# Active Markov Information-Theoretic Path Planning for Robotic Environmental Sensing

Kian Hsiang Low
Department of Computer Science
National University of Singapore
Republic of Singapore
lowkh@comp.nus.edu.sg

John M. Dolan and Pradeep Khosla
Robotics Institute
Carnegie Mellon University
Pittsburgh PA 15213 USA
jmd@cs.cmu.edu, pkk@ece.cmu.edu

## ABSTRACT

Recent research in multi-robot exploration and mapping has focused on sampling environmental fields, which are typically modeled using the Gaussian process (GP). Existing information-theoretic exploration strategies for learning GP-based environmental field maps adopt the non-Markovian problem structure and consequently scale poorly with the length of history of observations. Hence, it becomes computationally impractical to use these strategies for *in situ*, real-time active sampling. To ease this computational burden, this paper presents a Markov-based approach to efficient information-theoretic path planning for active sampling of GP-based fields. We analyze the time complexity of solving the Markov-based path planning problem, and demonstrate analytically that it scales better than that of deriving the non-Markovian strategies with increasing length of planning horizon. For a class of exploration tasks called the transect sampling task, we provide theoretical guarantees on the active sampling performance of our Markov-based policy, from which ideal environmental field conditions and sampling task settings can be established to limit its performance degradation due to violation of the Markov assumption. Empirical evaluation on real-world temperature and plankton density field data shows that our Markov-based policy can generally achieve active sampling performance comparable to that of the widely-used non-Markovian greedy policies under less favorable realistic field conditions and task settings while enjoying significant computational gain over them.

## Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Markov processes, stochastic processes; I.2.8 [**Problem Solving, Control Methods, and Search**]: Dynamic programming; I.2.9 [**Robotics**]: Autonomous vehicles

## General Terms

Algorithms, Performance, Experimentation, Theory

## Keywords

Multi-robot exploration and mapping, adaptive sampling, active learning, Gaussian process, non-myopic path planning

**Cite as:** Active Markov Information-Theoretic Path Planning for Robotic Environmental Sensing, Kian Hsiang Low, John M. Dolan, and Pradeep Khosla, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.
Copyright © 2011, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

## 1. INTRODUCTION

Research in multi-robot exploration and mapping has recently progressed from building occupancy grids [14] to sampling spatially varying environmental phenomena [5, 6, 8], in particular, environmental fields (e.g., plankton density, pollutant concentration, temperature fields) that are characterized by *continuous-valued, spatially correlated* measurements (see Fig. 1). Exploration strategies for building occupancy grid maps usually operate under the assumptions of (a) *discrete*, (b) *independent* cell occupancies, which impose, respectively, the following limitations for learning environmental field maps: these strategies (a) cannot be fully informed by the continuous field measurements and (b) cannot exploit the spatial correlation structure of an environmental field for selecting observation paths. As a result, occupancy grid mapping strategies are not capable of selecting the most informative observation paths for learning an environmental field map.

Furthermore, occupancy grid mapping strategies typically assume that range sensing is available. In contrast, many *in situ* environmental and ecological sensing applications (e.g., monitoring of ocean phenomena, forest ecosystems, or pollution) permit only point-based sensing, thus making a high-resolution sampling of the entire field impractical in terms of resource costs (e.g., energy consumption, mission time). In practice, the resource cost constraints restrict the spatial coverage of the observation paths. Fortunately, the spatial correlation structure of an environmental field enables a map of the field (in particular, its unobserved areas) to be learned using the point-based observations taken along the resource-constrained paths. To learn this map, a commonly-used approach in spatial statistics [15] is to assume that the environmental field is realized from a probabilistic model called the *Gaussian process* (GP) (Section 3.2). More importantly, the GP model allows an environmental field to be formally characterized and consequently provides formal measures of mapping uncertainty (e.g., based on mean-squared error [5] or entropy criterion [6]) for directing a robot team to explore highly uncertain areas of the field. In this paper, we focus on using the entropy criterion to measure mapping uncertainty.

How then does a robot team plan the most informative resource-constrained observation paths to minimize the mapping uncertainty of an environmental field? To address this, the work of [6] has proposed an information-theoretic multi-robot exploration strategy that selects non-myopic observation paths with maximum entropy. Interestingly, this work has established an equivalence result that the maximum-entropy paths selected by such a strategy can achieve the

dual objective of minimizing the mapping uncertainty defined using the entropy criterion. When this strategy is applied to sampling a GP-based environmental field, it can be reduced to solving a non-Markovian, deterministic planning problem called the *information-theoretic multi-robot adaptive sampling problem* ($i$MASP) (Section 3). Due to the non-Markovian problem structure of $i$MASP, its state size grows exponentially with the length of planning horizon. To alleviate this computational difficulty, an anytime heuristic search algorithm called Learning Real-Time A$^*$ [1] is used to solve $i$MASP approximately. However, this algorithm does not guarantee the performance of its induced exploration policy. We have also observed through experiments that when the joint action space of the robot team is large or the planning horizon is long, it no longer produces a good policy fast enough. Even after incurring a huge amount of time and space to improve the search, its resulting policy still performs worse than the widely-used non-Markovian greedy policy, the latter of which can be derived efficiently by solving the myopic formulation of $i$MASP (Section 3.3).

Though the anytime and greedy algorithms provide some computational relief to solving $i$MASP (albeit approximately), they inherit $i$MASP's non-Markovian problem structure and consequently scale poorly with the length of history of observations. Hence, it becomes computationally impractical to use these non-Markovian path planning algorithms for *in situ*, real-time active sampling performed (a) at high resolution (e.g., due to high sensor sampling rate or large sampling region), (b) over dynamic features of interest (e.g., algal blooms, oil spills), (c) with resource cost constraints (e.g., energy consumption, mission time), or (d) in the presence of dynamically changing external forces translating the robots (e.g., ocean drift on autonomous boats), thus requiring fast replanning. For example, the deployment of autonomous underwater vehicles (AUVs) and boats for ocean sampling poses the above challenges/issues among others [3].

To ease this computational burden, this paper proposes a principled Markov-based approach to efficient information-theoretic path planning for active sampling of GP-based environmental fields (Section 4), which we develop by assuming the Markov property in $i$MASP planning. To the probabilistic robotics community, such a move to achieve time efficiency is probably anticipated. However, the Markov property is often imposed without carefully considering or formally analyzing its consequence on the performance degradation while operating in non-Markovian environments. In particular, to what extent does the environmental structure affect the performance degradation due to violation of the Markov assumption? Motivated by this lack of treatment, our work in this paper is novel in demonstrating both theoretically and empirically the extent of which the degradation of active sampling performance depends on the spatial correlation structure of an environmental field. An important practical consequence is that of establishing environmental field conditions under which the Markov-based approach performs well relative to the non-Markovian $i$MASP-based policy while enjoying significant computational gain over it. The specific contributions of our work include:

- analyzing the time complexity of solving the Markov-based information-theoretic path planning problem, and showing analytically that it scales better than that of deriving the non-Markovian strategies with increasing length of planning horizon (Section 4.1);
- providing theoretical guarantees on the active sampling performance of our Markov-based policy (Section 4.2) for a class of exploration tasks called the *transect sampling task* (Section 2), from which various ideal environmental field conditions and sampling task settings can be established to limit its performance degradation;
- empirically evaluating the active sampling performance and time efficiency of our Markov-based policy on real-world temperature and plankton density field data under less favorable realistic environmental field conditions and sampling task settings (Section 5).

