} centered at points in the optimal tour T^* . By the triangle inequality, the optimal TSP tour through T' gives us a bound on the length of the optimal tour:

$$\text{len}(T^*) \geq \text{len}(T'). \quad (12)$$

Since T' is a $2r_{\max}$ -packing, the minimum distance between any two vertices in T' is $2r_{\max}$. Thus,

$$\text{len}(T') \geq 2r_{\max}|T'|. \quad (13)$$

Since T^* is a r_{\max} -covering of D (Lemma 5), T' is a $3r_{\max}$ -covering (Lemma 8). Then, $\bigcup_{x \in T'} B(x, 3r_{\max}) \supset D$ which implies

$$\begin{aligned} \text{area} \left(\bigcup_{x \in T'} B(x, 3r_{\max}) \right) &\geq \text{area}(D) \\ \implies \sum_{x \in T'} \text{area}(B(x, 3r_{\max})) &\geq \text{area}(D) \\ \implies |T'| &\geq \frac{\text{area}(D)}{9\pi r_{\max}^2}. \end{aligned} \quad (14)$$

Combining Equations (12), (13), and (14) gives us the result. \square

The guarantee is a consequence of Lemmas 7 and 9.

Theorem 2. HEXCOVERTOUR is an α -approximation algorithm for the shortest tour problem where for an arbitrarily small $\epsilon > 0$,

$$\alpha := 70.2 \frac{r_{\max}}{r_{\min}} + \epsilon.$$

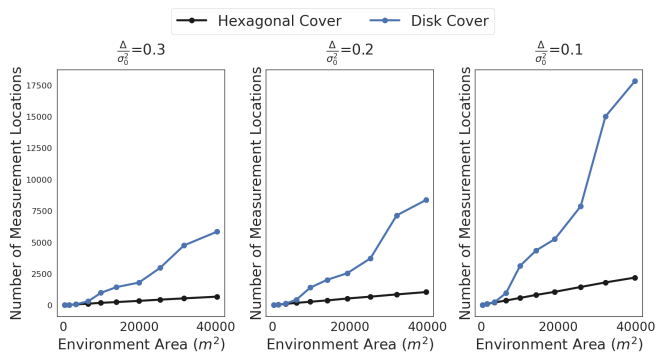


Fig. 3. Comparison of the number of measurement locations for Problem 1 versus the environment size. DISKCOVER uses roughly 6 times as many measurements as HEXCOVER.

This improves the previous approximation ratio of $9.33 + O\left(\frac{r_{\max}^2}{r_{\min}}\right)$ [8]. Note that if we use good heuristic solvers, such as the Lin-Kernighan heuristic [32], we can reduce our factor by (roughly) 1.5 leading to a $(46.8 \frac{r_{\max}}{r_{\min}} + \epsilon)$ -approximation.

V. SIMULATIONS

In this section, we demonstrate the effectiveness of HEXCOVER, HEXCOVERTOUR over DISKCOVER, DISKCOVERTOUR [8] across environments of different sizes.

Experimental Setup: We follow the setup in [8] where a Gaussian Process was fit to a real world dataset of organic matter measurements in an agricultural field. The authors computed $L = 8.33$ meters, $\sigma_0 = 12.87$, and $\sigma^2 = 0.0361$. Since the covariance function is stationary, only the relative distances between points matter. This enables us to consider different environment sizes similar to the setup in [18]. We consider rectangular environments whose area ranges from 400 to 40000 square metres. Further, we consider three regimes of desired accuracy: $\Delta/\sigma_0^2 = 0.3, 0.2, 0.1$. These correspond to keeping the estimation error under 30%, 20%, and 10% of the initial value σ_0^2 , respectively. The experiments are run on an AMD Ryzen 7 2700 processor.

Remark. The DISKCOVER and DISKCOVERTOUR algorithms end up placing samples outside the environment, which is not practically possible. In the simulations considered, we remove these samples and the redundant measurement locations that arise (as mentioned in the paper [8]). In addition, for DISKCOVERTOUR, we use approximation algorithms for the TSP instead of the proposed lawn-mower tours. This change leads to shorter tour lengths in practice.

