Performance Guarantees for Information Theoretic Active Inference

Jason L. Williams, John W. Fisher III and Alan S. Willsky
Laboratory for Information and Decision Systems and
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
jlwilliams@alum.mit.edu, fisher@csail.mit.edu, willsky@mit.edu

Abstract

In many estimation problems, the measurement process can be actively controlled to alter the information received. The control choices made in turn determine the performance that is possible in the underlying inference task. In this paper, we discuss performance guarantees for heuristic algorithms for adaptive measurement selection in sequential estimation problems, where the inference criterion is mutual information. We also demonstrate the performance of our tighter online computable performance guarantees through computational simulations.

1 Introduction

Active sensing is motivated by modern sensors which can be controlled to observe different aspects of an underlying probabilistic process. For example, if we use cameras to track people in buildings, we can steer the camera to focus or zoom on different people or places; in a sensor network, we can choose to activate and deactivate different nodes and different sensing modalities within a particular node; or in a medical diagnosis problem we can choose which tests to administer to a patient. In each of these cases, our control choices impact the information that we receive in our observation, and thus the performance achieved in the underlying inference task.

A commonly used performance objective in active sensing is mutual information (MI) (*e.g.*, [1]). Denoting the quantity that we are aiming to infer as X and the observation resulting from control choice u as z^u, the MI between X and z^u is defined as the expected reduction in the entropy produced by the observation [2], *i.e.*, $I(X; z^u) = H(X) - H(X|z^u) = H(z^u) - H(z^u|X)$.[1] Since $H(X)$ is independent of the control choice u, choosing u to maximize $I(X; z^u)$ is equivalent to minimizing the uncertainty in X as measured by the conditional entropy $H(X|z^u)$. We also refer to the MI performance objective (which we seek to maximize) as the *reward function*.

In different problems, the collection of subsets of observations from which one may choose can have a dramatically different structure. One common structure involves selection of any K-element subset of a set of observations, *e.g.*, the subset of sensors to activate in a sensor network application [3, 4]. Another structure is one in which there is a single sensor which can operate in one mode at each time increment; the resulting selection structure is one in which we may choose a single element from each of a series of observation sets, each of which corresponds to a different time instant [5, 6]. Application areas range from network fault detection and diagnosis [7] to environmental sensing [3] to object tracking [1, 5]. The greedy heuristic, which at each stage chooses the observation which maximizes the MI with X conditioned on the already selected observations, is used widely across the breadth of these applications [1, 3, 4, 5, 7].

Recent work [8] has applied results from [9] to establish that, when the selection structure is such that any subset of observations with cardinality $\leq K$ may be chosen, the greedy heuristic achieves a total MI of within a constant multiple $(1-1/e) \approx 0.632$ of the optimal subset of observations. Our analysis extends this to the larger class of problems involving sequential processes, providing the surprising result that in sequential problems, under quite general assumptions one may select the control for the current time instant neglecting future observation opportunities, and still have performance within a multiple 0.5 of the optimal. Furthermore, the online computable bounds demonstrated in Sections 2.4 and 3.3 can be significantly stronger in

[1] Note that when we condition on a random variable (such as a yet unrealized observation) the conditional entropy involves an expectation over the distribution of that random variable.

certain circumstances. Several new results relating to closed loop operation are presented in Section 4.

The guarantees we develop are based upon submodularity, the same property exploited in [8, 9, 11]. Submodularity captures the notion that as we select more observations, the value of the remaining unselected observations decreases, *i.e.*, the notion of diminishing returns.

Definition 1. *A set function f is submodular if $f(\mathcal{C} \cup \mathcal{A}) - f(\mathcal{A}) \geq f(\mathcal{C} \cup \mathcal{B}) - f(\mathcal{B}) \; \forall \; \mathcal{B} \supseteq \mathcal{A}$.*

It was established in [8] that, assuming that observations are independent conditioned on the quantity to be estimated (herein referred to as the *state*), MI is a submodular function of the observation selection set. In many applications the requirement for all observations to be independent conditioned on the state is not overly restrictive, since the definition of state may be expanded to include latent variables that provide the required conditional independence. The simple result that we will utilize from submodularity is that $I(x; z^{\mathcal{C}}|z^{\mathcal{A}}) \geq I(x; z^{\mathcal{C}}|z^{\mathcal{B}}) \; \forall \; \mathcal{B} \supseteq \mathcal{A}$.

2 A Simple Performance Guarantee

To commence, consider a simple sequential problem involving two time steps, where at each step we must choose a single observation (*e.g.*, in which mode to operate a sensor) from a different set of observations. The goal is to maximize the information obtained about an underlying quantity X. Let $\{o_1, o_2\}$ denote the optimal choice for the two stages, *i.e.*, that which maximizes $I(X; z_1^{u_1}, z_2^{u_2})$ over possible choices for $\{u_1, u_2\}$. Let $\{g_1, g_2\}$ denote the choice made by the greedy heuristic, where $g_1 = \arg\max_{u_1} I(X; z_1^{u_1})$ and $g_2 = \arg\max_{u_2} I(X; z_2^{u_2}|z_1^{g_1})$ (where conditioning is on the random variable $z_1^{g_1}$, not on the resulting observation value). Then the following analysis establishes a performance guarantee for the greedy algorithm:

$$\begin{aligned} I(X; z_1^{o_1}, z_2^{o_2}) &\overset{(a)}{\leq} I(X; z_1^{g_1}, z_2^{g_2}, z_1^{o_1}, z_2^{o_2}) \\ &\overset{(b)}{=} I(X; z_1^{g_1}) + I(X; z_2^{g_2}|z_1^{g_1}) \\ &\quad + I(X; z_1^{o_1}|z_1^{g_1}, z_2^{g_2}) \\ &\quad + I(x; z_2^{o_2}|z_1^{g_1}, z_2^{g_2}, z_1^{o_1}) \\ &\overset{(c)}{\leq} I(X; z_1^{g_1}) + I(X; z_2^{g_2}|z_1^{g_1}) \\ &\quad + I(X; z_1^{o_1}) + I(x; z_2^{o_2}|z_1^{g_1}) \\ &\overset{(d)}{\leq} 2I(X; z_1^{g_1}) + 2I(X; z_2^{g_2}|z_1^{g_1}) \\ &\overset{(e)}{=} 2I(X; z_1^{g_1}, z_2^{g_2}) \end{aligned} \tag{1}$$

where (a) results from the nondecreasing property of MI, (b) is an application of the MI chain rule, (c) results from submodularity (assuming that all observations are independent conditioned on X), (d) from the definition of the greedy heuristic, and (e) from a reverse application of the chain rule. Thus the optimal performance can be no more than twice that of the greedy heuristic, or, conversely, the performance of the greedy heuristic is at least half that of the optimal.[2]

Theorem 1 presents this result in its most general form; the proof directly follows the above steps. The following assumption establishes the basic structure: we have N sets of observations, and we can select a specified number of observations from each set in an arbitrary order.

Assumption 1. *There are N sets of observations, $\{\{z_1^1, \ldots, z_1^{n_1}\}, \; \{z_2^1, \ldots, z_2^{n_2}\}, \; \ldots, \; \{z_N^1, \ldots, z_N^{n_N}\}\}$, which are mutually independent conditioned on the quantity to be estimated (X). Any k_i observations can be chosen out of the i-th set ($\{z_i^1, \ldots, z_i^{n_i}\}$). The sequence $(w_1, \ldots, w_M)$ (where $w_i \in \{1, \ldots, N\} \; \forall \; i$) specifies the order in which we visit observation sets using the greedy heuristic (i.e., in the i-th stage we select a previously unselected observation out of the w_i-th set).*

The abstraction of the observation set sequence $(w_1, \ldots, w_M)$ allows us to visit observation sets more than once (allowing us to select multiple observations from each set) and in any order. The greedy heuristic operating on this structure is defined below, followed by the general form of the guarantee.

Definition 2. *The greedy heuristic operates according to the following rule:*

$$g_j = \underset{u \in \{1, \ldots, n_{w_j}\}}{\arg\max} \; I(X; z_{w_j}^u | z_{w_1}^{g_1}, \ldots, z_{w_{j-1}}^{g_{j-1}})$$

Theorem 1. *Under Assumption 1, the greedy heuristic in Definition 2 has performance guaranteed by the following expression:*

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_M}^{o_M}) \leq 2I(X; z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M})$$

where $\{z_{w_1}^{o_1}, \ldots, z_{w_M}^{o_M}\}$ is the optimal set of observations, i.e., the one which maximizes $I(X; z_{w_1}^{u_1}, \ldots, z_{w_M}^{u_M})$ over the possible choices for $\{u_1, \ldots, u_M\}$.

The proof of the theorem can be found in [10]; it directly follows the steps in Eq. (1).

2.1 Comparison to matroid guarantee

The prior work using matroids [11] provides another algorithm with the same guarantee for problems of this

[2] Note that this is considering only open loop control; we will discuss closed loop control in Section 4.

structure. However, to achieve the guarantee on matroids it is necessary to consider every observation at every stage of the problem. Computationally, it is far more desirable to be able to proceed in a dynamic system by selecting observations at time k considering only the observations available at that time, disregarding future time steps (indeed, countless previous works, such as [1], do just that). The freedom of choice of the order in which we visit observation sets in Theorem 1 extends the performance guarantee to this commonly used sequential selection structure.

2.2 Tightness of bound

The bound derived in Theorem 1 can be arbitrarily close to tight, as the following example shows.

Example 1. *Let $X = [a, b]^T$ where a and b are independent binary random variables with $P(a = 0) = P(a = 1) = 0.5$ and $P(b = 0) = 0.5 - \epsilon$; $P(b = 1) = 0.5 + \epsilon$ for some $\epsilon > 0$. We have two sets of observations with $n_1 = 2$, $n_2 = 1$ and $k_1 = k_2 = 1$. In the first set of observations we may measure $z_1^1 = a$ for reward $I(X; z_1^1) = H(a) = 1$, or $z_1^2 = b$ for reward $I(X; z_1^2) = H(b) = 1 - \delta(\epsilon)$, where $\delta(\epsilon) > 0 \ \forall \ \epsilon > 0$, and $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. At the second stage we have one choice, $z_2^1 = a$. Our walk is $w = (1, 2)$, i.e., we visit the first set of observations once, followed by the second set. The greedy algorithm selects at the first stage to observe $z_1^1 = a$, as it yields a higher reward (1) than $z_1^2 = b$ $(1 - \delta(\epsilon))$. At the second stage, the algorithm already has the exact value for a, hence the observation at the second stage yields zero reward. The total reward is 1. The optimal sequence selects observation $z_1^2 = b$ for reward $1 - \delta(\epsilon)$, and then gains a reward of 1 from the second observation z_2^1. The total reward is $2 - \delta(\epsilon)$. By choosing ϵ arbitrarily close to zero, we may make the ratio of optimal reward to greedy reward, $2 - \delta(\epsilon)$, arbitrarily close to 2.*