## 2. TRANSECT SAMPLING TASK

Fig. 1 illustrates the transect sampling task introduced in [12, 13] previously. A temperature field is spatially distributed over a 25 m $\times$ 150 m transect that is discretized into a $5 \times 30$ grid of sampling locations comprising 30 columns, each of which has 5 sampling locations. It can be observed that the number of columns is much greater than the number of sampling locations in each column; this observed property is assumed to be consistent with every other transect. The robots are constrained to simultaneously explore forward one column at a time from the leftmost to the rightmost column of the transect such that each robot samples one location per column for a total of 30 locations. So, each robot's action space given its current location consists of moving to any of the 5 locations in the adjacent column on its right. The number of robots is assumed not to be larger than the number of sampling locations per column. We assume that an adversary chooses the starting robot locations in the leftmost column and the robots will only know them at the time of deployment; such an adversary can be the dynamically changing external forces translating the robots (e.g., ocean drift on autonomous boats) or the unknown obstacles occupying potential starting locations. The robots are allowed to end at any location in the rightmost column.

In practice, the constraint on exploring forward in a transect sampling task permits the planning of less complex observation paths that can be achieved more reliably, using less sophisticated control algorithms, and by robots with limited maneuverability (e.g., unmanned aerial vehicles, autonomous boats and AUVs [10]). For practical applications, while the robot is in transit from its current location to a distant planned waypoint [3, 13], this task can be performed to collect the most informative observations during transit. In monitoring of ocean phenomena and freshwater quality along rivers, the transect can span a plankton density or temperature field drifting at a constant rate from right to left and the autonomous boats are tasked to explore within a line perpendicular to the drift. As another example, the transect can be the bottom surface of ship hull or other maritime structure to be inspected and mapped by AUVs.

## 3. NON-MARKOVIAN PATH PLANNING

### 3.1 Notations and Preliminaries

Let $\mathcal{U}$ be the domain of the environmental field representing a set of sampling locations in the transect such that each location $u \in \mathcal{U}$ yields a measurement $z_u$. The columns of the transect are indexed in an increasing order from left to right with the leftmost column being indexed '0'. Each planning stage is associated with a column from which every robot in

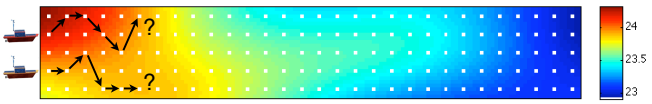



**Figure 1: Transect sampling task on a temperature field (measured in °C) spatially distributed over a 25 m × 150 m transect that is discretized into a 5 × 30 grid of sampling locations (white dots).**

the team selects and takes an observation (i.e., comprising a pair of location and its measurement). Let $k$ denote the number of robots in the team. In each stage $i$, the team of $k$ robots then collects from column $i$ a total of $k$ observations, which are denoted by a pair of vectors $x_i$ of $k$ locations and $z_{x_i}$ of the corresponding measurements. Let $x_{0:i}$ and $z_{x_{0:i}}$ denote vectors comprising the histories of robots' sampling locations and corresponding measurements over stages 0 to $i$ (i.e., concatenations of $x_0, x_1, \ldots, x_i$ and $z_{x_0}, z_{x_1}, \ldots, z_{x_i}$), respectively. Let $Z_u$, $Z_{x_i}$, and $Z_{x_{0:i}}$ be random measurements that are associated with the realizations $z_u$, $z_{x_i}$, and $z_{x_{0:i}}$, respectively.

## 3.2 Gaussian Process-Based Environmental Field

The GP model can be used to formally characterize an environmental field as follows: the environmental field is defined to vary as a realization of a GP. Let $\{Z_u\}_{u\in\mathcal{U}}$ denote a GP, i.e., every finite subset of $\{Z_u\}_{u\in\mathcal{U}}$ has a multivariate Gaussian distribution [9]. The GP is fully specified by its mean $\mu_u \triangleq \mathbb{E}[Z_u]$ and covariance $\sigma_{uv} \triangleq \mathrm{cov}[Z_u, Z_v]$ for all $u, v \in \mathcal{U}$. We assume that the GP is second-order stationary, i.e., it has a constant mean and a stationary covariance structure (i.e., $\sigma_{uv}$ is a function of $u - v$ for all $u, v \in \mathcal{U}$). In particular, its covariance structure is defined by the widely-used squared exponential covariance function [9]

$$\sigma_{uv} \triangleq \sigma_s^2 \exp\left\{-\frac{1}{2}(u-v)^\top M^{-2}(u-v)\right\} + \sigma_n^2\delta_{uv} \qquad (1)$$

where $\sigma_s^2$ is the signal variance, $\sigma_n^2$ is the noise variance, $M$ is a diagonal matrix with length-scale components $\ell_1$ and $\ell_2$ in the horizontal and vertical directions of a transect, respectively, and $\delta_{uv}$ is a Kronecker delta of value 1 if $u = v$, and 0 otherwise. Intuitively, the signal and noise variances describe, respectively, the intensity and noise of the field measurements while the length-scale can be interpreted as the approximate distance to be traversed in a transect for the field measurement to change considerably [9]; it therefore controls the degree of spatial correlation or "similarity" between field measurements. In this paper, the mean and covariance structure of the GP are assumed to be known. Given that the robot team has collected observations $x_0, z_{x_0}, x_1, z_{x_1}, \ldots, x_i, z_{x_i}$ over stages 0 to $i$, the distribution of $Z_u$ remains Gaussian with the following posterior mean and covariance

$$\mu_{u|x_{0:i}} = \mu_u + \Sigma_{ux_{0:i}}\Sigma_{x_{0:i}x_{0:i}}^{-1}\{z_{x_{0:i}} - \mu_{x_{0:i}}\}^\top \qquad (2)$$

$$\sigma_{uv|x_{0:i}} = \sigma_{uv} - \Sigma_{ux_{0:i}}\Sigma_{x_{0:i}x_{0:i}}^{-1}\Sigma_{x_{0:i}v} \qquad (3)$$

where $\mu_{x_{0:i}}$ is a row vector with mean components $\mu_w$ for every location $w$ of $x_{0:i}$, $\Sigma_{ux_{0:i}}$ is a row vector with covariance components $\sigma_{uw}$ for every location $w$ of $x_{0:i}$, $\Sigma_{x_{0:i}v}$ is a column vector with covariance components $\sigma_{wv}$ for every location $w$ of $x_{0:i}$, and $\Sigma_{x_{0:i}x_{0:i}}$ is a covariance matrix with components $\sigma_{wy}$ for every pair of locations $w, y$ of $x_{0:i}$. Note that the posterior mean $\mu_{u|x_{0:i}}$ (2) is the best unbiased predictor of the measurement $z_u$ at unobserved location $u$. An important property of GP is that the posterior covariance $\sigma_{uv|x_{0:i}}$ (3) is independent of the observed measurements $z_{x_{0:i}}$; this property is used to reduce $i$MASP to a deterministic planning problem as shown later.

## 3.3 Deterministic $i$MASP Planning

Supposing the robot team starts in locations $x_0$ of leftmost column 0, an exploration policy is responsible for directing it to sample locations $x_1, x_2, \ldots, x_{t+1}$ of the respective columns $1, 2, \ldots, t+1$ to form the observation paths. Formally, a non-Markovian policy is denoted by $\pi \triangleq \langle \pi_0(x_{0:0} = x_0), \pi_1(x_{0:1}), \ldots, \pi_t(x_{0:t})\rangle$ where $\pi_i(x_{0:i})$ maps the history $x_{0:i}$ of robots' sampling locations to a vector $a_i \in \mathcal{A}(x_i)$ of robots' actions in stage $i$ (i.e., $a_i \leftarrow \pi_i(x_{0:i})$), and $\mathcal{A}(x_i)$ is the joint action space of the robots given their current locations $x_i$. We assume that the transition function $\tau(x_i, a_i)$ *deterministically* (i.e., no localization uncertainty) moves the robots to their next locations $x_{i+1}$ in stage $i+1$ (i.e., $x_{i+1} \leftarrow \tau(x_i, a_i)$). Putting $\pi_i$ and $\tau$ together yields the assignment $x_{i+1} \leftarrow \tau(x_i, \pi_i(x_{0:i}))$.