A. Sample Placement

The results for the sample placement problem are shown in Figure 3. In each subplot, the number of measurements (y-axis) is plotted against the environment size (x-axis). The difference across the subplots is the error tolerance. As the error tolerance decreases, the number of measurements required will increase for any algorithm. In each subplot, we observe that DISKCOVER uses more measurements than HEXCOVER (ours). In small environments, the difference is negligible as it does not take many measurements to get a

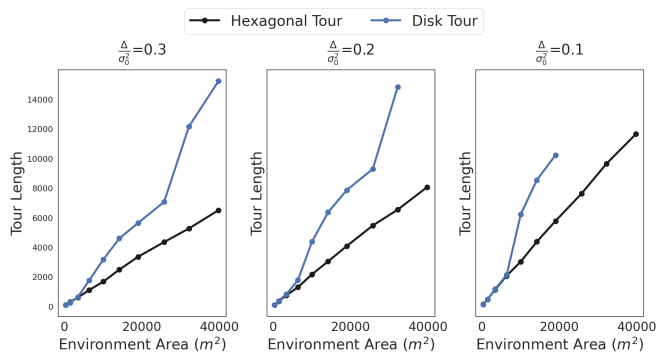


Fig. 4. Comparison of the tour length vs environment size for Problem 2. HEXCOVERTOUR plans tours of lengths roughly half that of DISKCOVERTOUR.

feasible solution. However, as the environment size increases, DISKCOVER uses roughly 6 times as many measurements as HEXCOVER. Further, the number of measurements used by HEXCOVER increases roughly linearly with the environment size. This is not the case for DISKCOVER whose measurement set size grows much faster.

B. Shortest Tours

The results for the shortest tour problem are shown in Figure 4. In each subplot, the tour length (y-axis) is plotted against the environment size (x-axis). As before, the difference across the subplots is the error tolerance. As the error tolerance decreases, the number of measurements required will increase, which subsequently increases the tour length.

In each subplot, we see that DISKCOVERTOUR produces tours of length larger than HEXCOVERTOUR (ours). For small environments, the difference in tour length is negligible. However, for larger environments, HEXCOVERTOUR produces tours that are roughly half the length of tours produced by DISKCOVERTOUR. Further, in the second and third plots, one may notice missing data points for DISKCOVERTOUR. This is because the number of vertices become too large (>8000) to run Christofides' algorithm in a reasonable amount of time and memory. In contrast, HEXCOVERTOUR scales better with the environment size.

VI. CONCLUSIONS & FUTURE DIRECTIONS

We considered the problem of finding the subset of measurement locations and the shortest tour in a random field that guarantees a desired level of prediction accuracy. We provided approximation algorithms for both problems based on hexagonal tiling of the Euclidean plane. In simulations, our algorithm outperforms prior work across environments of varying size. We also provided a counterexample to disprove a claim on a lower bound for the sample placement problem. Looking forward, we are interested in two extensions. The first involves an importance weight associated with every field location leading to a weighted estimation error constraint requiring a new approach. Second, we are interested in the multi-robot problem for field estimation where we are given a team of robots tasked with accurate field mapping.

