Information-Theoretic Approach to Efficient Adaptive Path Planning for Mobile Robotic Environmental Sensing

Kian Hsiang Low[†] and **John M. Dolan**[†§] and **Pradeep Khosla**[†§]
Department of Electrical and Computer Engineering[†], Robotics Institute[§]
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
{bryanlow, jmd}@cs.cmu.edu, pkk@ece.cmu.edu

Abstract

Recent research in robot exploration and mapping has focused on sampling environmental hotspot fields. This exploration task is formalized by Low, Dolan, and Khosla (2008) in a sequential decision-theoretic planning under uncertainty framework called MASP. The time complexity of solving MASP approximately depends on the map resolution, which limits its use in large-scale, high-resolution exploration and mapping. To alleviate this computational difficulty, this paper presents an information-theoretic approach to MASP (iMASP) for efficient adaptive path planning; by reformulating the cost-minimizing iMASP as a reward-maximizing problem, its time complexity becomes independent of map resolution and is less sensitive to increasing robot team size as demonstrated both theoretically and empirically. Using the reward-maximizing dual, we derive a novel adaptive variant of maximum entropy sampling, thus improving the induced exploration policy performance. It also allows us to establish theoretical bounds quantifying the performance advantage of optimal adaptive over non-adaptive policies and the performance quality of approximately optimal vs. optimal adaptive policies. We show analytically and empirically the superior performance of iMASP-based policies for sampling the log-Gaussian process to that of policies for the widely-used Gaussian process in mapping the hotspot field. Lastly, we provide sufficient conditions that, when met, guarantee adaptivity has no benefit under an assumed environment model.

Introduction

Recent research in multi-robot exploration and mapping (Low, Dolan, and Khosla 2008; Singh et al. 2007) has focused on sampling environmental fields, some of which typically feature a few small *hotspots* in a large region (Webster and Oliver 2007). Such a *hotspot field* often arises in environmental and ecological sensing applications such as precision agriculture, mineral prospecting, monitoring of ocean phenomena, forest ecosystems, pollution, or contamination. In particular, the hotspot field (e.g., plankton density and mineral distribution in Fig. 2) is characterized by *continuous, positively skewed, spatially correlated* measurements with the hotspots exhibiting extreme measurements and much higher spatial variability than the rest of the field. With limited (e.g., point-based) robot sensing range, a complete coverage becomes impractical in terms of resource costs (e.g., energy consumption). So, to accurately map the field, the hotspots have to be sampled at a higher resolution.

The hotspot field discourages static sensor placement (Guestrin, Krause, and Singh 2005) because a large number of sensors has to be positioned to detect and refine the sampling of hotspots. If these static sensors are not placed in any hotspot initially, they cannot reposition by themselves to locate one. In contrast, a robot team is capable of performing high-resolution hotspot sampling due to its mobility. Hence, it is desirable to build a mobile robot team that can actively explore to map a hotspot field.

To learn a hotspot field map, the *exploration strategy* of the robot team has to plan resource-constrained observation paths that minimize the map uncertainty of the hotspot field. To achieve this, the recent work of Low, Dolan, and Khosla (2008) has proposed such a strategy that plans non-myopic adaptive paths to minimize the uncertainty of a spatial model of the hotspot field. In particular, both (a) modeling and (b) planning components are designed to fully exploit the environmental structure in order to yield a high-quality map: (a) The hotspot field is assumed to be realized from a non-parametric probabilistic model called the log-Gaussian process, which can provide a formal measure of map uncertainty and more importantly, characterize the abovementioned hotspot field measurements well; (b) The exploration task is formalized in a sequential decision-theoretic planning under uncertainty framework, which we call the *multi-robot adaptive sampling problem* (MASP). So, MASP can be viewed as a sequential, non-myopic version of active learning. In contrast to finite-state Markov decision problems, MASP adopts a more complex but realistic continuous-state, *non-Markovian* problem structure so that its induced exploration policy can be informed by the complete history of continuous, spatially correlated observations for selecting paths. It is unique in unifying formulations of exploration problems along the entire adaptivity (see Def. 2) spectrum, thus subsuming existing non-adaptive formulations and allowing the performance advantage of a more adaptive policy to be theoretically realized. Through MASP, it is demonstrated that a more adaptive strategy can exploit clustering phenomena in a hotspot field to produce

lower map uncertainty.

However, MASP is besieged by a serious computational drawback due to its measure of map uncertainty using the mean-squared error criterion. Consequently, the time complexity of solving MASP (approximately) depends on the map resolution, which limits its practical use in large-scale, high-resolution exploration and mapping.

The principal contribution of this paper is to alleviate this computational difficulty through an information-theoretic approach to MASP (iMASP) for efficient adaptive path planning, which measures map uncertainty based on the entropy criterion instead. Unlike MASP, reformulating the cost-minimizing iMASP as a reward-maximizing problem causes its time complexity of being solved approximately to be independent of the map resolution and less sensitive to larger robot team size as demonstrated both theoretically and empirically in this paper. Additional contributions stemming from this reward-maximizing formulation include:

- transforming the commonly-used non-adaptive maximum entropy sampling problem (Shewry and Wynn 1987) into a novel adaptive variant, thus improving the performance of the induced exploration policy;
- establishing theoretical bounds to quantify the performance advantage of optimal adaptive over non-adaptive exploration policies, and the performance quality of approximately optimal vs. optimal adaptive policies;
- given an assumed environment model (e.g., occupancy grid map), establishing sufficient conditions that, when met, guarantee adaptivity provides no benefit; and
- showing analytically and empirically the superior performance of iMASP-based policies for sampling the log-Gaussian process (ℓGP) to that of policies for the widely-used Gaussian process (GP) (Guestrin, Krause, and Singh 2005; Shewry and Wynn 1987; Singh et al. 2007) in mapping the hotspot field.

Related Work. Beyond its computational gain, iMASP retains the beneficial properties of MASP: it is novel in the class of model-based exploration strategies to perform both wide-area coverage and hotspot sampling. The former considers sparsely sampled areas to be of high uncertainty and thus spreads the observations evenly across the environmental field. The latter expects areas of high uncertainty to contain highly-varying measurements and hence produces clustered observations. Like MASP, iMASP also covers the entire adaptivity spectrum, thus subsuming the existing non-adaptive entropy-based problem formulation (Shewry and Wynn 1987). In contrast, all other model-based strategies (Meliou et al. 2007; Singh et al. 2007) are non-adaptive and achieve only wide-area coverage; they are observed to perform well only with smoothly-varying fields. Similar to MASP, iMASP can plan non-myopic multi-robot paths, which are more desirable than greedy or single-robot paths (Meliou et al. 2007; Singh et al. 2007).

Cost-Minimizing Problem Formulations

We formalize here the information-theoretic exploration problems at the two extremes of the adaptivity spectrum. Exploration problems residing within the spectrum can be formalized in a similar manner. Note that the use of the entropy criterion in non-myopic active learning is not new but is limited to the non-adaptive problem formulation (Shewry and Wynn 1987), which is presented here as a comparison to the novel adaptive problem formulation. It can be observed that the resulting cost-minimizing formulations differ from that of MASP by only the entropy criterion. However, as we shall see in a later section, their reward-maximizing dual formulations are significantly different from that of MASP in terms of interpretation and computational complexity.

Notation and Preliminaries. Let $\mathcal{X}$ be the domain of the hotspot field corresponding to a finite set of grid cell locations. An observation taken (e.g., by a single robot) at stage i comprises a pair of location $x_i \in \mathcal{X}$ and its measurement z_{x_i}. More generally, k observations taken (e.g., by k robots or 1 robot taking k observations) at stage i can be represented by a pair of vectors $\mathbf{x}_i$ of k locations and $\mathbf{z}_{\mathbf{x}_i}$ of the corresponding measurements.

Definition 1 (Posterior Data) *The posterior data d_i at stage $i > 0$ comprises*

- *the prior data $d_0 = \langle \mathbf{x}_0, \mathbf{z}_{\mathbf{x}_0} \rangle$ available at stage 0, and*
- *a complete history of observations $\mathbf{x}_1, \mathbf{z}_{\mathbf{x}_1}, \ldots, \mathbf{x}_i, \mathbf{z}_{\mathbf{x}_i}$ induced by k observations per stage over stages 1 to i.*

Let $\mathbf{x}_{0:i}$ and $\mathbf{z}_{\mathbf{x}_{0:i}}$ denote vectors comprising the location and measurement components of the posterior data d_i (i.e., concatenations of $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_i$ and $\mathbf{z}_{\mathbf{x}_0}, \mathbf{z}_{\mathbf{x}_1}, \ldots, \mathbf{z}_{\mathbf{x}_i}$), respectively. Let $\overline{\mathbf{x}}_{0:i}$ denote the vector comprising locations of domain $\mathcal{X}$ not observed in d_i, and $\mathbf{z}_{\overline{\mathbf{x}}_{0:i}}$ be the vector comprising the corresponding measurements. Let Z_{x_i}, $\mathbf{Z}_{\mathbf{x}_i}$, $\mathbf{Z}_{\mathbf{x}_{0:i}}$, $\mathbf{Z}_{\overline{\mathbf{x}}_{0:i}}$ be the random measurements corresponding to the respective realizations z_{x_i}, $\mathbf{z}_{\mathbf{x}_i}$, $\mathbf{z}_{\mathbf{x}_{0:i}}$, $\mathbf{z}_{\overline{\mathbf{x}}_{0:i}}$.

