

# Experiments in Underwater Feature Tracking with Performance Guarantees Using a Small AUV

Benjamin Biggs  
Virginia Tech  
babiggs@vt.edu

Hans He  
Virginia Tech  
hjh2bs@vt.edu

James McMahon  
US Naval Research Laboratory  
Acoustics Division, Code 7130  
james.mcmahon@nrl.navy.mil

Daniel J. Stilwell  
Virginia Tech  
stilwell@vt.edu

**Abstract**—We present the results of experiments performed using a small autonomous underwater vehicle to determine the location of an isobath within a bounded area. The primary contribution of this work is to implement and integrate several recent developments real-time planning for environmental mapping, and to demonstrate their utility in a challenging practical example. We model the bathymetry within the operational area using a Gaussian process and propose a reward function that represents the task of mapping a desired isobath. As is common in applications where plans must be continually updated based on real-time sensor measurements, we adopt a receding horizon framework where the vehicle continually computes near-optimal paths. The sequence of paths does not, in general, inherit the optimality properties of each individual path. Our real-time planning implementation incorporates recent results that lead to performance guarantees for receding-horizon planning.

## I. INTRODUCTION

Determining the location of a depth contour or isobath is an important element of quickly finding navigable paths in waterways. The conventional approach to underwater feature mapping relies on performing an exhaustive search over an area where a human operator believes features of interest are located. Path planning methods that seek to maximize information gain by planning sequences of locations that maximize a specific reward function are available to attempt to track underwater features more quickly than exhaustive search methods. However, many of these methods are susceptible to poor performance caused by myopic planning horizons when using receding horizon methods or local maxima in the reward function when using gradient-based optimization approaches. The primary contribution of this paper is the presentation of the results of experiments in depth contour estimation using path planning techniques with performance guarantees. We integrate several recent developments for real-time isobath localization, all of which can be easily modified for similar real-time mapping applications: (1) use of a sparse Gaussian process for representing the bathymetry, which leads to a continuous representation of the environment

\*This work was supported by the Office of Naval Research via grants N00014-18-1-2627, and N00014-19-1-2194. The work of J. McMahon is supported by the Office of Naval Research through the NRL Base Program.

James McMahon is with the US Naval Research Laboratory, Code 7130, Washington D.C., USA

Benjamin Biggs, Hans He, and Daniel Stilwell are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA



Fig. 1: 690 vehicles being loaded for transportation to the operational area.

along with rigorous assessment of uncertainty, and (2) use of a so-called terminal reward in the optimization problem that generates each near-optimal path, which ensures that the cumulative reward attained by the AUV from a sequence of near-optimal paths is no worse than a desired lower bound. An important contribution of this work is to describe specific implementation solutions needed to integrate and implement these ideas for real-time operation on an AUV.

Isobath localization, like front tracking, can be formulated as an informative path planning (IPP) problem [1]. The goal of the IPP problem is to find a path that maximizes an information utility function while not exceeding a budget constraint on the cost of the path. In general, the computational complexity associated with informative path planning grows exponentially with the length of the path. Heuristic planning methods may be used to produce reasonable results [2]–[4]. However, it is desirable to plan paths that are as nearly optimal as possible. Despite a host of techniques developed to provide approximate or suboptimal solutions [5]–[8], receding horizon methods are often used regardless of the planning techniques used within the shorter planning horizon [9]–[11].

A receding horizon path is constructed from segments of near-optimal short paths that are greedily selected. However, the sequence of path segments in a receding horizon implementation do not inherit the near-optimality of each

individual segment. In prior work [12], [13] the authors showed how to append a specific terminal reward to the short horizon optimization problem used for each path segment such that the reward of the overall receding horizon path is guaranteed to exceed a desirable lower bound. This result was demonstrated for the explicit reward function associated with a robotic search problem in a discrete environment through numerical experiments. In this work, we demonstrate the utility of adding a terminal reward to a different class of reward function based on level set estimation using Gaussian process models. We report on the results of challenging real-world experiments accomplished using a real autonomous underwater vehicle (AUV) performing on-line planning using live data.