2.3 Online version of guarantee

Modifying step (c) of Eq. (1), we can also obtain an online performance guarantee, which will often be substantially tighter in practice (as demonstrated in Section 2.4). The online bound will tend to be tight in cases where the amount of information remaining after choosing the set of observations is small.

Theorem 2. *Under the same assumptions as Theorem 1, for each $i \in \{1, \ldots, N\}$ define $\bar{k}_i = \min\{k_i, n_i - k_i\}$, and for each $j \in \{1, \ldots, \bar{k}_i\}$ define*

$$\bar{g}_i^j = \underset{u \in \{1,\ldots,n_i\} - \{\bar{g}_i^l | l<j\}}{\arg\max} I(X; z_i^u | z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M}) \quad (2)$$

Then the following two performance guarantees, which are computable online, apply:

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_M}^{o_M}) \leq I(X; z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M}) + \sum_{i=1}^{N} \sum_{j=1}^{\bar{k}_i} I(X; z_i^{\bar{g}_i^j} | z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M}) \quad (3)$$

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_M}^{o_M}) \leq I(X; z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M}) + \sum_{i=1}^{N} \bar{k}_i I(X; z_i^{\bar{g}_i^1} | z_{w_1}^{g_1}, \ldots, z_{w_M}^{g_M}) \quad (4)$$

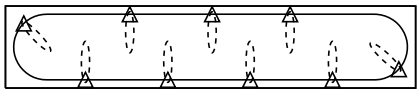
The proof of this result may be found in [10]. The quantity $\bar{g}_i^j$ represents the j-th next best observation in the i-th set, conditioned on the observation choices made in the greedy selection. Determining these quantities and their rewards requires no more reward evaluations than the initial calculation of the greedy choices. The reward values of these observations serve to bound the information remaining given the choice made by the greedy heuristic. The count $\bar{k}_i$ is an upper bound to the number of observations out of the i-th set that are in the optimal choice but not in the greedy choice. The online bound can be used to calculate an upper bound for the optimal reward starting from *any* sequence of observation choices, not just the choice made by the greedy heuristic in Definition 2, $(g_1, \ldots, g_M)$.

2.4 Example of online guarantee

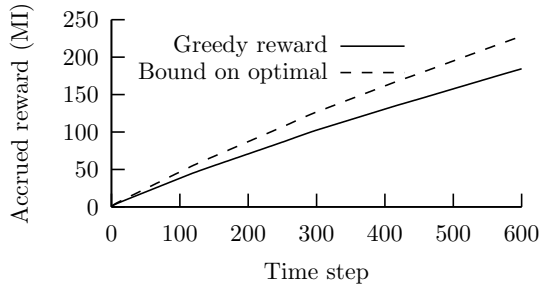
Suppose that we are using a surface vehicle travelling at a constant velocity along a fixed path (as illustrated in Fig. 1(a)) to map the depth of the ocean floor in a particular region. Assume that, at any position on the path (such as the points denoted by '△'), we may steer our sensor to measure the depth of any point within a given region around the current position (as depicted by the dotted ellipses), and that we receive a linear measurement of the depth corrupted by Gaussian noise with variance R. Suppose that we model the depth of the ocean floor as a Gauss-Markov random field with a 500×100 thin membrane grid model where neighboring node attractions are uniformly equal to q. One cycle of the vehicle path takes 300 time steps to complete. The number of observation choices available (*i.e.*, the number of cells inside the ellipses in Fig. 1(a)) varies between 380 and 400 due to changing geometry as the vehicle moves.

Defining the state X to be the vector containing one element for each cell in the 500×100 grid, the problem can be seen to fit into the structure of Assumption 1 (with $w_i = i$ and $k_i = 1 \ \forall \ i$). The selection algorithm simply selects at each stage the most informative observation conditioned on the observations previously chosen. A single observation of the cell directly beneath the sensing platform is used as initialization to obtain a full rank information matrix.

(a) Region boundary and vehicle path



(b) Reward of greedy heuristic and bound on optimal



(c) Factor of optimality from online guarantee

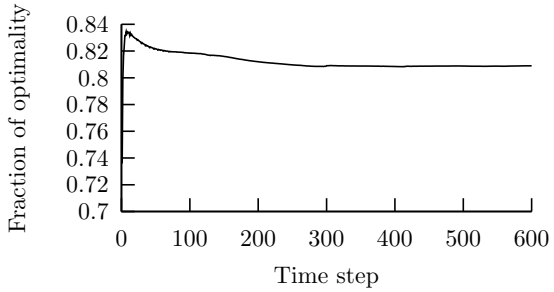


Figure 1: (a) shows region boundary and vehicle path (counter-clockwise, starting from the left end of the lower straight segment). When the vehicle is located at a '△' mark, any one grid element with center inside the surrounding dotted ellipse may be measured. (b) graphs reward accrued by the greedy heuristic after different periods of time, and the bound on the optimal sequence for the same time period. (c) shows the ratio of these two curves, providing the factor of optimality guaranteed by the bound.