Definition 2 (Characterizing Adaptivity) *Suppose prior data d_0 are available and n new locations are to be explored. Then, an exploration strategy is*

- **adaptive** *if its policy to select each vector $\mathbf{x}_{i+1}$ of k new locations depends only on the previously sampled data d_i for $i = 0, \ldots, n/k - 1$. So, this strategy selects k observations per stage over n/k stages. If $k = 1$, this strategy is strictly adaptive. Increasing k makes it partially adaptive;*
- **non-adaptive** *if its policy to select each new location x_{i+1} for $i = 0, \ldots, n-1$ is independent of the measurements $z_{x_1}, \ldots, z_{x_n}$. As a result, all n new locations $x_1, \ldots, x_n$ can be selected prior to exploration. That is, this strategy selects all n observations in a single stage.*

Objective Function. The exploration objective is to plan observation paths that minimize the uncertainty of mapping the hotspot field. To achieve this, we use the entropy criterion to measure map uncertainty. Given the posterior data d_n, the *posterior map entropy* of domain $\mathcal{X}$ can be represented by the posterior joint entropy of the measurements $\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}$ at the unobserved locations $\overline{\mathbf{x}}_{0:n}$:

$$\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}} | d_n] \triangleq -\int f(\mathbf{z}_{\overline{\mathbf{x}}_{0:n}} | d_n) \log f(\mathbf{z}_{\overline{\mathbf{x}}_{0:n}} | d_n) \, d\mathbf{z}_{\overline{\mathbf{x}}_{0:n}} \quad (1)$$

where f denotes a probability density function.

Value Function. If only the prior data d_0 are available, an exploration strategy has to produce a policy for selecting observation paths that minimize the *expected* posterior

map entropy instead. This policy must then collect the optimal observations $\mathbf{x}_1, \mathbf{z}_{\mathbf{x}_1}, \ldots, \mathbf{x}_n, \mathbf{z}_{\mathbf{x}_n}$ during exploration to form posterior data d_n. The value under an exploration policy π is defined to be the expected posterior map entropy (i.e., expectation of (1)) when starting in d_0 and following π thereafter:

$$\begin{aligned} V_0^{\pi}(d_0) &\stackrel{\triangle}{=} \mathbb{E}\{\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]|d_0, \pi\} \\ &= \int f(\mathbf{z}_{\mathbf{x}_{1:n}}|d_0, \pi)\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\mathbf{x}_{1:n}} \ . \end{aligned} \tag{2}$$

The strategies of Guestrin, Krause, and Singh (2005) and Singh et al. (2007) have optimized a closely related *mutual information* criterion that measures the expected entropy reduction of unobserved locations $\overline{\mathbf{x}}_{0:n}$ by observing $\mathbf{x}_{1:n}$ (i.e., $\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_0] - \mathbb{E}\{\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]|d_0\}$). This is deficient for the exploration objective because mutual information may be maximized by a choice of $\mathbf{x}_{1:n}$ inducing a very large prior entropy $\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_0]$ but not necessarily the smallest expected posterior map entropy $\mathbb{E}\{\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]|d_0\}$.

In the next two subsections, we will describe how the adaptive and non-adaptive exploration policies can be derived to minimize the expected posterior map entropy (2).

Adaptive Exploration. The adaptive policy π for directing a team of k robots is structured to collect k observations per stage over a finite planning horizon of n stages. This implies each robot observes 1 location per stage and is thus constrained to explore at most n new locations over n stages. Formally, $\pi \stackrel{\triangle}{=} \langle \pi_0(d_0), \ldots, \pi_{n-1}(d_{n-1}) \rangle$ where $\pi_i : d_i \rightarrow \mathbf{a}_i$ maps data d_i to a vector of robots' actions $\mathbf{a}_i \in \mathcal{A}(\mathbf{x}_i)$ at stage i, and $\mathcal{A}(\mathbf{x}_i)$ is the joint action space of the robots given their current locations $\mathbf{x}_i$. We assume the transition function $\tau : \mathbf{x}_i \times \mathbf{a}_i \rightarrow \mathbf{x}_{i+1}$ *deterministically* moves the robots to their next locations $\mathbf{x}_{i+1}$ at stage $i+1$. Combining π_i and τ gives $\mathbf{x}_{i+1} \leftarrow \tau(\mathbf{x}_i, \pi_i(d_i))$. We can observe from this assignment that the sequential (i.e., stagewise) selection of k new locations $\mathbf{x}_{i+1}$ to be included in the observation paths depends only on the previously sampled data d_i along the paths for stage $i = 0, \ldots, n-1$. Hence, policy π is adaptive (Def. 2).

Solving the adaptive exploration problem iMASP(1) means choosing the adaptive policy π to minimize $V_0^{\pi}(d_0)$ (2), which we call the *optimal adaptive policy* π^1. That is, $V_0^{\pi^1}(d_0) = \min_{\pi} V_0^{\pi}(d_0)$. Plugging π^1 into (2) gives the n-stage dynamic programming equations:

$$\begin{aligned} V_i^{\pi^1}(d_i) &= \int f(\mathbf{z}_{\mathbf{x}_{i+1}}|d_i, \pi_i^1)\, V_{i+1}^{\pi^1}(d_{i+1})\, \mathrm{d}\mathbf{z}_{\mathbf{x}_{i+1}} \\ &= \int f(\mathbf{z}_{\tau(\mathbf{x}_i, \pi_i^1(d_i))}|d_i)\, V_{i+1}^{\pi^1}(d_{i+1})\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_i, \pi_i^1(d_i))} \\ &= \min_{\mathbf{a}_i \in \mathcal{A}(\mathbf{x}_i)} \int f(\mathbf{z}_{\tau(\mathbf{x}_i, \mathbf{a}_i)}|d_i)\, V_{i+1}^{\pi^1}(d_{i+1})\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_i, \mathbf{a}_i)} \\ V_n^{\pi^1}(d_n) &= \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n] \end{aligned} \tag{3}$$

for stage $i = 0, \ldots, n-1$. The first and second equalities follow from $f(\mathbf{z}_{\mathbf{x}_{1:n}}|d_0, \pi^1) = \Pi_{i=0}^{n-1} f(\mathbf{z}_{\mathbf{x}_{i+1}}|d_i, \pi_i^1)$ and $\mathbf{x}_{i+1} \leftarrow \tau(\mathbf{x}_i, \pi_i^1(d_i))$ respectively. Policy $\pi^1 = \langle \pi_0^1(d_0), \ldots, \pi_{n-1}^1(d_{n-1}) \rangle$ can be determined in a stagewise manner by

$$\pi_i^1(d_i) = \arg\min_{\mathbf{a}_i \in \mathcal{A}(\mathbf{x}_i)} \int f(\mathbf{z}_{\tau(\mathbf{x}_i, \mathbf{a}_i)}|d_i)\, V_{i+1}^{\pi^1}(d_{i+1})\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_i, \mathbf{a}_i)} \ .$$

Note that the optimal action $\pi_0^1(d_0)$ at stage 0 can be determined prior to exploration using prior data d_0. However, each action rule $\pi_i^1(d_i)$ at stage $i = 1, \ldots, n-1$ defines the optimal action to take in response to d_i, part of which (i.e., $\mathbf{x}_1, \mathbf{z}_{\mathbf{x}_1}, \ldots, \mathbf{x}_i, \mathbf{z}_{\mathbf{x}_i}$) are only observed during exploration.

Non-Adaptive Exploration. The non-adaptive policy π is structured to collect, in 1 stage, n observations per robot with a team of k robots. So, each robot is also constrained to explore at most n new locations, but they have to do this within 1 stage. Formally, $\pi \stackrel{\triangle}{=} \pi_0(d_0)$ where $\pi_0 : d_0 \rightarrow \mathbf{a}_{0:n-1}$ maps prior data d_0 to a vector $\mathbf{a}_{0:n-1}$ of action components concatenating a sequence of robots' actions $\mathbf{a}_0, \ldots, \mathbf{a}_{n-1}$. Combining π_0 and τ gives $\mathbf{x}_{1:n} \leftarrow \tau(\mathbf{x}_{0:n-1}, \pi_0(d_0))$. We can observe from this assignment that the selection of $k \times n$ new locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to form the observation paths are independent of the measurements $\mathbf{z}_{\mathbf{x}_1}, \ldots, \mathbf{z}_{\mathbf{x}_n}$ obtained along the paths during exploration. Hence, policy π is non-adaptive (Def. 2) and all new locations can be selected in a single stage prior to exploration.

Solving the non-adaptive exploration problem iMASP(n) involves choosing π to minimize $V_0^{\pi}(d_0)$ (2), which we call the *optimal non-adaptive policy* π^n (i.e., $V_0^{\pi^n}(d_0) = \min_{\pi} V_0^{\pi}(d_0)$). Plugging π^n into (2) gives the 1-stage equation:

$$\begin{aligned} V_0^{\pi^n}(d_0) &= \int f(\mathbf{z}_{\mathbf{x}_{1:n}}|d_0, \pi_0^n)\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\mathbf{x}_{1:n}} \\ &= \int f(\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \pi_0^n(d_0))}|d_0)\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \pi_0^n(d_0))} \\ &= \min_{\mathbf{a}_{0:n-1}} \int f(\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \mathbf{a}_{0:n-1})}|d_0)\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \mathbf{a}_{0:n-1})}. \end{aligned} \tag{4}$$

The second equality follows from $\mathbf{x}_{1:n} \leftarrow \tau(\mathbf{x}_{0:n-1}, \pi_0^n(d_0))$ described above. Policy $\pi^n = \pi_0^n(d_0)$ can therefore be determined in a single stage by $\pi_0^n(d_0) =$

$$\arg\min_{\mathbf{a}_{0:n-1}} \int f(\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \mathbf{a}_{0:n-1})}|d_0)\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_{0:n-1}, \mathbf{a}_{0:n-1})} \ .$$

Note that the optimal sequence of robots' actions $\pi_0^n(d_0)$ (i.e., optimal observation paths) can be determined prior to exploration since the prior data d_0 are available.