Throughout this work, we seek to rapidly map the 15 meter isobath within a bounded area of Claytor lake in Dublin, VA, USA. The specific region of Claytor lake where the experiments of this work are performed is shown in Figure 4 with the solid red lines indicating the boundaries of the operational area. The AUV used for these experiments is the Virginia Tech 690 vehicle shown in Figure 1.

The remainder of this paper is organized as follows. Gaussian process regression and the specific considerations and methods used in this work are presented in Section II. The specific reward function used for evaluating paths is presented in Section III. The path planning methods proposed in [13] and used in this work are presented and discussed in Section IV. The usage and relevant capabilities of the 690 vehicle are discussed in Section V. The experimental procedure and selection of parameters is presented in Section VI. Results and conclusions are presented in Section VII.

## II. GAUSSIAN PROCESS REGRESSION

In this section we provide a brief review of Gaussian process regression. We refer the reader to [14, Chapter 2] for a more thorough discussion of the topic.

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution [14, Definition 2.1]. A GP is completely specified by its mean and covariance functions  $m(p)$  and  $k(p, p')$  respectively and is written

$$f(p) \sim \mathcal{GP}(m(p), k(p, p')) \quad (1)$$

where  $f(p)$  is a real process. In this work, we assume that the bathymetry or depth of Claytor lake within the operational area may be modeled using a GP.

A sample or datum  $d = (p, z)$  is an input-measurement pair with  $p \in \mathcal{P}$  and  $z \in \mathbb{R}$ . Let  $\mathcal{D}_t$  be the set of all samples acquired up to time  $t$ . For any input  $p_i$ , an associated measurement  $z_i$  is related to  $p_i$  via

$$z_i = f(p_i) + \epsilon_i, \text{ for } i = 1, \dots, |\mathcal{D}_t| \quad (2)$$

where  $|\mathcal{D}|$  denotes the cardinality of the set  $\mathcal{D}$  and  $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$  is drawn from a zero mean Gaussian distribution with standard deviation  $\sigma_n$ .

Given the data set  $\mathcal{D}_t$ , one may wish to predict  $f(p^*)$  for an arbitrary input  $p^*$ . From [14, Chapter 2.7], the posterior

predictive distribution of  $f(p^*)$  conditioned on  $\mathcal{D}_t$  is Gaussian and specified by the mean  $\mu_{p^*|\mathcal{D}_t}$  and variance  $\sigma_{p^*|\mathcal{D}_t}^2$  given as

$$\mu_{p^*|\mathcal{D}_t} = m(p^*) + k(p^*)^\top (K_t + \sigma_n^2 I)^{-1} (Z_t - m(P_t)) \quad (3)$$

$$\sigma_{p^*|\mathcal{D}_t}^2 = k(p^*, p^*) - k(p^*)^\top (K_t + \sigma_n^2 I)^{-1} k(p^*). \quad (4)$$

$Z_t = [z_1, \dots, z_{|\mathcal{D}_t|}]^\top$  is the vector of all measurements contained in  $\mathcal{D}_t$  and  $P_t = \{p_1, \dots, p_{|\mathcal{D}_t|}\}$  is the set of associated inputs. The prior mean vector is  $[m(P_t)]_i = m(p_i)$ ,  $[k(p^*)]_i = k(p^*, p_i)$ ,  $[K_t]_{i,j} = k(p_i, p_j)$ , and  $I_t$  is the  $|\mathcal{D}_t| \times |\mathcal{D}_t|$  identity matrix.

When computing the predictive distribution at  $p^*$ , the matrix inversion  $(K_t + \sigma_n^2 I)^{-1}$  can become intractable as  $|\mathcal{D}_t|$  grows large. To address this complexity, we utilize a distance metric  $d(p, p')$  on  $\mathcal{P}$  to reduce the number of samples retained in  $\mathcal{D}_t$  and make predictions using local Gaussian processes. To reduce the number of samples retained in  $\mathcal{D}_t$ , a new sample  $d_c = (p_c, z_c)$  is only added to  $\mathcal{D}_t$  if  $d(p_c, p_i) \geq \delta_f$  for all  $p_i \in \mathcal{D}_t$ . For a predictive point  $p^*$ , the associated data is limited to a locality around  $p^*$  giving the local data set as  $\mathcal{D}_{p^*,t} = \{d_i \in \mathcal{D}_t | d(p_i, p^*) \leq \delta_c\}$ . The local predictive mean and variance are given by equations (3) and (4) using  $\mathcal{D}_{p^*,t}$  in the place of  $\mathcal{D}_t$ .