Fig. 1(b) shows the accrual of reward over time as well as the bound on the optimal sequence obtained using Theorem 2 for each time step when $q = 100$ and $R = 1/40$, while Fig. 1(c) shows the ratio between the achieved performance and the optimal sequence bound over time. The graph indicates that the greedy heuristic achieves at least $0.8\times$ the optimal reward. In this case, finding the optimal plan for the 600 time slots through full enumeration would require on the order of 10^{1556} reward evaluations. The tightness of the online bound depends on particular model characteristics: if $q = R = 1$, then the guarantee ratio is much closer to the value of the offline bound (*i.e.*, 0.5).

2.5 Errors in rewards

The previous analysis assumes that all reward values can be calculated exactly. While this is possible for some common classes of problems (such as linear Gaussian models), approximations are often necessary; one method of approximately evaluating the MI terms for closed loop selection problems in complex networks is proposed in [7]. The analysis in [8] can be easily extended to the algorithms described in this paper. As an example, consider the proof of Theorem 1, where the greedy heuristic is used with estimated MI rewards,

$$g_j = \arg\max_{g \in \{1,\ldots,n_{w_j}\}} \hat{I}(X; z^g_{w_j} | z^{g_1}_{w_1}, \ldots, z^{g_{j-1}}_{w_{j-1}})$$

and the error in the MI estimate is bounded by ϵ, *i.e.*,

$$|I(X; z^g_{w_j} | z^{g_1}_{w_1}, \ldots, z^{g_{j-1}}_{w_{j-1}}) - \hat{I}(X; z^g_{w_j} | z^{g_1}_{w_1}, \ldots, z^{g_{j-1}}_{w_{j-1}})| \leq \epsilon$$

From the proof of Theorem 1, we can easily show that

$$I(X; z^{o_1}_{w_1}, \ldots, z^{o_M}_{w_M}) \leq 2I(X; z^{g_1}_{w_1}, \ldots, z^{g_M}_{w_M}) + 2M\epsilon$$

Hence the deterioration in the performance guarantee is at most $2M\epsilon$.

3 Exploiting diffusiveness

In problems such as object tracking, the kinematic quantities of interest evolve according to a diffusive process, in which correlation between states at different time instants reduces as the time difference increases. Intuitively, one would expect that a greedy algorithm would be closer to optimal in situations in which the diffusion strength is high. This section develops a performance guarantee which exploits the diffusiveness of the underlying process to obtain a tighter bound on performance.

The general form of the result, stated in Theorem 3, deals with an arbitrary graph in the latent structure. The simpler cases involving trees and chains are discussed in the sequel. The theorem is limited to only choosing a single observation from each set; the proof of Theorem 3 exploits this fact. The basic model structure is set up in Assumption 2.

Assumption 2. *Let the latent structure which we seek to infer consist of an undirected graph $\mathcal{G}$ with nodes $X = \{x_1, \ldots, x_L\}$, with an arbitrary interconnection structure. Assume that each node has a set of observations $\{z^1_i, \ldots, z^{n_i}_i\}$, which are independent of each other and all other nodes and observations in the graph conditioned on x_i. We may select a single observation from each set. Let $(w_1, \ldots, w_L)$ be a sequence which determines the order in which nodes are visited ($w_i \in \{1, \ldots, L\}\ \forall\ i$); we assume that each node is visited exactly once.*

The results of Section 2 were applicable to any submodular, nondecreasing objective for which the reward of an empty set was zero. In this section, we exploit an additional property of mutual information which holds under Assumption 2, that for any set of conditioning observations $z^{\mathcal{A}}$:

$$\begin{aligned} I(X; z_i^j | z^{\mathcal{A}}) &= H(z_i^j | z^{\mathcal{A}}) - H(z_i^j | X, z^{\mathcal{A}}) \\ &= H(z_i^j | z^{\mathcal{A}}) - H(z_i^j | x_i) \\ &= I(x_i; z_i^j | z^{\mathcal{A}}) \end{aligned} \tag{5}$$

We then utilize this property in order to exploit process diffusiveness. The general form of the diffusive characteristic is stated in Assumption 3. This is a strong assumption that is difficult to establish globally for any given model; in Section 3.1 we present an online computable guarantee which exploits the characteristic to whatever extent it exists in a particular selection problem. In Section 3.2 we then specialize the assumption to cases where the latent graph structure is a tree or a chain.

Assumption 3. *Under the structure in Assumption 2, let the graph $\mathcal{G}$ have the diffusive property in which there exists $\alpha < 1$ such that for each $i \in \{1, \ldots, L\}$ and each observation $z_{w_i}^j$ at node x_{w_i},*

$$\begin{aligned} I(x_{\mathcal{N}(w_i)}; z_{w_i}^j | z_{w_1}^{g_1}, \ldots, z_{w_{i-1}}^{g_{i-1}}) \\ \leq \alpha I(x_{w_i}; z_{w_i}^j | z_{w_1}^{g_1}, \ldots, z_{w_{i-1}}^{g_{i-1}}) \end{aligned}$$

where $x_{\mathcal{N}(w_i)}$ denotes the neighbors of node x_{w_i} in the latent structure graph $\mathcal{G}$.