Reward-Maximizing Dual Formulations

In this section, we transform the cost-minimizing iMASP(1) (3) and iMASP(n) (4) into reward-maximizing problems and show their equivalence. The reward-maximizing iMASP(n) turns out to be the well-known *maximum entropy sampling* (MES) problem (Shewry and Wynn 1987):

$$U_0^{\pi^n}(d_0) = \max_{\mathbf{a}_{0:n-1}} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_{0:n-1}, \mathbf{a}_{0:n-1})}|d_0] \ , \tag{5}$$

which is a single-staged problem of selecting $k \times n$ new locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with maximum entropy to form the observation paths. This dual ensues from the equivalence result $V_0^{\pi^n}(d_0) = \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_0}|d_0] - U_0^{\pi^n}(d_0)$ relating cost-minimizing and reward-maximizing iMASP(n)'s in the non-adaptive exploration setting, which follows from the chain rule of entropy. This result says the original objective of minimizing expected posterior map entropy (i.e., $V_0^{\pi^n}(d_0)$ (4)) is equivalent to that of discharging from prior map entropy $\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_0}|d_0]$ the largest entropy into the selected paths (i.e., $U_0^{\pi^n}(d_0)$ (5)). Hence, their optimal non-adaptive policies coincide.

Our reward-maximizing iMASP(1) is a novel adaptive variant of MES. Unlike the cost-minimizing iMASP(1), it can be subject to convex analysis, which allows monotone-bounding approximations to be developed as shown later. It comprises the following n-stage dynamic programming equations:

$$U_i^{\pi^1}(d_i) = \max_{\mathbf{a}_i \in \mathcal{A}(\mathbf{x}_i)} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i] + \int f(\mathbf{z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i)\, U_{i+1}^{\pi^1}(d_{i+1})\, \mathrm{d}\mathbf{z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)} \quad (6)$$

$$U_t^{\pi^1}(d_t) = \max_{\mathbf{a}_t \in \mathcal{A}(\mathbf{x}_t)} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_t,\mathbf{a}_t)}|d_t]$$

for stage $i = 0, \ldots, t-1$ where $t = n-1$. Each stagewise reward reflects the entropy of k new locations $\mathbf{x}_{i+1}$ to be potentially selected into the paths. By maximizing the sum of expected rewards over n stages in (6), the reward-maximizing iMASP(1) absorbs the largest expected entropy into the selected paths. In the adaptive exploration setting, the cost-minimizing and reward-maximizing iMASP(1)'s are also equivalent (i.e., their optimal adaptive policies coincide):

Theorem 3 *$V_i^{\pi^1}(d_i) = \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i}}|d_i] - U_i^{\pi^1}(d_i)$ for stage $i = 0, \ldots, n-1$.*

The work of Low, Dolan, and Khosla (2008) has also provided an equivalence result to relate the cost-minimizing and reward-maximizing MASPs through the use of the variance decomposition formula in its induction proof. In contrast, the induction proof to Theorem 3 uses the chain rule of entropy, which entails a computational complexity reduction (not available to MASP) as described next.

In cost-minimizing iMASP(1), the time complexity of evaluating the cost (i.e., posterior map entropy (1)) depends on the domain size $|\mathcal{X}|$ for the environment models described in the next section. By transforming into the dual, the time complexity of evaluating each stagewise reward becomes independent of $|\mathcal{X}|$ because it reflects only the uncertainty of the new locations to be potentially selected into the observation paths. As a result, the runtime of the approximation algorithm proposed in a later section does not depend on the map resolution, which is clearly advantageous in large-scale, high-resolution exploration and mapping. In contrast, the reward-maximizing MASP (Low, Dolan, and Khosla 2008) utilizing the mean-squared error criterion does not share this computational advantage, as the time needed to evaluate each stagewise reward still depends on $|\mathcal{X}|$. We will evaluate this computational advantage using time complexity analysis in a later section.

Learning the Hotspot Field Map

Traditionally, a hotspot is defined as a location where its measurement exceeds a pre-defined extreme. But, hotspot locations do not usually occur in isolation but in clusters. So, it is useful to characterize hotspots with spatial properties. Accordingly, we define a hotspot field to vary as a realization of a spatial random field $\{Y_x > 0\}_{x \in \mathcal{X}}$ such that putting together the observed measurements of a realization $\{y_x\}_{x \in \mathcal{X}}$ gives a positively skewed 1D sample frequency distribution (e.g., Fig. 1b). In this section, we will highlight the problem with modeling the hotspot field directly using GP and explain how the ℓGP remedies this. We will also show analytically that the iMASP-based policy for sampling ℓGP is adaptive and exploits clustering phenomena but that for sampling GP lacks these properties.

Gaussian Process. A widely-used random field to model environmental phenomena is the GP (Guestrin, Krause, and Singh 2005; Meliou et al. 2007; Singh et al. 2007). The stationary assumption on the GP covariance structure is very sensitive to strong positive skewness of hotspot field measurements (e.g., Fig. 1b) and is easily violated by a few extreme ones (Webster and Oliver 2007). In practice, this can cause reconstructed fields to display large hotspots centered about a few extreme observations and prediction variances to be unrealistically small in hotspots, which are undesirable. So, if GP is used to model a hotspot field directly, it may not map well. To remedy this, a standard statistical practice is to take the log of the measurements (i.e., $Z_x = \log Y_x$) to remove skewness and extremity (e.g., Fig. 1c), and use GP to map the *log-measurements*. As a result, the entropy criterion (1) has to be optimized in the transformed log-scale.

We will apply iMASP(1) to sampling GP and determine if π^1 exhibits adaptive and hotspot sampling properties. Let $\{Z_x\}_{x \in \mathcal{X}}$ denote a GP, i.e., the joint distribution over any finite subset of $\{Z_x\}_{x \in \mathcal{X}}$ is Gaussian (Rasmussen and Williams 2006). The GP can be completely specified by its mean $\mu_{Z_x} \triangleq \mathbb{E}[Z_x]$ and covariance $\sigma_{Z_x Z_u} \triangleq \mathrm{cov}[Z_x, Z_u]$ for $x, u \in \mathcal{X}$. We adopt a common assumption that the GP is second-order stationary, i.e., it has a constant mean and a stationary covariance structure (i.e., $\sigma_{Z_x Z_u}$ is a function of $x - u$ for all $x, u \in \mathcal{X}$). In this paper, we assume that the mean and covariance structure of Z_x are known. Given d_n, the distribution of Z_x is Gaussian with posterior mean and covariance

$$\mu_{Z_x|d_n} = \mu_{Z_x} + \Sigma_{x\mathbf{x}_{0:n}} \Sigma_{\mathbf{x}_{0:n}\mathbf{x}_{0:n}}^{-1} \{\mathbf{z}_{\mathbf{x}_{0:n}} - \boldsymbol{\mu}_{\mathbf{Z}_{\mathbf{x}_{0:n}}}\}^\top \quad (7)$$

$$\sigma_{Z_x Z_u|d_n} = \sigma_{Z_x Z_u} - \Sigma_{x\mathbf{x}_{0:n}} \Sigma_{\mathbf{x}_{0:n}\mathbf{x}_{0:n}}^{-1} \Sigma_{\mathbf{x}_{0:n}u} \quad (8)$$

where, for every pair of locations v, w of $\mathbf{x}_{0:n}$, $\boldsymbol{\mu}_{\mathbf{Z}_{\mathbf{x}_{0:n}}}$ is a row vector with mean components μ_{Z_v}, $\Sigma_{x\mathbf{x}_{0:n}}$ is a row vector with covariance components $\sigma_{Z_x Z_v}$, $\Sigma_{\mathbf{x}_{0:n}u}$ is a column vector with covariance components $\sigma_{Z_v Z_u}$, and $\Sigma_{\mathbf{x}_{0:n}\mathbf{x}_{0:n}}$ is a covariance matrix with components $\sigma_{Z_v Z_w}$. An important property of $\sigma_{Z_x Z_u|d_n}$ is its independence of $\mathbf{z}_{\mathbf{x}_{1:n}}$.

Policy π^1 can be reduced to be *non-adaptive*: observe that each stagewise reward is independent of the measurements

$$\mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i] = \log \sqrt{(2\pi e)^k\, |\Sigma_{\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i}|} \quad (9)$$

where $\Sigma_{\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i}$ is a covariance matrix with components $\sigma_{Z_x Z_u|d_i}$, x, u of $\tau(\mathbf{x}_i, \mathbf{a}_i)$, that are independent of $\mathbf{z}_{\mathbf{x}_{1:n}}$. As a result, it follows from (6) that $U_i^{\pi^1}(d_i)$ and $\pi_i^1(d_i)$ are independent of $\mathbf{z}_{\mathbf{x}_{1:n}}$ for $i = 0, \ldots, n-1$. The expectations in iMASP(1) (6) can then be integrated out. As a result, iMASP(1) for sampling GP can be reduced to a 1-stage deterministic problem $U_0^{\pi^1}(d_0) = \sum_{i=0}^{n-1} \max_{\mathbf{a}_i} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i] = \max_{\mathbf{a}_0,\ldots,\mathbf{a}_{n-1}} \sum_{i=0}^{n-1} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i] = \max_{\mathbf{a}_{0:n-1}} \mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_{0:n-1},\mathbf{a}_{0:n-1})}|d_0] = U_0^{\pi^n}(d_0)$. This indicates the induced optimal values from solving iMASP(1) and iMASP(n) are equal. So, π^1 offers no performance advantage over π^n.