REFERENCES

- [1] R. Webster and M. A. Oliver, *Geostatistics for environmental scientists*. John Wiley & Sons, 2007.
- [2] D. T. Connor, K. Wood, P. G. Martin, S. Goren, D. Megson-Smith, Y. Verbelen, I. Chyzhevskiy, S. Kirieiev, N. T. Smith, T. Richardson, et al., "Radiological Mapping of Post-Disaster Nuclear Environments using Fixed-Wing Unmanned Aerial Systems: a Study from Chernobyl," *Frontiers in Robotics and AI*, p. 149, 2020.
- [3] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.
- [4] A. Das and D. Kempe, "Algorithms for Subset Selection in Linear Regression," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, 2008, pp. 45–54.
- [5] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, "Efficient Informative Sensing using Multiple Robots," *Journal of Artificial Intelligence Research*, vol. 34, pp. 707–755, 2009.
- [6] M. L. Stein, *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.
- [7] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT Press, 2006, vol. 2, no. 3.
- [8] V. Suryan and P. Tokekar, "Learning a Spatial Field in Minimum Time with a Team of Robots," *IEEE Transactions on Robotics*, vol. 36, no. 5, pp. 1562–1576, 2020.
- [9] A. Krause, A. Singh, and C. Guestrin, "Near-optimal Sensor Placements in Gaussian Processes: Theory, Efficient algorithms and Empirical Studies." *Journal of Machine Learning Research*, vol. 9, no. 2, 2008.
- [10] N. Cao, K. H. Low, and J. M. Dolan, "Multi-Robot Informative Path Planning for Active Sensing of Environmental Phenomena: a Tale of Two Algorithms," in *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, 2013, pp. 7–14.
- [11] C.-W. Ko, J. Lee, and M. Queyranne, "An Exact Algorithm for Maximum Entropy Sampling," *Operations Research*, vol. 43, no. 4, pp. 684–691, 1995.
- [12] S. W. Feng, K. Gao, J. Gong, and J. Yu, "Sensor Placement for Globally Optimal Coverage of 3D-Embedded Surfaces," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 8600–8606.
- [13] J. Binney and G. S. Sukhatme, "Branch and Bound for Informative Path Planning," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 2147–2154.
- [14] S. Dutta, N. Wilde, and S. L. Smith, "Informative Path Planning in Random Fields via Mixed Integer Programming," *Conference on Decision and Control (CDC)*, 2022, To appear.
- [15] C. Chekuri and M. Pal, "A recursive greedy algorithm for walks in directed graphs," in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, 2005, pp. 245–253.
- [16] J. Binney, A. Krause, and G. S. Sukhatme, "Informative path planning for an autonomous underwater vehicle," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 4791–4796.
- [17] —, "Optimizing Waypoints for Monitoring Spatiotemporal Phenomena," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 873–888, 2013.
- [18] S. Dutta, N. Wilde, and S. L. Smith, "An Improved Greedy Algorithm for Subset Selection in Linear Estimation," in *2022 European Control Conference (ECC)*, 2022, pp. 1067–1072.
- [19] C. Xiao and J. Wachs, "Nonmyopic Informative Path Planning Based on Global Kriging Variance Minimization," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1768–1775, 2022.
- [20] G. A. Hollinger and G. S. Sukhatme, "Sampling-based Motion Planning for Robotic Information Gathering," in *Robotics: Science and Systems*, vol. 3, no. 5, 2013, pp. 1–8.
- [21] A. A. Meera, M. Popović, A. Millane, and R. Siegwart, "Obstacle-aware Adaptive Informative Path Planning for UAV-based Target Search," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 718–724.
- [22] P. Tokekar, J. Vander Hook, D. Mulla, and V. Isler, "Sensor Planning for a Symbiotic UAV and UGV System for Precision Agriculture," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1498–1511, 2016.
- [23] W. Chen and L. Liu, "Informative Planning in the Presence of Outliers," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 5311–5318.
- [24] H. Zhu, J. J. Chung, N. R. Lawrance, R. Siegwart, and J. Alonso-Mora, "Online Informative Path Planning for Active Information Gathering of a 3D Surface," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 1488–1494.
- [25] M. Popović, T. Vidal-Calleja, G. Hitz, I. Sa, R. Siegwart, and J. Nieto, "Multiresolution Mapping and Informative Path Planning for UAV-based Terrain Monitoring," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1382–1388.
- [26] M. Popović, T. Vidal-Calleja, J. J. Chung, J. Nieto, and R. Siegwart, "Informative Path Planning for Active Field Mapping under Localization Uncertainty," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10751–10757.
- [27] J. Wagberg, D. Zachariah, T. Schon, and P. Stoica, "Prediction performance after learning in gaussian process regression," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1264–1272.
- [28] A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta, "Robust Submodular Observation Selection." *Journal of Machine Learning Research*, vol. 9, no. 12, 2008.
- [29] L. Fejes, "Über die dichteste kugellagerung," *Mathematische Zeitschrift*, vol. 48, no. 1, pp. 676–684, 1942.
- [30] H.-C. Chang and L.-C. Wang, "A Simple Proof of Thue's Theorem on Circle Packing," *arXiv preprint arXiv:1009.4322*, 2010.
- [31] Y. Wu, "Lecture Notes on ECE 598: Information-theoretic Methods in High-Dimensional Statistics," <http://www.stat.yale.edu/~yw562/teaching/fit-stats.pdf>, 2016.
- [32] B. H. Korte and J. Vygen, *Combinatorial optimization*. Springer, 2011, vol. 1.