Assumption 3 states that the information which the observation $z_{w_i}^j$ contains about x_{w_i} is discounted by a factor of at least α when compared to the information it contains about the remainder of the graph. Theorem 3 uses this property to bound the loss of optimality associated with the greedy choice to be a factor of $(1 + \alpha)$ rather than 2.

Theorem 3. *Under Assumptions 2 and 3, the performance of the greedy heuristic in Definition 2 satisfies the following guarantee:*

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_L}^{o_L}) \leq (1 + \alpha) I(X; z_{w_1}^{g_1}, \ldots, z_{w_L}^{g_L})$$

Proof. The proof follows an induction on the following expression, which trivially holds for $j = 1$,

$$\begin{aligned} I(X; z_{w_1}^{o_1}, \ldots, z_{w_L}^{o_L}) \leq (1 + \alpha) I(X; z_{w_1}^{g_1}, \ldots, z_{w_{j-1}}^{g_{j-1}}) \\ + I(X; z_{w_j}^{o_j}, \ldots, z_{w_L}^{o_L} | z_{w_1}^{g_1}, \ldots, z_{w_{j-1}}^{g_{j-1}}) \end{aligned} \tag{6}$$

The steps necessary to establish the induction step are similar to those in Eq. (1); details may be found in [10]. □

3.1 Online guarantee

For many models the diffusive property is difficult to establish globally. Following from step (d) of Theorem 3, one may obtain an online computable bound which does not require the property of Assumption 3 to hold globally, but exploits it to whatever extent it exists in a particular selection problem.

Theorem 4. *Under the model of Assumption 2, but not requiring the diffusive property of Assumption 3, the following performance guarantee, which can be computed online, applies to the greedy heuristic of Definition 2:*

$$\begin{aligned} I(X; z_{w_1}^{o_1}, \ldots, z_{w_L}^{o_L}) \leq I(X; z_{w_1}^{g_1}, \ldots, z_{w_L}^{g_L}) + \\ \sum_{j=1}^{L} I(x_{\mathcal{N}(w_j)}; z_{w_j}^{g_j} | z_{w_1}^{g_1}, \ldots, z_{w_{j-1}}^{g_{j-1}}) \end{aligned}$$

The proof of this result may be found in [10].

3.2 Specialization to trees and chains

In the common case where the latent structure $X = \{x_1, \ldots, x_L\}$ forms a tree, we may avoid including *all* neighbors of a node in the condition of Assumption 3 and in the result of Theorem 4, replacing it instead with the *parent* in the tree. An additional requirement on the sequence $(w_1, \ldots, w_L)$ is necessary to exploit the tree structure, namely that the walk must be "bottom-up", *i.e.*, no node may be visited before all of its children have been visited. Theorems 3 and 4 both hold under these modified assumptions; the proofs pass directly once $x_{\mathcal{N}(w_j)}$ is replaced by $x_{\pi(w_j)}$ in step (d). Details may be found in [10]. The most common application of the diffusive model is in Markov chains (a special case of a tree), where the i-th node corresponds to time i. In this case, the sequence is simply $w_i = i$, *i.e.*, we visit the nodes in time order. Choosing the final node in the chain to be the tree root, this sequence respects the bottom-up requirement, and the diffusive requirement becomes:

$$I(x_{k+1}; z_k^j | z_1^{g_1}, \ldots, z_{k-1}^{g_{k-1}}) \leq \alpha I(x_k; z_k^j | z_1^{g_1}, \ldots, z_{k-1}^{g_{k-1}}) \tag{7}$$

3.3 Example of online diffusive guarantee

Consider an object which moves in two dimensions according to a Gaussian random walk:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{w}_k$$

where $\boldsymbol{w}_k \sim \mathcal{N}\{\boldsymbol{w}_k; \mathbf{0}, \mathbf{I}\}$. The initial position of the object is distributed according to $\boldsymbol{x}_0 \sim \mathcal{N}\{\boldsymbol{x}_0; \mathbf{0}, \mathbf{I}\}$.

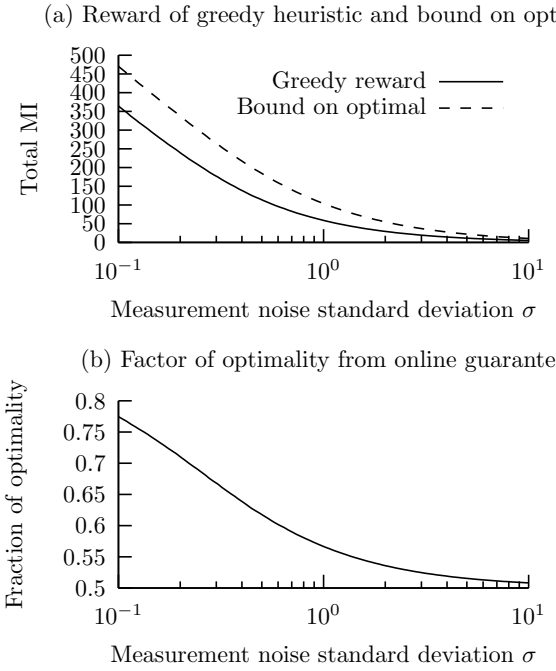


Figure 2: (a) shows average total reward accrued by the greedy heuristic in the 200 time steps for different diffusion strength values (q), and the bound on optimal obtained through the tree version of Theorem 4. (b) shows the ratio of these curves, providing the factor of optimality guaranteed by the bound.