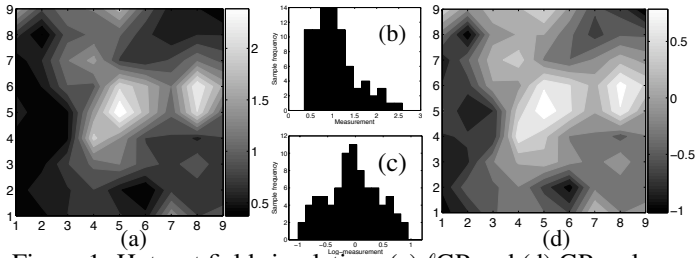

Figure 1: Hotspot field simulation: (a) ℓGP and (d) GP realizations with their 1D sample frequency distributions shown, respectively, in (b) and (c).

Based on the above analysis, the following sufficient conditions, when met, guarantee that adaptivity has no benefit under an assumed environmental model:

Theorem 4 *If $\mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i]$ is independent of $\mathbf{z}_{\mathbf{x}_{1:n}}$ for stage $i = 0,\ldots,n-1$, iMASP(1) and π^1 can be reduced to be single-staged and non-adaptive, respectively.*

For example, Theorem 4 also holds for the simple case of an *occupancy grid map* modeling an obstacle-ridden environment, which typically assumes z_x for $x \in \mathcal{X}$ to be independent. As a result, $\mathbb{H}[\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i]$ can be reduced to a sum of prior entropies over the unobserved locations $\tau(\mathbf{x}_i,\mathbf{a}_i)$, which are independent of $\mathbf{z}_{\mathbf{x}_{1:n}}$.

Policy π^1 performs *wide-area coverage* only: to maximize stagewise rewards (9), π^1 selects new locations with large posterior variance for observation. If we assume isotropic covariance structure (i.e., the covariance $\sigma_{Z_x Z_u}$ decreases monotonically with $||x-u||$) (Rasmussen and Williams 2006), the posterior data d_i provide the least amount of information on unobserved locations that are far away from all observed locations. As a result, the posterior variance of unobserved locations in sparsely sampled regions are still largely unreduced by the posterior data d_i from the observed locations. Hence, by exploring the sparsely sampled areas, a large expected entropy can be absorbed into the selected observation paths. Using the observations selected from wide-area coverage, the field of *original* measurements may not be mapped well because the under-sampled hotspots with extreme, highly-varying measurements contribute considerably to map entropy in the original scale, as discussed below.

Log-Gaussian Process. To map the original, rather than the log-, measurements directly, it is a conventional practice in geostatistics to use the ℓGP. Consequently, the entropy criterion (1) is optimized in the original scale. To do this, let $\{Y_x\}_{x\in\mathcal{X}}$ denote a ℓGP: if $Z_x = \log Y_x$, $\{Z_x\}_{x\in\mathcal{X}}$ is a GP. So, the positive-valued $Y_x = \exp\{Z_x\}$ denotes the original random measurement at location x. It is straightforward to derive the predictive properties of ℓGP from that of GP as shown in (Low, Dolan, and Khosla 2008).

A ℓGP can model a field with hotspots that exhibit much higher spatial variability than the rest of the field: Figs. 1a and 1d compare the realizations of ℓGP and GP; the GP realization results from taking the log of the ℓGP measurements. This does not just dampen the extreme measurements, but also dampens and amplifies the difference between extreme and small measurements respectively, thus removing the positive skew (compare Figs. 1b and 1c). Compared to the GP realization, the ℓGP one thus exhibits higher spatial variability within hotspots but lower variability in the rest of the field. This intuitively explains why wide-area coverage suffices for GP but hotspot sampling is further needed for ℓGP.

Policy π^1 is *adaptive*: observe that each stagewise reward depends on the previously sampled data d_i:

$$\mathbb{H}[\mathbf{Y}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i] = \log\sqrt{(2\pi e)^k|\Sigma_{\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i}|} + \boldsymbol{\mu}_{\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i}\mathbf{1}^\top \quad (10)$$

where $\boldsymbol{\mu}_{\mathbf{Z}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i}$ is a mean vector with components $\mu_{Z_x|d_i}$ for x of $\tau(\mathbf{x}_i,\mathbf{a}_i)$. Since $\mu_{Z_x|d_i}$ depends on d_i by (7), $\mathbb{H}[\mathbf{Y}_{\tau(\mathbf{x}_i,\mathbf{a}_i)}|d_i]$ depends on d_i. Consequently, it follows from (6) that $U_i^{\pi^1}(d_i)$ and $\pi_i^1(d_i)$ depend on d_i for $i = 0,\ldots,n-1$. Hence, π^1 is *adaptive*.

Policy π^1 performs both *hotspot sampling* and *wide-area coverage*: to maximize stagewise rewards (10), π^1 selects new locations with large Gaussian posterior variance and mean for observation. So, it directs exploration towards sparsely sampled areas and hotspots.

Value-Function Approximations

Strictly Adaptive Exploration. With a team of $k > 1$ robots, π^1 collects $k > 1$ observations per stage, thus becoming *partially adaptive*. We will now derive the optimal *strictly adaptive* policy (in particular, for sampling ℓGP), which, among policies of all adaptivity, selects paths with the largest expected entropy. By Def. 2, a strictly adaptive policy has to be structured to collect only 1 observation per stage. To achieve strict adaptivity, iMASP(1) (6) can be revised as follows: (a) The space $\mathcal{A}(\mathbf{x}_i)$ of simultaneous joint actions is reduced to a constrained set $\mathcal{A}'(\mathbf{x}_i)$ of joint actions that allows one robot to move to observe a new location and the other robots stay put. This tradeoff for strict adaptivity allows $\mathcal{A}'(\mathbf{x}_i)$ to grow linearly, rather than exponentially, with the number of robots; (b) We constrain each robot to explore a path of at most n new adjacent locations; this can be viewed as an energy consumption constraint on each robot. The horizon then spans $k \times n$, rather than n, stages, which reflects the additional time of exploration incurred by strict adaptivity; (c) If $\mathbf{a}_i \in \mathcal{A}'(\mathbf{x}_i)$, the assignment $\mathbf{x}_{i+1} \leftarrow \tau(\mathbf{x}_i,\mathbf{a}_i)$ moves one chosen robot to a new location x_{i+1} while the other unselected robots stay put at their current locations. Then, only one component of $\mathbf{x}_i$ is changed to x_{i+1} to form $\mathbf{x}_{i+1}$; the other components of $\mathbf{x}_{i+1}$ are unchanged from $\mathbf{x}_i$. Hence, there is only one unobserved component $Y_{x_{i+1}}$ in $\mathbf{Y}_{\mathbf{x}_{i+1}}$; the other components of $\mathbf{Y}_{\mathbf{x}_{i+1}}$ are already observed in the previous stages and can be found in d_i. As a result, the probability distribution of $\mathbf{Y}_{\mathbf{x}_{i+1}}$ can be simplified to a univariate $Y_{x_{i+1}}$.

These revisions of iMASP(1) yield the strictly adaptive exploration problem called iMASP($\frac{1}{k}$):

$$\begin{aligned} U_i(d_i) &= \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \int f(y_{x_{i+1}}|d_i)\, U_{i+1}(d_{i+1})\, \mathrm{d}y_{x_{i+1}} \\ &= \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \mathbb{E}[U_{i+1}(d_i, x_{i+1}, Y_{x_{i+1}})|d_i] \\ U_t(d_t) &= \max_{\mathbf{a}_t\in\mathcal{A}'(\mathbf{x}_t)} \mathbb{H}[Y_{x_{t+1}}|d_t] \end{aligned} \quad (11)$$

for stage $i = 0,\ldots,t-1$ where $t = kn-1$. Without ambiguity, we omit the superscript $\pi^{\frac{1}{k}}$ (i.e., the optimal strictly

adaptive policy) from the optimal value functions above.

Since $Y_{x_{i+1}}$ is continuous, it entails infinite state transitions. So, $\mathbb{E}[U_{i+1}(d_i, x_{i+1}, Y_{x_{i+1}})|d_i]$ has to be evaluated in closed form for iMASP($\frac{1}{k}$) to be solved exactly. This can be performed for $t = 1$. When $t > 1$, the expectation of the optimal value function results in an integral that is too complex to be evaluated. Hence, we will resort to approximating iMASP($\frac{1}{k}$) as described below. For ease of exposition, we will revert to using $Z_{x_{i+1}} = \log Y_{x_{i+1}}$ for ℓGP from now on.

Approximately Optimal Exploration. To approximate iMASP($\frac{1}{k}$), we will first approximate the expectation in (11) from below and above using the ν-fold generalized Jensen and Edmundson-Madansky (EM) bounds respectively (Huang, Ziemba, and Ben-Tal 1977). To do this, we need the following convexity result for iMASP($\frac{1}{k}$) (11):

Lemma 5 *$U_i(d_i)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ for $i = 0, \ldots, t$.*

Let the support of $Z_{x_{i+1}}$ given d_i be $\mathcal{Z}^{\nu}_{x_{i+1}}$ that is partitioned into ν disjoint intervals $\mathcal{Z}^{[j]}_{x_{i+1}} = [\underline{z}^{[j-1]}_{x_{i+1}}, \overline{z}^{[j]}_{x_{i+1}}]$ for $j = 1, \ldots, \nu$. Then,

$$\sum_{j=1}^{\nu} \underline{p}^{[j]}_{x_{i+1}} U_{i+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}}) \leq \mathbb{E}[U_{i+1}(d_i, x_{i+1}, Z_{x_{i+1}})|d_i] \leq \sum_{j=0}^{\nu} \overline{p}^{[j]}_{x_{i+1}} U_{i+1}(d_i, x_{i+1}, \overline{z}^{[j]}_{x_{i+1}}) \quad (12)$$

where $\underline{p}^{[j]}_{x_{i+1}} \triangleq \int_{\mathcal{Z}^{[j]}_{x_{i+1}}} f(z_{x_{i+1}}|d_i)\, \mathrm{d}z_{x_{i+1}}$ and