The work of [6] has proposed a non-Markovian policy $\pi^*$ that selects non-myopic observation paths with maximum entropy for sampling a GP-based field. To know how $\pi^*$ is derived, we first define the value under a policy $\pi$ to be the entropy of observation paths when starting in $x_0$ and following $\pi$ thereafter:

$$\begin{aligned} V_0^\pi(x_0) &\triangleq \mathbb{H}[Z_{x_{1:t+1}}|Z_{x_0}, \pi] \\ &= -\int f(z_{x_{0:t+1}}|\pi)\log f(z_{x_{1:t+1}}|z_{x_0}, \pi)\, dz_{x_{0:t+1}} \end{aligned} \qquad (4)$$

where $f$ denotes a Gaussian probability density function. When a non-Markovian policy $\pi$ is plugged into (4), the following $(t+1)$-stage recursive formulation results from the chain rule for entropy and $x_{i+1} \leftarrow \tau(x_i, \pi_i(x_{0:i}))$:

$$\begin{aligned} V_i^\pi(x_{0:i}) &= \mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}, \pi_i] + V_{i+1}^\pi(x_{0:i+1}) \\ &= \mathbb{H}[Z_{\tau(x_i,\pi_i(x_{0:i}))}|Z_{x_{0:i}}] + V_{i+1}^\pi((x_{0:i}, \tau(x_i, \pi_i(x_{0:i})))) \\ V_t^\pi(x_{0:t}) &= \mathbb{H}[Z_{x_{t+1}}|Z_{x_{0:t}}, \pi_t] \\ &= \mathbb{H}[Z_{\tau(x_t,\pi_t(x_{0:t}))}|Z_{x_{0:t}}] \end{aligned} \qquad (5)$$

for stage $i = 0, \ldots, t-1$ such that each stagewise posterior entropy (i.e., of the measurements $Z_{x_{i+1}}$ to be observed in stage $i+1$ given the history of measurements $Z_{x_{0:i}}$ observed from stages 0 to $i$) reduces to

$$\mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}] = \frac{1}{2}\log\,(2\pi e)^k|\Sigma_{x_{i+1}|x_{0:i}}| \qquad (6)$$

where $\Sigma_{x_{i+1}|x_{0:i}}$ is a covariance matrix with components $\sigma_{uv|x_{0:i}}$ for every pair of locations $u, v$ of $x_{i+1}$, each of which is independent of observed measurements $z_{x_{0:i}}$ by (3), as discussed above. So, $\mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}]$ can be evaluated in closed form, and the value functions (5) only require the history of robots' sampling locations $x_{0:i}$ as inputs but not that of corresponding measurements $z_{x_{0:i}}$.

Solving $i$MASP involves choosing $\pi$ to maximize $V_0^\pi(x_0)$ (5), which yields the optimal policy $\pi^*$. Plugging $\pi^*$ into (5) gives the $(t+1)$-stage dynamic programming equations:

$$\begin{aligned} V_i^{\pi^*}(x_{0:i}) &= \max_{a_i\in\mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_{0:i}}] + V_{i+1}^{\pi^*}((x_{0:i}, \tau(x_i, a_i))) \\ V_t^{\pi^*}(x_{0:t}) &= \max_{a_t\in\mathcal{A}(x_t)} \mathbb{H}[Z_{\tau(x_t,a_t)}|Z_{x_{0:t}}] \end{aligned} \qquad (7)$$

for stage $i = 0, \ldots, t-1$. Since each stagewise posterior entropy $\mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_{0:i}}]$ (6) can be evaluated in closed form as explained above, $i$MASP for sampling the GP-based field (7) reduces to a deterministic planning problem. Furthermore, it turns out to be the well-known maximum entropy sampling problem [11] as demonstrated in [6]. Policy $\pi^* = \langle \pi_0^*(x_{0:0}), \ldots, \pi_t^*(x_{0:t}) \rangle$ can be determined by

$$\begin{array}{ll} \pi_i^*(x_{0:i}) = & \arg\max_{a_i \in \mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_{0:i}}] + V_{i+1}^{\pi^*}((x_{0:i}, \tau(x_i,a_i))) \\ \pi_t^*(x_{0:t}) = & \arg\max_{a_t \in \mathcal{A}(x_t)} \mathbb{H}[Z_{\tau(x_t,a_t)}|Z_{x_{0:t}}] \end{array} \tag{8}$$

for stage $i = 0, \ldots, t-1$. Similar to the optimal value functions (7), $\pi^*$ only requires the history of robots' sampling locations as inputs. So, $\pi^*$ can generate the maximum-entropy paths prior to exploration.

Solving the myopic formulation of $i$MASP (7) is often considered to ease computation (Section 4.1), which entails deriving the non-Markovian greedy policy $\pi^G = \langle \pi_0^G(x_{0:0}), \ldots, \pi_t^G(x_{0:t}) \rangle$ where, for stage $i = 0, \ldots, t$,

$$\pi_i^G(x_{0:i}) = \arg\max_{a_i \in \mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_{0:i}}] \ . \tag{9}$$

The work of [2] has proposed a non-Markovian greedy policy $\pi^M = \langle \pi_0^M(x_{0:0}), \ldots, \pi_t^M(x_{0:t}) \rangle$ to approximately maximize the closely related mutual information criterion:

$$\pi_i^M(x_{0:i}) = \arg\max_{a_i \in \mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_{0:i}}] - \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{\overline{x}_{0:i+1}}] \tag{10}$$

for stage $i = 0, \ldots, t$ where $\overline{x}_{0:i+1}$ denotes the vector comprising locations of domain $\mathcal{U}$ not found in $(x_{0:i}, \tau(x_i, a_i))$. It is shown in [2] that $\pi^M$ greedily selects new sampling locations that maximize the increase in mutual information. As noted in [6], this strategy is deficient in that it may not necessarily minimize the mapping uncertainty defined using the entropy criterion. More importantly, it suffers a huge computational drawback: the time needed to derive $\pi^M$ depends on the map resolution (i.e., $|\mathcal{U}|$) (Section 4.1).

# 4. MARKOV-BASED PATH PLANNING

The Markov property assumes that the measurements $Z_{x_{i+1}}$ to be observed next in stage $i+1$ depends only on the current measurements $Z_{x_i}$ observed in stage $i$ and is conditionally independent of the past measurements $Z_{x_{0:i-1}}$ observed from stages 0 to $i-1$. That is, $f(z_{x_{i+1}}|z_{x_{0:i}}) = f(z_{x_{i+1}}|z_{x_i})$ for all $z_{x_0}, z_{x_1}, \ldots, z_{x_{i+1}}$. As a result, $\mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}]$ (6) can be approximated by $\mathbb{H}[Z_{x_{i+1}}|Z_{x_i}]$. It is therefore straightforward to impose the Markov assumption on $i$MASP (7), which yields the following dynamic programming equations for the Markov-based path planning problem:

$$\begin{array}{ll} \widetilde{V}_i(x_i) = & \max_{a_i \in \mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_i}] + \widetilde{V}_{i+1}(\tau(x_i,a_i)) \\ \widetilde{V}_t(x_t) = & \max_{a_t \in \mathcal{A}(x_t)} \mathbb{H}[Z_{\tau(x_t,a_t)}|Z_{x_t}] \ . \end{array} \tag{11}$$

for stage $i = 0, \ldots, t-1$. Consequently, the Markov-based policy $\widetilde{\pi} = \langle \widetilde{\pi}_0(x_0), \ldots, \widetilde{\pi}_t(x_t) \rangle$ can be determined by

$$\begin{array}{ll} \widetilde{\pi}_i(x_i) = & \arg\max_{a_i \in \mathcal{A}(x_i)} \mathbb{H}[Z_{\tau(x_i,a_i)}|Z_{x_i}] + \widetilde{V}_{i+1}(\tau(x_i,a_i)) \\ \widetilde{\pi}_i(x_t) = & \arg\max_{a_t \in \mathcal{A}(x_t)} \mathbb{H}[Z_{\tau(x_t,a_t)}|Z_{x_t}] \ . \end{array} \tag{12}$$

## 4.1 Time Complexity: Analysis & Comparison

THEOREM 1. *Let $\mathcal{A} \triangleq \mathcal{A}(x_0) = \ldots = \mathcal{A}(x_t)$. Deriving the Markov-based policy $\widetilde{\pi}$ (12) for the transect sampling task requires $\mathcal{O}(|\mathcal{A}|^2(t+k^4))$ time.*

Note that $|\mathcal{A}| = {}^r\mathrm{C}_k = \mathcal{O}(r^k)$ where $r$ is the number of sampling locations per column and $k \leq r$ as assumed in Section 2. Though $|\mathcal{A}|$ is exponential in the number $k$ of robots, $r$ is expected to be small in a transect, which prevents $|\mathcal{A}|$ from growing too large.