Noisy bearing observations are available from four sensors positioned at $(\pm 100, \pm 100)$, but only one observation may be utilized at any instant. Simulations were run for 200 time steps. The total reward and the bound obtained from the tree version of Theorem 4 are shown in Fig. 2(a) as a function of the measurement noise standard deviation (in degrees). The results demonstrate that the performance guarantee becomes stronger as the measurement noise decreases; the same effect occurs if the observation noise is held constant and the dynamics noise increased. Fig. 2(b) shows the ratio of the greedy performance to the upper bound on optimal, demonstrating that the greedy heuristic is guaranteed to be within a factor of 0.77 of optimal with a measurement standard deviation of 0.1 degrees.

In this example, we utilized the closed loop greedy heuristic examined in Section 4, hence it was necessary to use multiple Monte Carlo simulations to compute the online guarantee. Tracking was performed using an extended Kalman filter, hence the bounds are approximate (the EKF variances were used to calculate the rewards). In this scenario, the low degree of nonlinearity in the observation model provides confidence that the inaccuracy in the rewards is insignificant.

4 Closed loop control

The preceding analysis concentrates on an open loop control structure, *i.e.*, it assumes that all observation choices are made before any observation values are received. Greedy heuristics are often applied in a closed loop setting, in which an observation is chosen, and then its value is received before the next choice is made.

The performance guarantees of Theorems 1 and 3 both apply to the *expected* performance of the greedy heuristic operating in a closed loop fashion, *i.e.*, in expectation the closed loop greedy heuristic achieves at least half the reward of the optimal *open* loop selection. Theorem 5 establishes the result of Theorem 1 for the closed loop heuristic. The same process can be used to establish a closed loop version of Theorem 3. To obtain the closed loop guarantee, we need to exploit an additional characteristic of mutual information:

$$I(X; z^{\mathcal{A}} | z^{\mathcal{B}}) = \int I(X; z^{\mathcal{A}} | z^{\mathcal{B}} = \zeta) p_{z^{\mathcal{B}}}(\zeta) \mathrm{d}\zeta \qquad (8)$$

While the results are presented in terms of mutual information, they apply to any other objective which meets the previous requirements as well as Eq. (8).

We define $\hbar_j = (u_1, z^{u_1}_{w_1}, u_2, z^{u_2}_{w_2}, \ldots, u_{j-1}, z^{u_{j-1}}_{w_{j-1}})$ to be the history of all observation actions chosen, and the resulting observation values, *prior to* stage j (this constitutes all the information which we can utilize in choosing our action at time j). Accordingly, $\hbar_1 = \emptyset$, and $\hbar_{j+1} = (\hbar_j, u_j, z^{u_j}_{w_j})$. The greedy heuristic operating in closed loop is defined in Definition 3.

Definition 3. *Under the same assumptions as Theorem 1, define the closed loop greedy heuristic policy* μ^g*:*

$$\mu^g_j(\hbar_j) = \arg\max_{u \in \{1,\ldots,n_{w_j}\}} I(X; z^u_{w_j} | \hbar_j) \qquad (9)$$

We use the convention that conditioning on $\hbar_i$ in an MI expression is always on the *value*, and hence if $\hbar_i$ contains elements which are random variables we will always include an explicit expectation operator. The expected reward-to-go from stage j to the end of the planning horizon for the greedy heuristic $\mu^g_j(\hbar_j)$ commencing from the history $\hbar_j$ is denoted as:

$$J^{\mu^g}_j(\hbar_j) = I(X; z^{\mu^g_j(\hbar_j)}_{w_j}, \ldots, z^{\mu^g_N(\hbar_N)}_{w_N} | \hbar_j) \qquad (10)$$

$$= \mathrm{E}\left[\sum_{i=j}^{N} I(X; z^{\mu^g_i(\hbar_i)}_{w_i} | \hbar_i) \,\middle|\, \hbar_j \right] \qquad (11)$$

The expectation in Eq. (11) is over the random variables corresponding to the actions

$\{\mu_{j+1}^g(\hbar_{j+1}), \ldots, \mu_N^g(\hbar_N)\}$,[3] along with the observations resulting from the actions, $\{z_{w_j}^{\mu_j^g(\hbar_j)}, \ldots, z_{w_N}^{\mu_N^g(\hbar_N)}\}$, where $\hbar_i$ is the concatenation of the previous history sequence $\hbar_{i-1}$ with the new observation action $\mu_i^g(\hbar_i)$ and the new observation value $z_{w_i}^{\mu_i^g(\hbar_i)}$. The expected reward of the greedy heuristic over the full planning horizon is $J_1^{\mu^g}(\emptyset)$. We also define the expected reward accrued by the greedy heuristic up to and including stage j, commencing from an empty history sequence (*i.e.*, $\hbar_1 = \emptyset$), as:

$$J_{\rightarrow j}^{\mu^g} = \mathrm{E}\left[\sum_{i=1}^{j} I(X; z_{w_i}^{\mu_i^g(\hbar_i)} | \hbar_i)\right] \tag{12}$$

This gives rise to the recursive relationship:

$$J_{\rightarrow j}^{\mu^g} = \mathrm{E}[I(X; z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j)] + J_{\rightarrow j-1}^{\mu^g} \tag{13}$$

Comparing Eq. (11) with Eq. (12), we have $J_{\rightarrow N}^{\mu^g} = J_1^{\mu^g}(\emptyset)$. We define $J_{\rightarrow 0}^{\mu^g} = 0$.