$\underline{z}^{[j]}_{x_{i+1}} \triangleq \frac{1}{\underline{p}^{[j]}_{x_{i+1}}} \int_{\mathcal{Z}^{[j]}_{x_{i+1}}} z_{x_{i+1}} f(z_{x_{i+1}}|d_i)\, \mathrm{d}z_{x_{i+1}}$ for $j = 1, \ldots, \nu$,

$\overline{p}^{[j]}_{x_{i+1}} \triangleq \underline{p}^{[j]}_{x_{i+1}} \frac{\underline{z}^{[j]}_{x_{i+1}} - \overline{z}^{[j-1]}_{x_{i+1}}}{\overline{z}^{[j]}_{x_{i+1}} - \overline{z}^{[j-1]}_{x_{i+1}}} + \underline{p}^{[j+1]}_{x_{i+1}} \frac{\overline{z}^{[j+1]}_{x_{i+1}} - \underline{z}^{[j+1]}_{x_{i+1}}}{\overline{z}^{[j+1]}_{x_{i+1}} - \overline{z}^{[j]}_{x_{i+1}}}$ for $j = 0, \ldots, \nu$, and $\underline{p}^{[0]}_{x_{i+1}} := \underline{p}^{[\nu+1]}_{x_{i+1}} := \underline{z}^{[0]}_{x_{i+1}} := \underline{z}^{[\nu+1]}_{x_{i+1}} := \overline{z}^{[-1]}_{x_{i+1}} := 0$. By increasing ν to refine the partition, the bounds can be improved.

The upper approximate problem $\bar{i}$MASP($\frac{1}{k}$) can be constructed from iMASP($\frac{1}{k}$) (11) by replacing the expectation with the upper EM bound (12) to yield the optimal value functions $\overline{U}^{\nu}_i(d_i)$ for $i = 0, \ldots, t$. Similarly, the lower approximate problem $\underline{i}$MASP($\frac{1}{k}$) can be constructed from iMASP($\frac{1}{k}$) (11) by replacing the expectation with the lower Jensen bound (12) to yield the optimal value functions $\underline{U}^{\nu}_i(d_i)$ for $i = 0, \ldots, t$ and optimal policy $\underline{\pi}^{\frac{1}{k}}$.

The next result uses the induced optimal values from solving the lower and upper approximate problems to monotonically bound the maximum expected entropy achieved by the optimal strictly adaptive policy $\pi^{\frac{1}{k}}$:

Theorem 6 *If $\mathcal{Z}^{\nu+1}_{x_{i+1}}$ is obtained by splitting one of the intervals in $\mathcal{Z}^{\nu}_{x_{i+1}}$, $\underline{U}^{\nu}_i(d_i) \leq \underline{U}^{\nu+1}_i(d_i) \leq U_i(d_i) \leq \overline{U}^{\nu+1}_i(d_i) \leq \overline{U}^{\nu}_i(d_i)$ for $i = 0, \ldots, t$.*

A previous result of Low, Dolan, and Khosla (2008) has guaranteed that $\underline{\pi}^{\frac{1}{k}}$ can achieve an expected entropy not worse than $\underline{U}^{\nu}_0(d_0)$. But, that result does not account for how much it differs from the maximum expected entropy achieved by $\pi^{\frac{1}{k}}$. With the upper bound of Theorem 6, this error difference can be bounded:

Corollary 7 *$\underline{\pi}^{\frac{1}{k}}$ is guaranteed to achieve an expected entropy that is not more than $\overline{U}^{\nu}_0(d_0) - \underline{U}^{\nu}_0(d_0)$ from the maximum expected entropy $U_0(d_0)$ achieved by $\pi^{\frac{1}{k}}$.*

Bounds on Performance Advantage of Adaptive Exploration. A previous result of Low, Dolan, and Khosla (2008) has established the performance advantage of optimal adaptive over non-adaptive policies. Realizing the extent of such an advantage is important if adaptivity incurs a cost. In particular, we are interested in quantifying the performance difference between the strictly adaptive $\pi^{\frac{1}{k}}$ and the non-adaptive π^n. This performance advantage of $\pi^{\frac{1}{k}}$ over π^n is defined as the difference of their achieved maximum expected entropies $U_0(d_0) - U_0^{\pi^n}(d_0)$. Using the induced optimal values from solving the approximate problems (Theorem 6), the advantage $U_0(d_0) - U_0^{\pi^n}(d_0)$ can be bounded between $\underline{U}^{\nu}_0(d_0) - U_0^{\pi^n}(d_0)$ and $\overline{U}^{\nu}_0(d_0) - U_0^{\pi^n}(d_0)$. A large lower bound $\underline{U}^{\nu}_0(d_0) - U_0^{\pi^n}(d_0)$ implies $\pi^{\frac{1}{k}}$ is to be preferred. A small upper bound $\overline{U}^{\nu}_0(d_0) - U_0^{\pi^n}(d_0)$ implies π^n performs close to that of $\pi^{\frac{1}{k}}$ and should be preferred if it is more costly to deploy $\pi^{\frac{1}{k}}$. For GP, this advantage is zero as $\pi^{\frac{1}{k}}$ can be reduced to be non-adaptive as shown previously.

Real-Time Dynamic Programming. For our bounding approximation scheme, the state size grows exponentially with the number of stages. This is due to the nature of dynamic programming problems (e.g., $\underline{i}$MASP($\frac{1}{k}$)), which takes into account all possible states. To alleviate this computational difficulty, we modify the anytime algorithm URTDP of Low, Dolan, and Khosla (2008) based on $\underline{i}$MASP($\frac{1}{k}$), which can guarantee its policy performance in real time. It simulates greedy exploration paths through a large state space, resulting in desirable properties of focused search and good anytime behavior. The greedy exploration is guided by computationally efficient, informed initial heuristic bounds independent of state size.

In URTDP (Algorithm 1), each simulated path involves an alternating selection of actions and their corresponding outcomes till the last stage. Each action is selected based on the upper bound (line 3). For each encountered state, the algorithm maintains both lower and upper bounds, which are used to derive the uncertainty of its corresponding optimal value function. It exploits them to guide future searches in an informed manner; it explores the next state/outcome with the greatest amount of uncertainty (lines 4-5). Then, the algorithm backtracks up the path to update the upper heuristic bounds using $\max_{\mathbf{a}_i} \overline{Q}_i(\mathbf{a}_i, d_i)$ (line 11) where

$$\overline{Q}_i(\mathbf{a}_i, d_i) \triangleq \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu} \underline{p}^{[j]}_{x_{i+1}} \overline{U}_{i+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})$$

and the lower bounds via $\max_{\mathbf{a}_i} \underline{Q}_i(\mathbf{a}_i, d_i)$ (line 12) where

$$\underline{Q}_i(\mathbf{a}_i, d_i) \triangleq \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu} \underline{p}^{[j]}_{x_{i+1}} \underline{U}_{i+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})\,.$$

When an exploration policy is requested, we provide the greedy policy induced by the lower bound. The policy performance has a similar guarantee to Corollary 7.

We will show that the time complexity of SIMULATED-PATH(d_0, t) is independent of map resolution but the same

procedure in (Low, Dolan, and Khosla 2008) is not. It is also less sensitive to increasing robot team size. Assuming no prior data and $|\mathcal{A}'(\mathbf{x}_i)| = \Delta$, the time needed to evaluate the stagewise rewards $\mathbb{H}[Y_{x_{i+1}}|d_i]$ for all Δ new locations x_{i+1} (i.e., using Cholesky factorization) is $\mathcal{O}(t^3 + \Delta t^2)$, which is independent of $|\mathcal{X}|$ and results in $\mathcal{O}(t(t^3 + \Delta(t^2 + \nu)))$ time to run SIMULATED-PATH(d_0, t). In contrast, the time needed to evaluate the stagewise rewards in (Low, Dolan, and Khosla 2008) is $\mathcal{O}(t^3+\Delta(t^2+|\mathcal{X}|t)+|\mathcal{X}|t^2)$, which depends on $|\mathcal{X}|$ and entails $\mathcal{O}(t(t^3+\Delta(t^2+|\mathcal{X}|t+\nu)+|\mathcal{X}|t^2))$ time to run the same procedure. When the joint action set size Δ increases with larger robot team size, the time to run the procedure in (Low, Dolan, and Khosla 2008) increases faster than that of ours due to the gradient factor $|\mathcal{X}|t$ involving large domain size. In the next section, we will report the time taken to run this procedure empirically.

URTDP(d_0, t):

while $\overline{U}_0(d_0) - \underline{U}_0(d_0) > \alpha$ **do** SIMULATED-PATH(d_0, t)

SIMULATED-PATH(d_0, t):

1: $i \leftarrow 0$
2: **while** $i < t$ **do**
3: $\quad \mathbf{a}_i^* \leftarrow \arg\max_{\mathbf{a}_i} \overline{Q}_i(\mathbf{a}_i, d_i)$
4: $\quad \forall j, \quad \Xi_j \leftarrow p^{[j]}_{x^*_{i+1}}\{\overline{U}_{i+1}(d_i, x^*_{i+1}, \underline{z}^{[j]}_{x^*_{i+1}}) - \underline{U}_{i+1}(d_i, x^*_{i+1}, \underline{z}^{[j]}_{x^*_{i+1}})\}$
5: $\quad z \leftarrow$ sample from distribution at points $\underline{z}^{[j]}_{x^*_{i+1}}$ of probability $\Xi_j / \sum_k \Xi_k$
6: $\quad d_{i+1} \leftarrow d_i, x^*_{i+1}, z$
7: $\quad i \leftarrow i + 1$
8: $\overline{U}_i(d_i) \leftarrow \max_{\mathbf{a}_i} \mathbb{H}[Y_{x_{i+1}}|d_i]$, $\underline{U}_i(d_i) \leftarrow \overline{U}_i(d_i)$
9: **while** $i > 0$ **do**
10: $\quad i \leftarrow i - 1$
11: $\quad \overline{U}_i(d_i) \leftarrow \max_{\mathbf{a}_i} \overline{Q}_i(\mathbf{a}_i, d_i)$
12: $\quad \underline{U}_i(d_i) \leftarrow \max_{\mathbf{a}_i} \underline{Q}_i(\mathbf{a}_i, d_i)$

Algorithm 1: URTDP (α is user-specified bound).