Throughout this work, we define the input space  $\mathcal{P}$  to be two dimensional euclidean space ( $\mathbb{R}^2$ ) with distance metric  $d(p, p') = \|p - p'\|$  being the euclidean distance. We use the squared exponential kernel function

$$k(p, p') = \sigma_f^2 \exp \left\{ -\frac{\|p - p'\|^2}{2l_c^2} \right\} \quad (5)$$

as the prior for the covariance function. The prior mean function used in this work is presented in Section VI.

## III. OBJECTIVE FUNCTION

We wish to determine the level set  $\mathcal{L}(l^*) = \{p \in \mathcal{P} | f(p) = l^*\}$ . To direct the selection of locations that should be measured in order to obtain information regarding this level set, we require an objective function that encodes uncertainty regarding whether  $p^*$  is an element of  $\mathcal{L}(l^*)$ . To provide such a function, we use a measure of ambiguity presented in [15] which is closely related to the straddle heuristic presented in [16].

*Definition 1 (Ambiguity):* The ambiguity associated with a point  $p^*$  with respect to a level  $l$  is

$$a_{p^*|\mathcal{D}_t} = \beta \sigma_{p^*|\mathcal{D}_t} - |\mu_{p^*|\mathcal{D}_t} - l|. \quad (6)$$

Concretely, if  $\beta = 2$  and  $a_{p^*|\mathcal{D}_t} < 0$ , then the model indicates that the probability that  $f(p^*)$  falls on the opposite side of  $l$  as  $\mu_{p^*|\mathcal{D}_t}$  is less than 2.5 percent.

Similar to [15] where batches of points are selected for sampling, we seek to select a path or a sequence of points to sample. Considering the ambiguity of sets of points requires some slight modifications to account for locations that will be sampled while no samples have yet been acquired.

*Definition 2 (Anticipated Ambiguity):* Let  $\mathcal{P}^* = \{p_0^*, \dots, p_n^*\}$  and  $\mathcal{P}_i^* = \{p_0^*, \dots, p_{i-1}^*\}$  with  $\mathcal{P}_0^* = \emptyset$ . Then

the anticipated ambiguity of  $p_i^* \in \mathcal{P}^*$ , with respect to a level  $l$  is

$$a'_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*} = \beta \sigma_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*} - |\mu_{p^*|\mathcal{D}_t} - l| \quad (7)$$

where  $\sigma_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*}$  is the posterior predictive standard deviation computed using equation (4) where  $\mathcal{P}_t$  is augmented with  $\mathcal{P}_i^*$ .

Having defined the anticipated ambiguity for a set of points, we define our reward function to be the anticipated reduction in ambiguity at a point with respect to a level  $l$ .

$$r_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*} = \max(a'_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*}, 0) - \max(a'_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^* \cup p_i^*}, 0). \quad (8)$$

Note that  $a'_{p_i^*|\mathcal{D}_t, \mathcal{P}_0^*} = a_{p_i^*|\mathcal{D}_t}$ . Also note that without the inclusion of the max function, the reward simply becomes  $\beta[\sigma_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^*} - \sigma_{p_i^*|\mathcal{D}_t, \mathcal{P}_i^* \cup p_i^*}]$ , which is uninformative regarding  $\mathcal{L}(l)$ . Intuitively, the reward places the most value on points  $p^*$  where  $f(p^*)$  is expected to be close to  $l$  and the anticipated standard deviation at  $p^*$  is expected to be reduced significantly. That is,  $f(p^*)$  may be close to  $l$ , but if the anticipated reduction in variance is low, there is little to be gained by sampling that location.

#### IV. RECEDING HORIZON PATH PLANNING

As in [13], we consider that the state of a mobile search agent may be represented by the tuple  $s = \{p, x\}$  where  $x$  gives all states of the mobile search agent excluding the position states  $p$ . We say that a state  $s_i$  is feasible from the state  $s_{i-1}$  if there exists an action  $a$  available to the search agent such that applying  $a$  to the state  $s_{i-1}$  results in the new state  $s_i$ . An  $n$ -length path may then be represented as a sequence of states  $\gamma_n(s_0) = \{s_0, s_1, \dots, s_n\}$ . The set of all feasible  $n$ -length paths beginning at state  $s_0$  is denoted  $\Gamma_n(s_0)$ .