In contrast, deriving $i$MASP-based policy $\pi^*$ (8) requires $\mathcal{O}(|\mathcal{A}|^t t^2 k^4)$ time. Deriving greedy policies $\pi^G$ (9) and $\pi^M$ (10) incur, respectively, $\mathcal{O}(|\mathcal{A}|t^4k^3 + |\mathcal{A}|^2tk^4)$ and $\mathcal{O}(|\mathcal{A}|t|\mathcal{U}|^3 + |\mathcal{A}|^2tk^4) = \mathcal{O}(|\mathcal{A}|t^4r^3 + |\mathcal{A}|^2tk^4)$ time to compute the observation paths over all $|\mathcal{A}|$ possible choices of starting robot locations. Clearly, all the non-Markovian strategies do not scale as well as our Markov-based approach with increasing length $t+1$ of planning horizon or number $t+2$ of columns, which is expected to be large. As demonstrated empirically (Section 5), the Markov-based policy $\widetilde{\pi}$ can be derived faster than $\pi^G$ and $\pi^M$ by more than an order of magnitude; this computational advantage is boosted further for transect sampling tasks with unknown starting robot locations.

## 4.2 Performance Guarantees

We will first provide a theoretical guarantee on how the Markov-based policy $\widetilde{\pi}$ (12) performs relative to the non-Markovian $i$MASP-based policy $\pi^*$ (8) for the case of 1 robot. This key result follows from our intuition that when the horizontal spatial correlation becomes small, exploiting the past measurements for path planning should hardly improve the active sampling performance in a transect sampling task, thus favoring the Markov-based policy. Though this intuition is simple, supporting it with formal theoretical results (and their corresponding proofs reported elsewhere [7]) turns out to be non-trivial as shown below.

Recall the Markov assumption that $\mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}]$ (6) is to be approximated by $\mathbb{H}[Z_{x_{i+1}}|Z_{x_i}]$. This prompts us to first consider bounding the difference of these posterior entropies that ensues from the Markov property:

$$\begin{aligned} \mathbb{H}[Z_{x_{i+1}}|Z_{x_i}] - \mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}] &= \frac{1}{2}\log\frac{\sigma^2_{x_{i+1}|x_i}}{\sigma^2_{x_{i+1}|x_{0:i}}} \\ &= \frac{1}{2}\log\left(1 - \frac{\sigma^2_{x_{i+1}|x_i} - \sigma^2_{x_{i+1}|x_{0:i}}}{\sigma^2_{x_{i+1}|x_i}}\right)^{-1} \geq 0 \ . \end{aligned} \tag{13}$$

This difference can be interpreted as the reduction in uncertainty of the measurements $Z_{x_{i+1}}$ to be observed next in stage $i+1$ by observing the past measurements $Z_{x_{0:i-1}}$ from stages 0 to $i-1$ given the current measurements $Z_{x_i}$ observed in stage $i$. This difference is small if $Z_{x_{0:i-1}}$ does not contribute much to the reduction in uncertainty of $Z_{x_{i+1}}$ given $Z_{x_i}$. It (13) is often known as the conditional mutual information of $Z_{x_{i+1}}$ and $Z_{x_{0:i-1}}$ given $Z_{x_i}$ denoted by

$$\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}] \triangleq \mathbb{H}[Z_{x_{i+1}}|Z_{x_i}] - \mathbb{H}[Z_{x_{i+1}}|Z_{x_{0:i}}] \ ,$$

which is of value 0 if the Markov property holds.

The results to follow assume that the transect is discretized into a grid of sampling locations. Let $\omega_1$ and $\omega_2$ denote the horizontal and vertical grid discretization widths (i.e., separations between adjacent sampling locations), respectively. Let $\ell_1' \triangleq \ell_1/\omega_1$ and $\ell_2' \triangleq \ell_2/\omega_2$ represent the normalized horizontal and vertical length-scale components, respectively.

The following lemma bounds the variance reduction term $\sigma^2_{x_{i+1}|x_i} - \sigma^2_{x_{i+1}|x_{0:i}}$ in (13):

LEMMA 2. *Let* $\xi \triangleq \exp\left\{-\frac{1}{2\ell_1'^2}\right\}$ *and* $\rho \triangleq 1 + \frac{\sigma_n^2}{\sigma_s^2}$. *If* $\xi < \frac{\rho}{i}$, *then* $0 \leq \sigma^2_{x_{i+1}|x_i} - \sigma^2_{x_{i+1}|x_{0:i}} \leq \frac{\sigma_s^2\xi^4}{\frac{\rho}{i} - \xi}$.

The next lemma is fundamental to the subsequent results on the active sampling performance of Markov-based policy $\widetilde{\pi}$. It provides bounds on $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$, which follow immediately from (13), Lemma 2, and the lower bound $\sigma^2_{x_{i+1}|x_i} = \sigma^2_{x_{i+1}} - (\sigma_{x_{i+1}x_i})^2/\sigma^2_{x_i} \geq \sigma_s^2 + \sigma_n^2 - \sigma_s^2\xi^2$ :

LEMMA 3. *If* $\xi < \frac{\rho}{i}$, *then* $0 \leq \mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}] \leq \Delta(i)$ *where* $\Delta(i) \triangleq \frac{1}{2}\log\left(1 - \frac{\xi^4}{(\frac{\rho}{i} - \xi)(\rho - \xi^2)}\right)^{-1}$.

REMARK. *If* $j \leq s$, *then* $\Delta(j) \leq \Delta(s)$ *for* $j, s = 0, \ldots, t$.

From Lemma 3, since $\Delta(i)$ bounds $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$ from above, a small $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$ can be guaranteed by making $\Delta(i)$ small. From the definition of $\Delta(i)$, there are a few ways to achieve a small $\Delta(i)$: (a) $\Delta(i)$ depends on $\ell_1'$ through $\xi$. As $\ell_1' \to 0^+$, $\xi \to 0^+$, by definition. Consequently, $\Delta(i) \to 0^+$. A small $\ell_1'$ can be obtained using a small $\ell_1$ and/or a large $\omega_1$, by definition; (b) $\Delta(i)$ also depends on the noise-to-signal ratio $\sigma_n^2/\sigma_s^2$ through $\rho$. Raising $\sigma_n^2$ or lowering $\sigma_s^2$ increases $\rho$, by definition. This, in turn, decreases $\Delta(i)$; (c) Since $i$ indicates the length of history of observations, the remark after Lemma 3 tells us that a shorter length produces a smaller $\Delta(i)$. To summarize, (a) environmental field conditions such as smaller horizontal spatial correlation and noisy, less intense fields, and (b) sampling task settings such as larger horizontal grid discretization width and shorter length of history of observations all contribute to smaller $\Delta(i)$, and hence smaller $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$. This analysis is important for understanding the practical implication of our theoretical results later. A limitation with using Lemma 3 is that of the sufficient condition $\xi < \rho/i$, which will hold if the field conditions and task settings realized above to make $\Delta(i)$ small are adequately satisfied.