Experiments and Discussion

This section evaluates, empirically, the approximately optimal strictly adaptive policy $\underline{\pi}^{\frac{1}{k}}$ on 2 real-world datasets exhibiting positive skew: (a) June 2006 plankton density data (Fig. 2a) of Chesapeake Bay bounded within lat. $38.481 - 38.591$N and lon. $76.487 - 76.335$W, and (b) potassium distribution data (Fig. 2d) of Broom's Barn farm spanning 520m by 440m. Each region is discretized into a 14×12 grid of sampling units. Each unit x is, respectively, associated with (a) plankton density y_x (chl-a) in mg m^{-3}, and (b) potassium level y_x (K) in mg l^{-1}. Each region comprises, respectively, (a) $|\mathcal{X}| = 148$ and (b) $|\mathcal{X}| = 156$ such units. Using a team of 2 robots, each robot is tasked to explore 9 adjacent units in its path including its starting unit. If only 1 robot is used, it is placed, respectively, in (a) top and (b) bottom starting unit, and samples all 18 units. Each robot's actions are restricted to move to the front, left, or right unit. We use the data of 20 randomly selected units to learn the hyperparameters (i.e., mean and covariance structure) of GP and ℓGP through maximum likelihood estimation (Rasmussen and Williams 2006). So, prior data d_0 comprise the randomly selected and robot starting units.

The performance of $\underline{\pi}^{\frac{1}{k}}$ is compared to the policies produced by four state-of-the-art exploration strategies: The *optimal non-adaptive policy π^n for GP* (Shewry and Wynn 1987) is produced by solving iMASP(n) (5). Similar to Theorem 4, it can be shown to be equivalent to the strictly adaptive $\pi^{\frac{1}{k}}$ for GP. Although $\underline{i}$MASP($\frac{1}{k}$) and iMASP(n) can be solved exactly, their state size grows exponentially with the number of stages. To alleviate this computational difficulty, we use anytime heuristic search algorithms URTDP (Algorithm 1) and Learning Real-Time A* to, respectively, solve $\underline{i}$MASP($\frac{1}{k}$) and iMASP(n) approximately. The *adaptive greedy policy for ℓGP* repeatedly chooses a reward-maximizing action (i.e., by repeatedly solving iMASP($\frac{1}{k}$) with $t = 0$ in (11)) to form the paths. The *non-adaptive greedy policy for GP* performs likewise but does it in the log-scale. In contrast to the above policies that optimize the entropy criterion (1), a non-adaptive greedy policy is proposed by Guestrin, Krause, and Singh (2005) to approximately maximize the mutual information (MI) criterion for GP; it repeatedly selects a new sampling location that maximizes the increase in MI. We call this the *MI-based policy*.

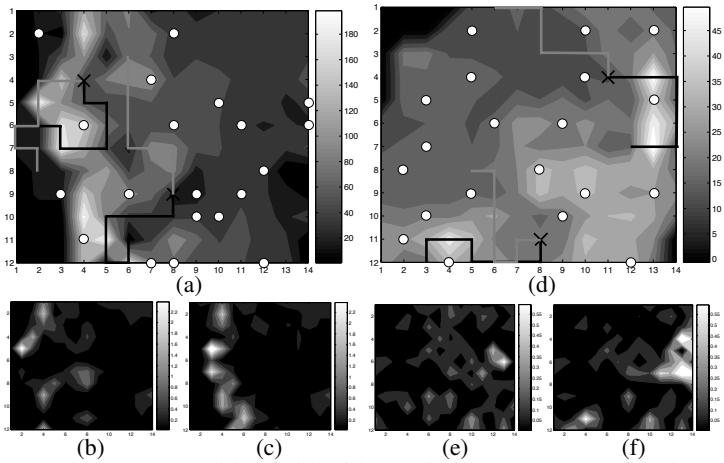

Figure 2: (a) chl-a field with prediction error maps for (b) strictly adaptive $\underline{\pi}^{1/k}$ and (c) non-adaptive π^n: 20 units (white circles) are randomly selected as prior data. The robots start at locations marked by '$\times$'s. The black and gray robot paths are produced by $\underline{\pi}^{1/k}$ and π^n respectively. (d-f) K field with error maps for $\underline{\pi}^{1/k}$ and π^n.

Performance metrics. Two metrics are used to evaluate the above policies: (a) *Posterior map entropy* (ENT) $\mathbb{H}[\mathbf{Y}_{\overline{\mathbf{x}}_{0:t}}|d_t]$ of domain $\mathcal{X}$ is the optimized criterion (1) measuring the posterior joint entropy of the original measurements $\mathbf{Y}_{\overline{\mathbf{x}}_{0:t}}$ at the unobserved locations $\overline{\mathbf{x}}_{0:t}$ where $t = 16$ (17) for the case of 2 (1) robots. A smaller ENT implies lower map uncertainty; (b) *Mean-squared relative error* (ERR) $|\mathcal{X}|^{-1}\sum_{x\in\mathcal{X}}\{(y_x-\mu_{Y_x|d_t})/\bar{\mu}\}^2$ measures the posterior map error from using the best unbiased predictor $\mu_{Y_x|d_t}$ (i.e., ℓGP posterior mean) (Low, Dolan, and Khosla 2008) of the measurement y_x to predict the hotspot field where $\bar{\mu} = |\mathcal{X}|^{-1}\sum_{x\in\mathcal{X}} y_x$. Although this criterion is not the one being optimized, it allows the use of ground truth measurements to evaluate if the field is being mapped accurately. A smaller ERR implies lower map prediction error.

Table 1 shows the results of various policies with different

Table 1: Performance comparison of information-theoretic policies for chl-a and K fields: 1R (2R) denotes 1 (2) robots.

Plankton density (chl-a) field		ENT		ERR	
Exploration policy	Model	1R	2R	1R	2R
Adaptive $\underline{\pi}^{1/k}$	ℓGP	381.37	376.19	0.1827	0.2319
Adaptive greedy	ℓGP	382.97	383.55	0.2919	0.2579
Non-adaptive π^n	GP	390.62	399.63	0.4145	0.3194
Non-adaptive greedy	GP	392.35	392.51	0.2994	0.3356
MI-based	GP	395.37	397.02	0.2764	0.2706
Potassium (K) field		ENT		ERR	
Exploration policy	Model	1R	2R	1R	2R
Adaptive $\underline{\pi}^{1/k}$	ℓGP	47.330	48.287	0.0299	0.0213
Adaptive greedy	ℓGP	61.080	56.181	0.0457	0.0302
Non-adaptive π^n	GP	67.084	59.318	0.0434	0.0358
Non-adaptive greedy	GP	58.704	64.186	0.0431	0.0335
MI-based	GP	59.058	67.390	0.0435	0.0343

assumed models and robot team sizes for chl-a and K fields. For $\underline{i}$MASP($\frac{1}{k}$) and iMASP(n), the results are obtained using the policies provided by the anytime algorithms after running 120000 simulated paths. The differences in results between policies have been verified using t-tests ($\alpha = 0.1$) to be statistically significant.

Plankton density data. The results show that the strictly adaptive $\underline{\pi}^{\frac{1}{k}}$ achieves lowest ENT and ERR as compared to the tested policies. From Fig. 2a, $\underline{\pi}^{\frac{1}{k}}$ moves the robots to sample the hotspots showing higher spatial variability whereas π^n moves them to sparsely sampled areas. Figs. 2b and 2c show, respectively, the prediction error maps resulting from $\underline{\pi}^{\frac{1}{k}}$ and π^n; the prediction error at each location x is measured using $|y_x - \mu_{Y_x|d_t}|/\bar{\mu}$. Locations with large errors are mostly concentrated in the left region where the field is highly-varying and contains higher measurements. Compared to $\underline{\pi}^{\frac{1}{k}}$, π^n incurs large errors at more locations in or close to hotspots, thus resulting in higher ERR.

We also compare the time needed to run the first 10000 SIMULATED-PATH(d_0, t)'s of our URTDP algorithm to that of Low, Dolan, and Khosla (2008), which are 115s and 10340s respectively for 2 robots (i.e., $90\times$ faster). They, respectively, take 66s and 2835s for 1 robot (i.e., $43\times$ faster). So, scaling to 2 robots incurs $1.73\times$ and $3.65\times$ more time for the respective algorithms. Policy $\underline{\pi}^{\frac{1}{k}}$ can already achieve the performance reported in Table 1 for 2 robots, and ENT of 389.23 and ERR of 0.231 for 1 robot. In contrast, the policy of Low, Dolan, and Khosla (2008) only improves to ENT of 377.82 (391.85) and ERR of 0.233 (0.252) for 2 (1) robots, which are slightly worse off.

Potassium distribution data. The results show again that $\underline{\pi}^{\frac{1}{k}}$ achieves lowest ENT and ERR. From Fig. 2d, $\underline{\pi}^{\frac{1}{k}}$ again moves the robots to sample the hotspots showing higher spatial variability whereas π^n moves them to sparsely sampled areas. Compared to $\underline{\pi}^{\frac{1}{k}}$, π^n incurs large errors at a greater number of locations in or close to hotspots as shown in Figs. 2e and 2f, thus resulting in higher ERR.