The expected reward of the tail of the optimal open loop observation sequence $(o_j, \ldots, o_N)$ commencing from the history $\hbar_j$ is denoted by:

$$J_j^o(\hbar_j) = I(X; z_{w_j}^{o_j}, \ldots, z_{w_N}^{o_N} | \hbar_j) \tag{14}$$

Using the MI chain rule and Eq. (8), this can be written recursively as:

$$J_j^o(\hbar_j) = I(X; z_{w_j}^{o_j} | \hbar_j) + \mathop{\mathrm{E}}_{z_{w_j}^{o_j} | \hbar_j} J_{j+1}^o[(\hbar_j, o_j, z_{w_j}^{o_j})] \tag{15}$$

where $J_{N+1}^o(\hbar_{N+1}) = 0$. The reward of the optimal open loop observation sequence over the full planning horizon is:

$$J_1^o(\emptyset) = I(X; z_{w_1}^{o_1}, \ldots, z_{w_N}^{o_N}) \tag{16}$$

We now seek to obtain a guarantee on the performance ratio between the optimal open loop observation sequence and the closed loop greedy heuristic. The proof of the theorem uses the following simple result, which is proven in [10].

Lemma 1. *Given the above definitions:*

$$\mathop{\mathrm{E}}_{z_{w_j}^{o_j} | \hbar_j} J_{j+1}^o[(\hbar_j, o_j, z_{w_j}^{o_j})] \leq J_{j+1}^o(\hbar_j)$$
$$\leq I(X; z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j) + \mathop{\mathrm{E}}_{z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j} J_{j+1}^o[(\hbar_j, \mu_j^g(\hbar_j), z_{w_j}^{\mu_j^g(\hbar_j)})]$$

[3]We assume a deterministic policy, hence the action at stage j is fixed given knowledge of $\hbar_j$.

Theorem 5. *Under the same assumptions as Theorem 1,*

$$J_1^o(\emptyset) \leq 2 J_1^{\mu^g}(\emptyset)$$

i.e., the expected reward of the closed loop greedy heuristic is at least half the reward of the optimal open loop policy.

Proof. To establish an induction, assume that

$$J_1^o(\emptyset) \leq 2 J_{\rightarrow j-1}^{\mu^g} + \mathrm{E}\, J_j^o(\hbar_j) \tag{17}$$

Noting that $\hbar_1 = \emptyset$, this trivially holds for $j = 1$ since $J_{\rightarrow 0}^{\mu^g} = 0$. Now, assuming that it holds for j, we show that it also holds for $(j+1)$. Applying Eq. (15),

$$J_1^o(\emptyset) \leq 2 J_{\rightarrow j-1}^{\mu^g} + \mathrm{E}\left\{ I(X; z_{w_j}^{o_j} | \hbar_j) + \mathop{\mathrm{E}}_{z_{w_j}^{o_j} | \hbar_j} J_{j+1}^o[(\hbar_j, o_j, z_{w_j}^{o_j})] \right\}$$

By the definition of the closed loop greedy heuristic (Definition 3),

$$I(X; z_{w_j}^{o_j} | \hbar_j) \leq I(X; z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j)$$

hence:

$$J_1^o(\emptyset) \leq 2 J_{\rightarrow j-1}^{\mu^g} + \mathrm{E}\left\{ I(X; z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j) + \mathop{\mathrm{E}}_{z_{w_j}^{o_j} | \hbar_j} J_{j+1}^o[(\hbar_j, o_j, z_{w_j}^{o_j})] \right\}$$

Applying Lemma 1, followed by Eq. (13):

$$J_1^o(\emptyset) \leq 2 J_{\rightarrow j-1}^{\mu^g} + \mathrm{E}\left\{ 2 I(X; z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j) + \mathop{\mathrm{E}}_{z_{w_j}^{\mu_j^g(\hbar_j)} | \hbar_j} J_{j+1}^o[(\hbar_j, \mu_j^g(\hbar_j), z_{w_j}^{\mu_j^g(\hbar_j)})] \right\}$$
$$= 2 J_{\rightarrow j}^{\mu^g} + \mathrm{E}\, J_{j+1}^o(\hbar_{j+1})$$

where $\hbar_{j+1} = (\hbar_j, \mu_j^g(\hbar_j), z_{w_j}^{\mu_j^g(\hbar_j)})$. This establishes the induction step.