The following theorem uses the induced optimal value $\widetilde{V}_0(x_0)$ from solving the Markov-based path planning problem (11) to bound the maximum entropy $V_0^{\pi^*}(x_0)$ of observation paths achieved by $\pi^*$ from solving $i$MASP (7):

THEOREM 4. *Let* $\epsilon_i \triangleq \sum_{s=i}^{t}\Delta(s) \leq (t-i+1)\Delta(t)$. *If* $\xi < \frac{\rho}{t}$, *then* $\widetilde{V}_i(x_i) - \epsilon_i \leq V_i^{\pi^*}(x_{0:i}) \leq \widetilde{V}_i(x_i)$ *for* $i = 0, \ldots, t$.

The above result is useful in providing an efficient way of knowing the maximum entropy $V_0^{\pi^*}(x_0)$, albeit approximately: the time needed to derive the two-sided bounds on $V_0^{\pi^*}(x_0)$ is linear in the length of planning horizon (Theorem 1) as opposed to exponential time required to compute the exact value of $V_0^{\pi^*}(x_0)$. Since the error bound $\epsilon_i$ is defined as a sum of $\Delta(s)$'s, we can rely on the above analysis of $\Delta(s)$ (see paragraph after Lemma 3) to improve this error bound: (a) environmental field conditions such as smaller horizontal spatial correlation and noisy, less intense fields, and (b) sampling task settings such as larger horizontal grid discretization width and shorter planning horizon (i.e., fewer transect columns) all improve this error bound.

In the main result below, the Markov-based policy $\widetilde{\pi}$ is guaranteed to achieve an entropy $V_0^{\widetilde{\pi}}(x_0)$ of observation paths (i.e., by plugging $\widetilde{\pi}$ into (5)) that is not more than $\epsilon_0$ from the maximum entropy $V_0^{\pi^*}(x_0)$ of observation paths achieved by policy $\pi^*$:

THEOREM 5. *If* $\xi < \frac{\rho}{t}$, *then policy* $\widetilde{\pi}$ *is* $\epsilon_0$*-optimal in achieving the maximum-entropy criterion, i.e.,* $V_0^{\pi^*}(x_0) - V_0^{\widetilde{\pi}}(x_0) \leq \epsilon_0$.

Again, since the error bound $\epsilon_0$ is defined as a sum of $\Delta(s)$'s, we can use the above analysis of $\Delta(s)$ to improve this bound: (a) environmental field conditions such as smaller horizontal spatial correlation and noisy, less intense fields, and (b) sampling task settings such as larger horizontal grid discretization width and shorter planning horizon (i.e., fewer transect columns) all result in smaller $\epsilon_0$, and hence improve the active sampling performance of Markov-based policy $\widetilde{\pi}$ relative to that of non-Markovian $i$MASP-based policy $\pi^*$. This not only supports our prior intuition (see first paragraph of this section) but also identifies other means of limiting the performance degradation of the Markov-based policy.

For the multi-robot case, a condition has to be imposed on the covariance structure of GP to obtain a similar guarantee:

$$|\sigma_{uv|x_{0:i}}| \leq |\sigma_{uv|x_m}| \tag{14}$$

for $m = 0, \ldots, i$ and any $u, v, x_0, x_1, \ldots, x_i \in \mathcal{U}$. Intuitively, (14) says that further conditioning does not make $Z_u$ and $Z_v$ more correlated. Note that (14) is satisfied if $u = v$.

Similar to Lemma 3 for the 1-robot case, we can bound $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$ for the multi-robot case but tighter conditions have to be satisfied:

LEMMA 6. *Let* $\ell_1' = \ell_2'$. *If* $\xi < \min(\frac{\rho}{ik}, \frac{\rho}{4k})$ *and* (14) *is satisfied, then* $0 \leq \mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}] \leq \Delta_k(i)$ *where* $\Delta_k(i) \triangleq \frac{k}{2}\log\left(1 - \frac{\xi^4}{(\frac{\rho}{ik} - \xi)(\rho - \frac{4k}{\rho}\xi^2)}\right)^{-1}$.

To improve the upper bound $\Delta_k(i)$, the above analysis of $\Delta(i)$ can be applied here as these two upper bounds are largely similar: (a) environmental field conditions such as smaller spatial correlation and noisy, less intense fields, and (b) sampling task settings such as larger grid discretization width and shorter planning horizon (i.e., fewer transect columns) all entail smaller $\Delta_k(i)$. Decreasing the number $k$ of robots also reduces $\Delta_k(i)$, thus yielding tighter bounds on $\mathbb{I}[Z_{x_{i+1}}; Z_{x_{0:i-1}}|Z_{x_i}]$. Using Lemma 6, we can derive guarantees similar to that of Theorems 4 and 5 on the performance of Markov-based policy $\widetilde{\pi}$ for the multi-robot case.

# 5. EXPERIMENTS AND DISCUSSION

In Section 4.2, we have highlighted the practical implication of our main theoretical result (i.e., Theorem 5), which establishes various environmental field conditions and sampling task settings to limit the performance degradation of Markov-based policy $\widetilde{\pi}$. This result, however, does not reveal whether $\widetilde{\pi}$ performs well (or not) under "seemingly" less favorable field conditions and task settings that do not jointly satisfy its sufficient condition $\xi < \rho/(tk)$. These include large spatial correlation, less noisy, highly intense fields, small grid discretization width, long planning horizon (i.e., many transect columns), and large number of robots. So, this section evaluates the active sampling performance

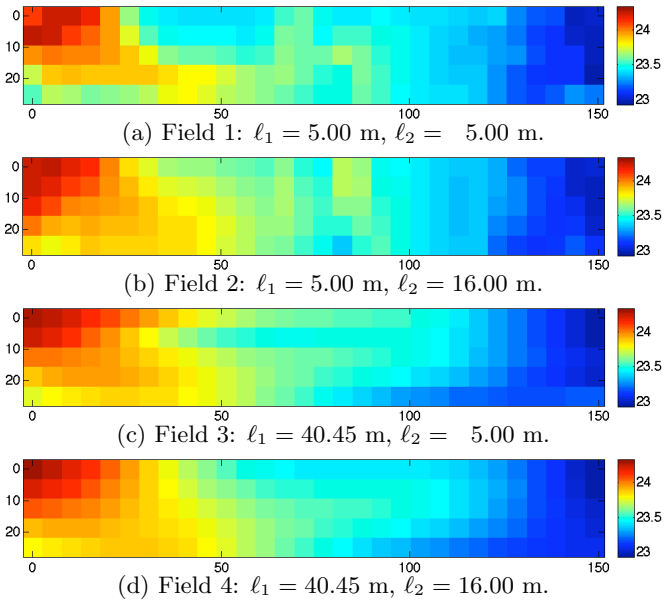
and time efficiency of $\widetilde{\pi}$ empirically on two real-world datasets under such field conditions and task settings as detailed below: (a) May 2009 temperature field data of Panther Hollow Lake in Pittsburgh, PA spanning 25 m by 150 m, and (b) June 2009 plankton density field data of Chesapeake Bay spanning 314 m by 1765 m.

Using maximum likelihood estimation (MLE) [9], the learned hyperparameters (i.e., horizontal and vertical length-scales, signal and noise variances) are, respectively, $\ell_1 = 40.45$ m, $\ell_2 = 16.00$ m, $\sigma_s^2 = 0.1542$, and $\sigma_n^2 = 0.0036$ for the temperature field, and $\ell_1 = 27.53$ m, $\ell_2 = 134.64$ m, $\sigma_s^2 = 2.152$, and $\sigma_n^2 = 0.041$ for the plankton density field. It can be observed that the temperature and plankton density fields have low noise-to-signal ratios $\sigma_n^2/\sigma_s^2$ of 0.023 and 0.019, respectively. Relative to the size of transect, both fields have large vertical spatial correlations, but only the temperature field has large horizontal spatial correlation.

The performance of Markov-based policy $\widetilde{\pi}$ is compared to non-Markovian policies produced by two state-of-the-art information-theoretic exploration strategies: greedy policies $\pi^G$ (9) and $\pi^M$ (10) proposed by [6] and [2], respectively. The non-Markovian policy $\pi^*$ that has to be derived approximately using Learning Real-Time A$^*$ is excluded from comparison due to the reason provided in Section 1.