To run 10000 SIMULATED-PATH(d_0, t)'s, our URTDP algorithm is $84\times$ ($48\times$) faster than that of Low, Dolan, and Khosla (2008) for 2 (1) robots. Scaling to 2 robots incurs $1.93\times$ and $3.37\times$ more time for the respective algorithms. Policy $\underline{\pi}^{\frac{1}{k}}$ can already achieve the performance reported in Table 1 for 1 and 2 robots. In contrast, the policy of Low, Dolan, and Khosla (2008) achieves worse ENT of 67.132 (55.015) for 2 (1) robots. It achieves worse ERR of 0.032 for 2 robots but better ERR of 0.025 for 1 robot.

Summary of test results. The above results show that the strictly adaptive $\underline{\pi}^{\frac{1}{k}}$ can learn the highest-quality hotspot field map (i.e., lowest ENT and ERR) among the tested state-of-the-art strategies. After evaluating whether MASP- vs. iMASP-based planners are time-efficient for real-time deployment, we observe that $\underline{\pi}^{\frac{1}{k}}$ can achieve mapping performance comparable to the policy of Low, Dolan, and Khosla (2008) using significantly less time, and the incurred planning time is also less sensitive to larger robot team size. Lastly, we see in Fig. 2 that the strictly adaptive $\underline{\pi}^{\frac{1}{k}}$ has exploited clustering phenomena (i.e., hotspots) to achieve lower ENT and ERR than that of the non-adaptive π^n.

Conclusion

This paper describes an information-theoretic approach to efficient adaptive path planning for active exploration and mapping of hotspot fields. We have shown that, like MASP, iMASP is capable of exploiting clustering phenomena to produce lower map uncertainty. In contrast to MASP, the time complexity of solving (reward-maximizing) iMASP approximately is independent of map resolution and is also less sensitive to increasing robot team size as demonstrated theoretically and empirically. This is clearly advantageous in large-scale, high-resolution exploration and mapping. The proposed approximation techniques can be generalized to solve iMASPs that utilize the full joint action space of the robot team, thus allowing the robots to move simultaneously at every stage and the mission time to be constrained.

Acknowledgments. We would like to thank Dr R. Webster from Rothamsted Research for providing the Broom's Barn Farm data.

References

Guestrin, C.; Krause, A.; and Singh, A. P. 2005. Near-optimal sensor placements in Gaussian processes. In *Proc. ICML*.

Huang, C. C.; Ziemba, W. T.; and Ben-Tal, A. 1977. Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Oper. Res.* 25:315–325.

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

Meliou, A.; Krause, A.; Guestrin, C.; Kaiser, W.; and Hellerstein, J. M. 2007. Nonmyopic informative path planning in spatio-temporal models. In *Proc. AAAI*, 602–607.

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

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

Singh, A.; Krause, A.; Guestrin, C.; Kaiser, W.; and Batalin, M. 2007. Efficient planning of informative paths for multiple robots. In *Proc. IJCAI*, 2204–2211.

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

Proofs

Theorem 3

Proof by induction on i that $V_i^{\pi^1}(d_i) = \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i}}|d_i] - U_i^{\pi^1}(d_i)$ for $i = n-1, \ldots, 0$.

Base case ($i = n-1$):

$$
\begin{aligned}
& V_{n-1}^{\pi^1}(d_{n-1}) \\
&= \min_{\mathbf{a}_{n-1} \in \mathcal{A}(\mathbf{x}_{n-1})} \int f(\mathbf{z}_{\mathbf{x}_n}|d_{n-1})\, V_n^{\pi^1}(d_n)\, \mathrm{d}\mathbf{z}_{\mathbf{x}_n} \\
&= \min_{\mathbf{a}_{n-1} \in \mathcal{A}(\mathbf{x}_{n-1})} \int f(\mathbf{z}_{\mathbf{x}_n}|d_{n-1})\, \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n}}|d_n]\, \mathrm{d}\mathbf{z}_{\mathbf{x}_n} \\
&= \min_{\mathbf{a}_{n-1} \in \mathcal{A}(\mathbf{x}_{n-1})} \mathbb{H}[\mathbf{Z}_{\mathbf{x}_n}, \mathbf{Z}_{\overline{\mathbf{x}}_{0:n}} \mid d_{n-1}] - \mathbb{H}[\mathbf{Z}_{\mathbf{x}_n}|d_{n-1}] \\
&= \min_{\mathbf{a}_{n-1} \in \mathcal{A}(\mathbf{x}_{n-1})} \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n-1}}|d_{n-1}] - \mathbb{H}[\mathbf{Z}_{\mathbf{x}_n}|d_{n-1}] \\
&= \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n-1}}|d_{n-1}] - \max_{\mathbf{a}_{n-1} \in \mathcal{A}(\mathbf{x}_{n-1})} \mathbb{H}[\mathbf{Z}_{\mathbf{x}_n}|d_{n-1}] \\
&= \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:n-1}}|d_{n-1}] - U_{n-1}^{\pi^1}(d_{n-1})\,.
\end{aligned}
$$

The first and second equalities follow from (3). The third equality is due to the chain rule for entropy (Cover and Thomas 1991). The last equality is due to (6). Hence, the base case is true.

Inductive case: Suppose that

$$V_i^{\pi^1}(d_i) = \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i}}|d_i] - U_i^{\pi^1}(d_i) \tag{13}$$

is true. We have to prove that $V_{i-1}^{\pi^1}(d_{i-1}) = \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i-1}}|d_{i-1}] - U_{i-1}^{\pi^1}(d_{i-1})$ is true.

$$
\begin{aligned}
& V_{i-1}^{\pi^1}(d_{i-1}) \\
&= \min_{\mathbf{a}_{i-1} \in \mathcal{A}(\mathbf{x}_{i-1})} \int f(\mathbf{z}_{\mathbf{x}_i}|d_{i-1})\, V_i^{\pi^1}(d_i)\, \mathrm{d}\mathbf{z}_{\mathbf{x}_i} \\
&= \min_{\mathbf{a}_{i-1} \in \mathcal{A}(\mathbf{x}_{i-1})} \int f(\mathbf{z}_{\mathbf{x}_i}|d_{i-1}) \left(\mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i}}|d_i] - U_i^{\pi^1}(d_i) \right) \mathrm{d}\mathbf{z}_{\mathbf{x}_i} \\
&= \min_{\mathbf{a}_{i-1} \in \mathcal{A}(\mathbf{x}_{i-1})} \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i-1}}|d_{i-1}] - \mathbb{H}[\mathbf{Z}_{\mathbf{x}_i}|d_{i-1}] - \\
&\qquad \int f(\mathbf{z}_{\mathbf{x}_i}|d_{i-1})\, U_i^{\pi^1}(d_i)\, \mathrm{d}\mathbf{z}_{\mathbf{x}_i} \\
&= \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i-1}}|d_{i-1}] - \max_{\mathbf{a}_{i-1} \in \mathcal{A}(\mathbf{x}_{i-1})} \Big(\mathbb{H}[\mathbf{Z}_{\mathbf{x}_i}|d_{i-1}] + \\
&\quad \int f(\mathbf{z}_{\mathbf{x}_i}|d_{i-1})\, U_i^{\pi^1}(d_i)\, \mathrm{d}\mathbf{z}_{\mathbf{x}_i} \Big) \\
&= \mathbb{H}[\mathbf{Z}_{\overline{\mathbf{x}}_{0:i-1}}|d_{i-1}] - U_{i-1}^{\pi^1}(d_{i-1})\,.
\end{aligned}
$$

The first equality follows from (3). The second equality follows from (13). The third equality follows from linearity of expectation and the chain rule for entropy (Cover and Thomas 1991). The last equality is due to (6). Hence, the inductive case is true.

It is clear from above that the induced optimal adaptive policies from solving the cost-minimizing and reward-maximizing iMASP(1)'s coincide.

Equation 9

Since $f(\mathbf{Z}_{\mathbf{x}_{i+1}} = \mathbf{z}_{\mathbf{x}_{i+1}}|d_i) = \dfrac{\exp\left\{-\frac{1}{2}(\mathbf{z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}(\mathbf{z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})^\top\right\}}{\sqrt{(2\pi)^k\,|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|}}$,

$$
\begin{aligned}
& \mathbb{H}[\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i] \\
&= \mathbb{E}[-\log f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i)|d_i] \\
&= \mathbb{E}[\log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \\
&\quad \frac{1}{2}(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})^\top|d_i] \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \\
&\quad \frac{1}{2}\mathbb{E}[(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})^\top|d_i] \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \\
&\quad \frac{1}{2}\mathbb{E}[\mathrm{tr}(\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}))|d_i] \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \\
&\quad \frac{1}{2}\mathrm{tr}(\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}\mathbb{E}[(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})(\mathbf{Z}_{\mathbf{x}_{i+1}} - \mu_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i})|d_i]) \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \frac{1}{2}\mathrm{tr}(\Sigma^{-1}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}) \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \frac{1}{2}\mathrm{tr}(\mathbf{I}) \\
&= \log \sqrt{(2\pi)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \frac{k}{2} \\
&= \log \sqrt{(2\pi e)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|}\,.
\end{aligned}
$$

The fourth equality is due to the trace property $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

Equation 10

Using the Jacobian method of variable transformation,

$$
\begin{aligned}
f(\mathbf{Y}_{\mathbf{x}_{i+1}}|d_i) = \;& f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i) \prod_{x \in \mathcal{X}'} \frac{dZ_x}{dY_x} \\
= \;& f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i) \prod_{x \in \mathcal{X}'} \frac{1}{Y_x}
\end{aligned}
$$

where $\mathcal{X}' = \{x \mid x \text{ is a location component in } \mathbf{x}_{i+1}\}$. So,

$$
\begin{aligned}
& \mathbb{H}[\mathbf{Y}_{\mathbf{x}_{i+1}}|d_i] \\
&= \mathbb{E}[-\log f(\mathbf{Y}_{\mathbf{x}_{i+1}}|d_i)|d_i] \\
&= \mathbb{E}[-\log \left(f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i) \prod_{x \in \mathcal{X}'} \frac{1}{Y_x} \right) |d_i] \\
&= \mathbb{E}[-\log f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i) + \sum_{x \in \mathcal{X}'} \log Y_x|d_i] \\
&= \mathbb{E}[-\log f(\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i)|d_i] + \sum_{x \in \mathcal{X}'} \mathbb{E}[Z_x|d_i] \\
&= \log \sqrt{(2\pi e)^k|\Sigma_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}|} + \boldsymbol{\mu}_{\mathbf{Z}_{\mathbf{x}_{i+1}}|d_i}\mathbf{1}^\top\,.
\end{aligned}
$$

The fourth equality is due to the transformation $Z_x = \log Y_x$ and linearity of expectation. The fifth equality follows from (9).