Applying the induction step N times, we obtain:

$$J_1^o(\emptyset) \leq 2 J_{\rightarrow N}^{\mu^g} + \mathrm{E}\, J_{N+1}^o(\hbar_{N+1}) = 2 J_1^{\mu^g}(\emptyset)$$

since $J_{N+1}^o(\hbar_{N+1}) = 0$ and $J_{\rightarrow N}^{\mu^g} = J_1^{\mu^g}(\emptyset)$. □

We emphasize that this performance guarantee is for *expected* performance: it does not provide a guarantee for the change in entropy of every sample path. An online bound cannot be obtained on the basis of a single realization, although online bounds similar to Theorems 2 and 4 could be calculated through Monte Carlo simulation (to approximate the expectation). The guarantee for K-element subset selection in [8] can be extended similarly to closed loop selection; the analogous result is proven in [10].

4.1 Closed loop greedy vs closed loop optimal

While Theorem 5 provides a performance guarantee with respect to the optimal open loop sequence, there is no guarantee relating the performance of the closed loop greedy heuristic to the optimal closed loop controller, as the following example illustrates. One exception to this is linear Gaussian models, where closed loop policies can perform no better than open loop sequences, so that the open loop guarantee extends to closed loop performance.

Example 2. *Consider the following two-stage problem, where* $X = [a, b, c]^T$*, with* $a \in \{1, \ldots, N\}$*,* $b \in \{1, \ldots, N+1\}$*, and* $c \in \{1, \ldots, M\}$*. The prior distribution of each of these is uniform and independent. In the first stage, we may measure* $z_1^1 = a$ *for reward* $\log N$*, or* $z_1^2 = b$ *for reward* $\log(N+1)$*. In the second stage, we may choose* z_2^i*,* $i \in \{1, \ldots, N\}$*, where*

$$z_2^i = \begin{cases} c, & i = a \\ d, & otherwise \end{cases}$$

where d *is independent of* X*, and is uniformly distributed on* $\{1, \ldots, M\}$*. The greedy algorithm in the first stage selects the observation* $z_1^2 = b$*, as it yields a higher reward (*$\log(N+1)$*) than* $z_1^1 = a$ *(*$\log N$*). At the second stage, all options have the same reward,* $\frac{1}{N} \log M$*, so we choose one arbitrarily for a total reward of* $\log(N+1) + \frac{1}{N} \log M$*. The optimal algorithm in the first stage selects the observation* $z_1^1 = a$ *for reward* $\log N$*, followed by the observation* z_2^a *for reward* $\log M$*, for total reward* $\log N + \log M$*. The ratio of the greedy reward to the optimal reward is*

$$\frac{\log(N+1) + \frac{1}{N} \log M}{\log N + \log M} \to \frac{1}{N}, \; M \to \infty$$

Hence, by choosing N *and* M *to be large, we can obtain an arbitrarily small ratio between the greedy closed-loop reward and the optimal closed-loop reward.*

We conjecture that it may be possible to establish a closed loop performance guarantee for diffusive processes, but it is likely to be dramatically weaker than the bounds presented in this paper. In [7] it is established that the number of tests required by the closed loop greedy heuristic to perfectly learn the state of a collection of K binary random variables[4] from a set of possible tests (each of which depends on one or more of the underlying variables) is within a factor of $O(\log K)$ of the optimal (smallest) set. Our example demonstrates just how bad the factor can be in the related problem where the criterion is to maximize performance using a fixed set of resources (rather than minimizing resources to obtained a fixed level of performance), and where variables can be of arbitrary cardinality (rather than binary) and the set of observations from which we choose changes from iteration to iteration (rather than selecting at each stage from a fixed set of observations).

[4] or learn it as well as it can be learned using a fixed group of observations

5 Conclusion

The performance guarantees presented in this paper provide theoretical basis for simple heuristic algorithms that are widely used in practice. The guarantees apply to both open loop and closed loop operation, and are naturally tighter for diffusive processes.

Acknowledgements

This work was supported by MIT Lincoln Laboratory through ACC PO#3019934.

References

[1] K.J. Hintz and E.S. McVey, "Multi-process constrained estimation," *Systems, Man and Cybernetics, IEEE Trans on*, vol. 21, no. 1, pp. 237–244, 1991.

[2] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, NY, 1991.

[3] A. Krause, C. Guestrin, and A.P. Singh, "Near-optimal sensor placements in Gaussian processes," in *International Conference on Machine Learning*, August 2005.

[4] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, March 2002.

[5] C.M. Kreucher, A.O. Hero III, K. Kastella, and D. Chang, "Efficient methods of non-myopic sensor management for multitarget tracking," in *43rd IEEE Conference on Decision and Control*, December 2004.

[6] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden Markov model sensors," *Signal Processing, IEEE Trans on*, vol. 50, no. 6, pp. 1382–1397, 2002.

[7] A.X. Zheng, I. Rish, and A. Beygelzimer, "Efficient test selection in active diagnosis via entropy approximation," in *Uncertainty in AI*, July 2005.

[8] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Uncertainty in AI*, July 2005.

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

[10] J.L. Williams, *Information Theoretic Sensor Management*, Ph.D. thesis, Massachusetts Institute of Technology, February 2007.

[11] G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher, "An analysis of approximations for maximizing submodular set functions–II," in *Polyhedral combinatorics*, M.L. Balinski and A.J. Hoffman, Eds., vol. 8 of *Mathematical programming study*, pp. 73–87. Elsevier, 1978.