## 5.1 Performance Metrics

The tested policies are evaluated using the two metrics proposed in [6], which quantify the mapping uncertainty of the unobserved areas of the field differently: (a) The ENT($\pi$) metric measures the posterior joint entropy $\mathbb{H}[Z_{\overline{x}_{0:t+1}}|Z_{x_{0:t+1}}]$ of field measurements $Z_{\overline{x}_{0:t+1}}$ at unobserved locations $\overline{x}_{0:t+1}$ where $\overline{x}_{0:t+1}$ denotes the vector comprising locations of domain $\mathcal{U}$ not found in the sampled locations $x_{0:t+1}$ selected by policy $\pi$. Smaller ENT($\pi$) implies lower mapping uncertainty; (b) The ERR($\pi$) metric measures the mean-squared relative error $|\mathcal{U}|^{-1}\sum_{u\in\mathcal{U}}\{(z_u-\mu_{u|x_{0:t+1}})/\bar{\mu}\}^2$ resulting from using the observations (i.e., sampled locations $x_{0:t+1}$ and corresponding measurements $z_{x_{0:t+1}}$) selected by policy $\pi$ and the posterior mean $\mu_{u|x_{0:t+1}}$ (2) to predict the field where $\bar{\mu} = |\mathcal{U}|^{-1}\sum_{u\in\mathcal{U}} z_u$. Smaller ERR($\pi$) implies higher prediction accuracy. Two noteworthy differences distinguish these metrics: (a) The ENT($\pi$) metric exploits the spatial correlation between field measurements in the unobserved areas whereas the ERR($\pi$) metric implicitly assumes independence between them. As a result, unlike the ERR($\pi$) metric, the ENT($\pi$) metric does not overestimate the mapping uncertainty. To illustrate this, suppose the unknown field measurements are restricted to only two unobserved locations $u$ and $v$ residing in a highly uncertain area and they are highly correlated due to spatial proximity. The behavior of the ENT($\pi$) metric can be understood upon applying the chain rule for entropy (i.e., $\text{ENT}(\pi) = \mathbb{H}[Z_u, Z_v|Z_{x_{0:t+1}}] = \mathbb{H}[Z_u|Z_{x_{0:t+1}}] + \mathbb{H}[Z_v|Z_{x_{0:t+1}}, Z_u]$); the latter uncertainty term (i.e., posterior entropy of $Z_v$) is significantly reduced or "discounted" due to the high spatial correlation between $Z_u$ and $Z_v$. Hence, the mapping uncertainty of these two unobserved locations is not overestimated. A practical advantage of this metric is that it does not overcommit sensing resources; in the simple illustration above, a single observation at either location $u$ or $v$ suffices to learn both field measurements well. On the other hand, the ERR($\pi$) metric considers each location to be of high uncertainty due to the independence assumption; (b) In contrast to the ENT($\pi$) metric, the ERR($\pi$) metric can use ground truth measurements to evaluate if the field is being mapped accurately. Let $\text{ENTD}(\pi) \triangleq \text{ENT}(\widetilde{\pi}) - \text{ENT}(\pi)$ and $\text{ERRD}(\pi) \triangleq \text{ERR}(\widetilde{\pi}) - \text{ERR}(\pi)$. Decreasing ENTD($\pi$) improves the ENT($\widetilde{\pi}$) performance of $\widetilde{\pi}$ relative to that of $\pi$. Small $|\text{ENTD}(\pi)|$ implies that $\widetilde{\pi}$ achieves ENT($\widetilde{\pi}$) performance comparable to that of $\pi$. ERRD($\pi$) can be interpreted likewise. Additionally, we will consider the time taken to derive each policy as the third metric.



(a) Field 1: $\ell_1 = 5.00$ m, $\ell_2 = 5.00$ m.

(b) Field 2: $\ell_1 = 5.00$ m, $\ell_2 = 16.00$ m.

(c) Field 3: $\ell_1 = 40.45$ m, $\ell_2 = 5.00$ m.

(d) Field 4: $\ell_1 = 40.45$ m, $\ell_2 = 16.00$ m.

**Figure 2: Temperature fields (measured in °C) with varying horizontal length-scale $\ell_1$ and vertical length-scale $\ell_2$.**

## 5.2 Temperature Field Data

We will first investigate how varying spatial correlations (i.e., varying length-scales) of the temperature field affect the ENT($\pi$) and ERR($\pi$) performance of evaluated policies. The temperature field is discretized into a $5\times30$ grid of sampling locations as shown in Figs. 1 and 2d. The horizontal and/or vertical length-scales of the original field (i.e., field 4 in Fig. 2d) are reduced to produce modified fields 1, 2, and 3 (respectively, Figs. 2a, 2b, and 2c); we fix these reduced length-scales while learning the remaining hyperparameters (i.e., signal and noise variances) through MLE.

Table 1 shows the results of mean ENT($\pi$) and ERR($\pi$) performance of tested policies (i.e., averaged over all possible starting robot locations) with varying length-scales and number of robots. The ENT($\pi$) and ERR($\pi$) for all policies generally decrease with increasing length-scales (except ERR($\widetilde{\pi}$) for 1 robot from field 2 to 4) due to increasing spatial correlation between measurements, thus resulting in lower mapping uncertainty.

For the case of 1 robot, the observations are as follows: (a) When $\ell_2$ is kept constant (i.e., at 5 m or 16 m), reducing $\ell_1$ from 40.45 m to 5 m (i.e., from field 3 to 1 or field 4 to 2) decreases ENTD($\pi^G$), ERRD($\pi^G$), ENTD($\pi^M$), and ERRD($\pi^M$): when the horizontal correlation becomes small, it can no longer be exploited by the non-Markovian policies $\pi^G$ and $\pi^M$; (b) For field 3 with large $\ell_1$ and small $\ell_2$, ENTD($\pi^G$) and ENTD($\pi^M$) are large as the Markov property of $\widetilde{\pi}$ prevents it from exploiting the large horizontal

**Table 1: Comparison of ENT($\pi$) (left) and ERR($\pi$) ($\times 10^{-5}$) (right) performance for temperature fields that are discretized into $5 \times 30$ grids (Fig. 2).**

| 1 robot | Field | | | | 1 robot | Field | | | |
|---|---|---|---|---|---|---|---|---|---|
| Policy | 1 | 2 | 3 | 4 | Policy | 1 | 2 | 3 | 4 |
| $\widetilde{\pi}$ | -83 | -246 | -543 | -597 | $\widetilde{\pi}$ | 3.7040 | 0.5713 | 2.3680 | 0.5754 |
| $\pi^G$ | -82 | -246 | -554 | -598 | $\pi^G$ | 1.8680 | 0.5713 | 0.0801 | 0.0252 |
| $\pi^M$ | -80 | -211 | -554 | -596 | $\pi^M$ | 1.8433 | 0.5212 | 0.0701 | 0.0421 |
| 2 robots | Field | | | | 2 robots | Field | | | |
| Policy | 1 | 2 | 3 | 4 | Policy | 1 | 2 | 3 | 4 |
| $\widetilde{\pi}$ | -71 | -190 | -380 | -422 | $\widetilde{\pi}$ | 0.3797 | 0.2101 | 0.1171 | 0.0095 |
| $\pi^G$ | -72 | -190 | -382 | -425 | $\pi^G$ | 0.3526 | 0.2101 | 0.0150 | 0.0087 |
| $\pi^M$ | -68 | -131 | -382 | -421 | $\pi^M$ | 0.6714 | 0.1632 | 0.0148 | 0.0086 |
| 3 robots | Field | | | | 3 robots | Field | | | |
| Policy | 1 | 2 | 3 | 4 | Policy | 1 | 2 | 3 | 4 |
| $\widetilde{\pi}$ | -53 | -109 | -232 | -297 | $\widetilde{\pi}$ | 0.1328 | 0.0068 | 0.0063 | 0.0031 |
| $\pi^G$ | -53 | -109 | -215 | -297 | $\pi^G$ | 0.1312 | 0.0068 | 0.0059 | 0.0031 |
| $\pi^M$ | -53 | -73 | -214 | -255 | $\pi^M$ | 0.1080 | 0.1397 | 0.0055 | 0.0030 |