Lemma 5

We first show that $\mathbb{H}[Y_{x_{i+1}}|d_i]$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ for $i = 0, \ldots, t$. From (10), we know that

$$\mathbb{H}[Y_{x_{i+1}}|d_i] = \log \sqrt{2\pi e \sigma^2_{Z_{x_{i+1}}|d_i}} + \mu_{Z_{x_{i+1}}|d_i}\,.$$

From (7), the posterior mean $\mu_{Z_{x_{i+1}}|d_i}$ is an affine function of $\mathbf{z}_{\mathbf{x}_{0:i}}$. Hence, it is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ ((Boyd and Vandenberghe 2004), pp. 71). From (8), the posterior variance $\sigma^2_{Z_{x_{i+1}}|d_i}$ is independent of $\mathbf{z}_{\mathbf{x}_{0:i}}$. So, $\log\sqrt{2\pi e\sigma^2_{Z_{x_{i+1}}|d_i}}$ is a constant term. Therefore, $\mathbb{H}[Y_{x_{i+1}}|d_i]$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$.

We will revert to using $Z_{x_{i+1}}$ in iMASP($\frac{1}{k}$) (11) for ℓGP (i.e., by transforming $Z_{x_{i+1}} = \log Y_{x_{i+1}}$).

Proof by induction on i that $U_i(d_i)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ for $i = t, \ldots, 0$.

Base case ($i = t$): As proven above, $\mathbb{H}[Y_{x_{t+1}}|d_t]$ is convex in $\mathbf{z}_{\mathbf{x}_{0:t}}$. Then, the pointwise maximum of $\mathbb{H}[Y_{x_{t+1}}|d_t]$ (i.e., $\max_{\mathbf{a}_t\in\mathcal{A}'(\mathbf{x}_t)}\mathbb{H}[Y_{x_{t+1}}|d_t]$) is convex in $\mathbf{z}_{\mathbf{x}_{0:t}}$ ((Boyd and Vandenberghe 2004), pp. 81). Therefore, $U_t(d_t)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:t}}$. The base case is true.

Inductive case: Suppose that $U_{i+1}(d_{i+1})$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i+1}}$. We have to prove that $U_i(d_i)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$.

From (11), the expectation under the normal variable $Z_{x_{i+1}}$ with posterior mean $\mu_{Z_{x_{i+1}}|d_i}$ and variance $\sigma^2_{Z_{x_{i+1}}|d_i}$ can be expressed in terms of the standard normal variable $Z = (Z_{x_{i+1}} - \mu_{Z_{x_{i+1}}|d_i})/\sigma^2_{Z_{x_{i+1}}|d_i}$:

$$\int f(Z_{x_{i+1}} = z_{x_{i+1}}|d_i)\; U_{i+1}(d_i, x_{i+1}, z_{x_{i+1}})\;\mathrm{d}z_{x_{i+1}} = \int f(z)\; U_{i+1}(d_i, x_{i+1}, \mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z)\;\mathrm{d}z\;.$$

Since d_i and $\mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z$ are affine in $\mathbf{z}_{\mathbf{x}_{0:i}}$ and $U_{i+1}(d_{i+1})$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i+1}}$ by assumption, $U_{i+1}(d_i, x_{i+1}, \mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ because vector composition operation preserves convexity[1] ((Boyd and Vandenberghe 2004), pp. 86). Since $U_{i+1}(d_i, x_{i+1}, \mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ for each z, $\int f(z)\; U_{i+1}(d_i, x_{i+1}, \mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z)\;\mathrm{d}z$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ because integration preserves convexity ((Boyd and Vandenberghe 2004), pp. 79). So, $\int f(z_{x_{i+1}}|d_i)\; U_{i+1}(d_i, x_{i+1}, z_{x_{i+1}})\;\mathrm{d}z_{x_{i+1}}$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$. From above, $\mathbb{H}[Y_{x_{i+1}}|d_i]$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$. Then, the pointwise maximum of $\mathbb{H}[Y_{x_{i+1}}|d_i] + \int f(z_{x_{i+1}}|d_i)\; U_{i+1}(d_i, x_{i+1}, z_{x_{i+1}})\;\mathrm{d}z_{x_{i+1}}$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$. Therefore, $U_i(d_i)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$. The inductive case is true.

Theorem 6

Proof by induction on i that $\underline{U}_i^{\nu}(d_i) \le \underline{U}_i^{\nu+1}(d_i) \le U_i(d_i)$ for $i = t, \ldots, 0$.

Base case ($i = t$): $\underline{U}_t^{\nu}(d_t) = \underline{U}_t^{\nu+1}(d_t) = U_t(d_t) = \max_{\mathbf{a}_t\in\mathcal{A}'(\mathbf{x}_t)}\mathbb{H}[Y_{x_{t+1}}|d_t]$. Hence, the base case is true.

Inductive case: Suppose that $\underline{U}_{i+1}^{\nu}(d_{i+1}) \le \underline{U}_{i+1}^{\nu+1}(d_{i+1}) \le U_{i+1}(d_{i+1})$ is true. We have to prove that $\underline{U}_i^{\nu}(d_i) \le \underline{U}_i^{\nu+1}(d_i) \le U_i(d_i)$ is true.

We will first show that $\underline{U}_i^{\nu+1}(d_i) \le U_i(d_i)$.

$$\begin{aligned}
&\underline{U}_i^{\nu+1}(d_i)\\
&= \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu+1} \underline{p}^{[j]}_{x_{i+1}}\, \underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})\\
&\le \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu+1} \underline{p}^{[j]}_{x_{i+1}}\, U_{i+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})\\
&\le \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \mathbb{E}[U_{i+1}(d_i, x_{i+1}, Z_{x_{i+1}})|d_i]\\
&= U_i(d_i)\;.
\end{aligned}$$

The first inequality follows from assumption (i.e., $\underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}}) \le U_{i+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})$). The second inequality follows from Lemma 5 that $U_{i+1}(d_i, x_{i+1}, z_{x_{i+1}})$ is convex in $z_{x_{i+1}}$ for ℓGP, and the generalized Jensen bound (12).

We will now prove that $\underline{U}_i^{\nu}(d_i) \le \underline{U}_i^{\nu+1}(d_i)$.

$$\begin{aligned}
&\underline{U}_i^{\nu}(d_i)\\
&= \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu} \underline{p}^{[j]}_{x_{i+1}}\, \underline{U}_{i+1}^{\nu}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})\\
&\le \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{j=1}^{\nu} \underline{p}^{[j]}_{x_{i+1}}\, \underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})\\
&\le \max_{\mathbf{a}_i\in\mathcal{A}'(\mathbf{x}_i)} \mathbb{H}[Y_{x_{i+1}}|d_i] + \sum_{\ell=1}^{\nu+1} \underline{p}^{[\ell]}_{x_{i+1}}\, \underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, \underline{z}^{[\ell]}_{x_{i+1}})\\
&= \underline{U}_i^{\nu+1}(d_i)\;.
\end{aligned}$$

The first inequality follows from assumption (i.e., $\underline{U}_{i+1}^{\nu}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}}) \le \underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, \underline{z}^{[j]}_{x_{i+1}})$). We need the result that $\underline{U}_i^{\nu+1}(d_i)$ is convex in $\mathbf{z}_{\mathbf{x}_{0:i}}$ for $i = 0, \ldots, t$ for the second inequality to hold. The proof[2] is similar to that of Lemma 5. Consequently, since $\underline{U}_{i+1}^{\nu+1}(d_i, x_{i+1}, z_{x_{i+1}})$ is convex in $z_{x_{i+1}}$ and $\mathcal{Z}^{\nu+1}_{x_{i+1}}$ is obtained by splitting one of the intervals in $\mathcal{Z}^{\nu}_{x_{i+1}}$, the second inequality results. The inductive case is thus true.

The proof of $U_i(d_i) \le \overline{U}_i^{\nu+1}(d_i) \le \overline{U}_i^{\nu}(d_i)$ for $i = t, \ldots, 0$ is similar to the above except that the inequalities are reversed.

References for Proofs

Boyd, S., and Vandenberghe L. 2004. *Convex Optimization*. Cambridge Univ. Press.

Cover T., and Thomas J. 1991. *Elements of Information Theory*. John Wiley & Sons.

[1] Note that $U_{i+1}(d_{i+1})$ does not have to be non-decreasing in each argument because d_i and $\mu_{Z_{x_{i+1}}|d_i} + \sigma^2_{Z_{x_{i+1}}|d_i} z$ are affine in $\mathbf{z}_{\mathbf{x}_{0:i}}$.

[2] The approximate problems $\underline{i}$MASP($\frac{1}{k}$) and $\overline{i}$MASP($\frac{1}{k}$) differ from iMASP($\frac{1}{k}$) (11) by the non-negative weighted sum (instead of the expectation), which also preserves convexity.