**Table 2: Comparison of ENT($\pi$) (left) and ERR($\pi$) ($\times 10^{-5}$) (right) performance for temperature field that is discretized into $13 \times 75$ grid.**

| ENT($\pi$) | Number $k$ of robots | | | ERR($\pi$) | Number $k$ of robots | | |
|---|---|---|---|---|---|---|---|
| Policy | 1 | 2 | 3 | Policy | 1 | 2 | 3 |
| $\widetilde{\pi}$ | -4813 | -4284 | -3828 | $\widetilde{\pi}$ | 1.0287 | 0.0032 | 0.0015 |
| $\pi^G$ | -4813 | -4286 | -3841 | $\pi^G$ | 0.0082 | 0.0030 | 0.0024 |
| $\pi^M$ | -4808 | -4277 | -3825 | $\pi^M$ | 0.0087 | 0.0034 | 0.0019 |

correlation; (c) When $\ell_1$ is kept constant (i.e., at 5 m or 40.45 m), reducing $\ell_2$ from 16 m to 5 m (i.e., from field 2 to 1 or field 4 to 3) increases ERRD($\pi^G$) and ERRD($\pi^M$): when vertical correlation becomes small, it can no longer be exploited by $\widetilde{\pi}$, thus incurring larger ERR($\widetilde{\pi}$).

For the case of 2 robots, the observations are as follows: (a) |ENTD($\pi^G$)| and |ENTD($\pi^M$)| are small for all fields except for field 2 where $\widetilde{\pi}$ significantly outperforms $\pi^M$. In particular, when $\ell_2$ is kept constant (i.e., at 5 m or 16 m), reducing $\ell_1$ from 40.45 m to 5 m (i.e., from field 3 to 1 or field 4 to 2) decreases ENTD($\pi^G$), ENTD($\pi^M$), and ERRD($\pi^G$): this is explained in the first observation of 1-robot case; (b) For field 3 with large $\ell_1$ and small $\ell_2$, ERRD($\pi^G$) and ERRD($\pi^M$) are large: this is explained in the second and third observations of 1-robot case; (c) When $\ell_1$ is kept constant (i.e., at 5 m or 40.45 m), reducing $\ell_2$ from 16 m to 5 m (i.e., from field 2 to 1 or field 4 to 3) increases ERRD($\pi^G$): this is explained in the third observation of 1-robot case. This also holds for ERRD($\pi^M$) when $\ell_1$ is large.

For the case of 3 robots, it can be observed that $\widetilde{\pi}$ can achieve ENT($\widetilde{\pi}$) and ERR($\widetilde{\pi}$) performance comparable to (if not, better than) that of $\pi^G$ and $\pi^M$ for all fields.

To summarize the above observations on spatial correlation conditions favoring $\widetilde{\pi}$ over $\pi^G$ and $\pi^M$, $\widetilde{\pi}$ can achieve ENT($\widetilde{\pi}$) performance comparable to (if not, better than) that of $\pi^G$ and $\pi^M$ for all fields with any number of robots except for field 3 (i.e., of large $\ell_1$ and small $\ell_2$) with 1 robot as explained previously. Policy $\widetilde{\pi}$ can achieve comparable ERR($\widetilde{\pi}$) performance for field 2 (i.e., of small $\ell_1$ and large $\ell_2$) with 1 robot because $\widetilde{\pi}$ is capable of exploiting the large vertical correlation, and the small horizontal correlation cannot be exploited by $\pi^G$ and $\pi^M$. Policy $\widetilde{\pi}$ can also achieve comparable ERR($\widetilde{\pi}$) performance for all fields with 2 and 3 robots except for field 3 (i.e., of large $\ell_1$ and small $\ell_2$) with 2 robots. These observations reveal that (a) small horizontal and large vertical correlations are favorable to $\widetilde{\pi}$; (b) though large horizontal and small vertical correlations are not favorable to $\widetilde{\pi}$, this problem can be mitigated by increasing the number of robots. For more detailed analysis (e.g., visualization of planned observation paths and their corresponding error maps), the interested reader is referred to [4].

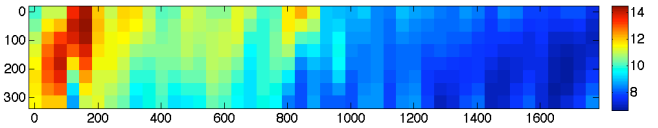



**Figure 3: Plankton density (chl-a) field (measured in $mg\ m^{-3}$) spatially distributed over a 314 m $\times$ 1765 m transect that is discretized into a $8 \times 45$ grid with $\ell_1 = 27.53$ m and $\ell_2 = 134.64$ m.**

We will now examine how the increase in resolution to $13\times 75$ grid affects the ENT($\pi$) and ERR($\pi$) performance of evaluated policies; the resulting grid discretization width and planning horizon are about $0.4\times$ smaller and $2.5\times$ longer, respectively. Table 2 shows the results of mean ENT($\pi$) and ERR($\pi$) performance of tested policies with varying number of robots, from which we can derive observations similar to that for temperature field 4 discretized into $5 \times 30$ grid: $\widetilde{\pi}$ can achieve ENT($\widetilde{\pi}$) and ERR($\widetilde{\pi}$) performance comparable to (if not, better than) that of $\pi^G$ and $\pi^M$ except for ERR($\widetilde{\pi}$) performance with 1 robot. So, increasing the grid resolution does not seem to noticeably degrade the active sampling performance of $\widetilde{\pi}$ relative to that of $\pi^G$ and $\pi^M$.

## 5.3 Plankton Density Field Data

Fig. 3 illustrates the plankton density field that is discretized into a $8 \times 45$ grid. Table 3 shows the results of mean ENT($\pi$) and ERR($\pi$) performance of tested policies with varying number of robots. The observations are as follows: $\widetilde{\pi}$ can achieve the same ENT($\widetilde{\pi}$) and ERR($\widetilde{\pi}$) performance as that of $\pi^G$ and superior ENT($\widetilde{\pi}$) performance over that of $\pi^M$ because small horizontal and large vertical correlations favor $\widetilde{\pi}$ as explained in Section 5.2. By increasing the number of robots (i.e., $k > 2$), $\widetilde{\pi}$ can achieve ERR($\widetilde{\pi}$) performance comparable to (if not, better than) that of $\pi^M$.

Table 4 shows the results of mean ENT($\pi$) and ERR($\pi$) performance of tested policies after increasing the resolution to $16 \times 89$ grid; the resulting grid discretization width and planning horizon are about $0.5\times$ smaller and $2\times$ longer, respectively. Similar observations can be obtained: $\widetilde{\pi}$ can achieve ENT($\widetilde{\pi}$) performance comparable to that of $\pi^G$ and superior ENT($\widetilde{\pi}$) performance over that of $\pi^M$. By deploying more than 1 robot, $\widetilde{\pi}$ can achieve ERR($\widetilde{\pi}$) performance comparable to (if not, better than) that of $\pi^G$ and $\pi^M$. Again, we can observe that increasing the grid resolution does not seem to noticeably degrade the active sampling performance of $\widetilde{\pi}$ relative to that of $\pi^G$ and $\pi^M$.

## 5.4 Incurred Policy Time

Fig. 4 shows the time taken to derive the tested policies for sampling the temperature and plankton density fields with varying number of robots and grid resolutions. It can be observed that the time taken to derive $\widetilde{\pi}$ is shorter than that needed to derive $\pi^G$ and $\pi^M$ by more than 1 and 4 orders of magnitude, respectively. It is important to point out that Fig. 4 reports the average time taken to derive $\pi^G$ and $\pi^M$ over all possible starting robot locations. So, if the starting robot locations are unknown, the incurred time to derive $\pi^G$ and $\pi^M$ have to be increased by ${}^r\mathrm{C}_k$-fold. In contrast, $\widetilde{\pi}$ caters to all possible starting robot locations. So, the incurred time to derive $\widetilde{\pi}$ is unaffected. These observations show a considerable computational gain of $\widetilde{\pi}$ over $\pi^G$ and $\pi^M$, which supports our time complexity analysis and comparison (Section 4). So, our Markov-based path planner

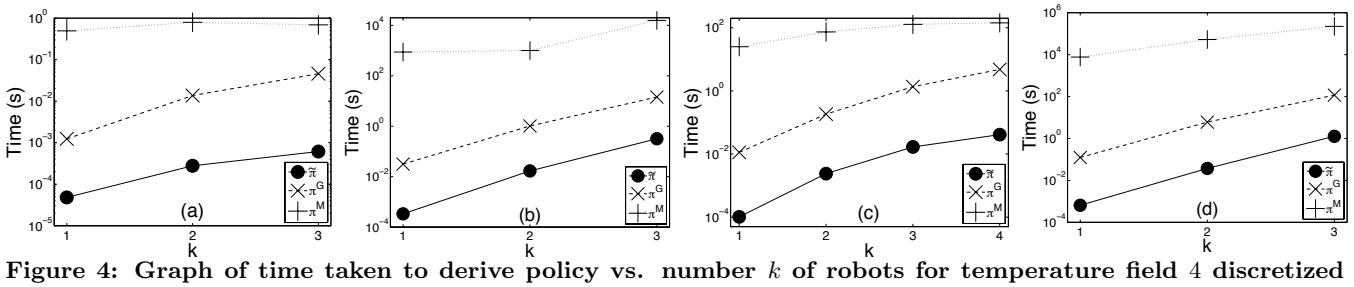



**Figure 4: Graph of time taken to derive policy vs. number $k$ of robots for temperature field 4 discretized into (a) $5 \times 30$ and (b) $13 \times 75$ grids and plankton density field discretized into (c) $8 \times 45$ and (d) $16 \times 89$ grids.**

**Table 3: Comparison of ENT($\pi$) (left) and ERR($\pi$) ($\times 10^{-3}$) (right) performance for plankton density field that is discretized into $8 \times 45$ grid.**

| ENT($\pi$) | Number $k$ of robots | | | |
|---|---|---|---|---|
| Policy | 1 | 2 | 3 | 4 |
| $\widetilde{\pi}$ | -359 | -322 | -196 | -121 |
| $\pi^G$ | -359 | -322 | -196 | -121 |
| $\pi^M$ | -230 | -186 | -70 | -11 |

| ERR($\pi$) | Number $k$ of robots | | | |
|---|---|---|---|---|
| Policy | 1 | 2 | 3 | 4 |
| $\widetilde{\pi}$ | 5.6124 | 2.2164 | 0.0544 | 0.0066 |
| $\pi^G$ | 5.6124 | 2.2164 | 0.0544 | 0.0066 |
| $\pi^M$ | 4.5371 | 0.5613 | 0.0472 | 0.0324 |

**Table 4: Comparison of ENT($\pi$) (left) and ERR($\pi$) ($\times 10^{-3}$) (right) performance for plankton density field that is discretized into $16 \times 89$ grid.**

| ENT($\pi$) | Number $k$ of robots | | |
|---|---|---|---|
| Policy | 1 | 2 | 3 |
| $\widetilde{\pi}$ | -4278 | -3949 | -3681 |
| $\pi^G$ | -4238 | -3964 | -3686 |
| $\pi^M$ | -4171 | -3840 | -3501 |

| ERR($\pi$) | Number $k$ of robots | | |
|---|---|---|---|
| Policy | 1 | 2 | 3 |
| $\widetilde{\pi}$ | 3.4328 | 0.0970 | 0.0546 |
| $\pi^G$ | 1.5648 | 0.1073 | 0.0643 |
| $\pi^M$ | 0.8186 | 0.0859 | 0.0348 |

is more time-efficient for *in situ*, real-time, high-resolution active sampling.

# 6. CONCLUSION

This paper describes an efficient Markov-based information-theoretic path planner for active sampling of GP-based environmental fields. We have provided theoretical guarantees on the active sampling performance of our Markov-based policy $\widetilde{\pi}$ for the transect sampling task, from which ideal environmental field conditions (i.e., small horizontal spatial correlation and noisy, less intense fields) and sampling task settings (i.e., large grid discretization width and short planning horizon) can be established to limit its performance degradation. Empirically, we have shown that $\widetilde{\pi}$ can generally achieve active sampling performance comparable to that of the widely-used non-Markovian greedy policies $\pi^G$ and $\pi^M$ under less favorable realistic field conditions (i.e., low noise-to-signal ratio) and task settings (i.e., small grid discretization width and long planning horizon) while enjoying huge computational gain over them. In particular, we have empirically observed that (a) small horizontal and large vertical correlations strongly favor $\widetilde{\pi}$; (b) though large horizontal and small vertical correlations do not favor $\widetilde{\pi}$, this problem can be mitigated by increasing the number of robots. In fact, deploying a large robot team often produces superior active sampling performance of $\widetilde{\pi}$ over $\pi^M$ in our experiments, not forgetting the computational gain of $> 4$ orders of magnitude. Our Markov-based planner can be used to efficiently achieve more general exploration tasks (e.g., boundary tracking and those in [5, 6]), but the guarantees provided here may not apply. For our future work, we will "relax" the Markov assumption by utilizing a longer (but not entire) history of observations in path planning. This can potentially improve the active sampling performance in fields of moderate to large horizontal correlation but does not incur as much time as that of non-Markovian policies.

# 7. REFERENCES


[1] R. Korf. Real-time heuristic search. *Artif. Intell.*, 42(2-3):189–211, 1990.

[2] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *JMLR*, 9:235–284, 2008.

[3] N. E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. Davis. Collective motion, sensor networks and ocean sampling. *Proc. IEEE*, 95(1):48–74, 2007.

[4] K. H. Low. *Multi-Robot Adaptive Exploration and Mapping for Environmental Sensing Applications.* Ph.D. Thesis, Technical Report CMU-ECE-2009-024, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 2009.

[5] K. H. Low, J. M. Dolan, and P. Khosla. Adaptive multi-robot wide-area exploration and mapping. In *Proc. AAMAS*, pages 23–30, 2008.

[6] K. H. Low, J. M. Dolan, and P. Khosla. Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing. In *Proc. ICAPS*, pages 233–240, 2009.

[7] K. H. Low, J. M. Dolan, and P. Khosla. Active Markov information-theoretic path planning for robotic environmental sensing. arXiv:1101.5632, 2011.

[8] K. H. Low, G. J. Gordon, J. M. Dolan, and P. Khosla. Adaptive sampling for multi-robot wide-area exploration. In *Proc. ICRA*, pages 755–760, 2007.

[9] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, MA, 2006.

[10] D. L. Rudnick, R. E. Davis, C. C. Eriksen, D. Fratantoni, and M. J. Perry. Underwater gliders for ocean research. *Mar. Technol. Soc. J.*, 38(2):73–84, 2004.

[11] M. C. Shewry and H. P. Wynn. Maximum entropy sampling. *J. Applied Stat.*, 14(2):165–170, 1987.

[12] A. Ståhl, Ringvall, and T. Lämås. Guided transect sampling for assessing sparse populations. *Forest Science*, 46(1):108–115, 2000.

[13] D. R. Thompson and D. Wettergreen. Intelligent maps for autonomous kilometer-scale science survey. In *Proc. i-SAIRAS*, 2008.

[14] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics.* MIT Press, Cambridge, MA, 2005.

[15] R. Webster and M. Oliver. *Geostatistics for Environmental Scientists.* John Wiley & Sons, Inc., NY, 2nd edition, 2007.