Using the point-wise reduction in ambiguity in equation (8) as the reward function, the value of a path is the sum of rewards obtained while traversing the path. That is, the anticipated reward for the path  $\gamma_n(s_k)$  is

$$J(\gamma_n(s_0)) = \sum_{i=0}^n r_{p_i|\mathcal{D}_t \cup \mathcal{P}_i} \quad (9)$$

where  $p_i \in s_i \in \gamma_n(s_k)$  is the point to be sampled and  $\mathcal{P}_i = \{p_0, \dots, p_{i-1}\}$  is the set of points that will have been sampled by step  $i$ . For notational simplicity we define

$$J_n(s_0) \triangleq J(\gamma_n(s_0)). \quad (10)$$

The goal of path planning is to determine a path  $\gamma_n^*(s_0)$  satisfying

$$\gamma_n^*(s_0) = \arg \max_{\gamma_n(s_0) \in \Gamma_n(s_0)} J_n(s_0). \quad (11)$$

In general, the complexity of (11) grows exponentially as the length of the desired path increases. One method of addressing the complexity associated with planning long paths is to employ receding horizon planning methods wherein a search agent solves for a short optimal or nearly optimal path, traverses a portion of the short path, and repeats the process until a path of the desired length has been traversed. The length of the computationally feasible short paths is termed

the *planning horizon*. The number of steps that the agent traverses along each short path before planning a new short path is termed the *execution horizon*.

While receding horizon path planning is commonly used in practice and generally yields good results, there are no optimality guarantees for paths produced using receding horizon methods despite the receding horizon paths being constructed from segments of optimal or nearly optimal short paths. In fact, receding horizon planning methods may produce worst possible solutions.

In prior works [12] and [13], it was shown that the appropriate use of a terminal reward when constructing receding horizon paths can guarantee a lower bound on the value of receding horizon paths produced. The primary tool in producing this lower bound is a lower bound on the *value-to-go* or a lower bound on the optimal value of the remainder of the path normally ignored during receding horizon planning.

*Definition 3:* [13, Definition 3.1][value-to-go] A lower bound on the *value-to-go* is the function  $B_l(s_n)$  that satisfies

$$B_l(s_n) \leq J_{l-n}^*(s_n) \quad (12)$$

for all  $n \leq l$  with  $B_l(s_l) = 0$ , and for the infinite case where  $l = \infty$ ,  $B_l(s_k) \rightarrow \infty$  as  $k \rightarrow \infty$ .

Using the value-to-go function  $B$ , the authors of [13] propose a new objective function as

$$\mathcal{V}(\gamma_n(s_0)) = J_n(s_0) + B_l(s_n) \quad (13)$$

with the abbreviated notation

$$\mathcal{V}_n(s_0) \triangleq \mathcal{V}(\gamma_n(s_0)). \quad (14)$$

Using the objective function of equation 13, the lower bound guarantee that we seek leverage in the field trials of this work, is given in Proposition 4.

*Proposition 4:* [13, Proposition 4.2] Suppose that at each planning step  $k$  the path  $\gamma_n(s_k)$  satisfies

$$\mathcal{V}_n(s_k) \geq \mathcal{V}_{n-m}(s_k) \quad (15)$$

where  $\mathcal{V}_{n-m}(s_k)$  is the value of the  $n - m$  remainder of the path produced at the previous planning step and

$$\mathcal{V}_n(s_0) \geq B_l(s_0) \quad (16)$$

then the  $l$ -length RH path  $\tilde{\gamma}_l(s_0)$  satisfies

$$J(\tilde{\gamma}_l(s_0)) \geq B_l(s_0). \quad (17)$$

Significantly, it is shown in Corollary 4.3 of [13] that the lower bound on the value-to-go may be defined using the value of the highest-reward path from a set of naive paths. This allows the receding horizon planner to leverage the strengths of distinct naive paths to produce a receding horizon path that will perform at least as well as the best naive solution and possibly much better.



Fig. 2: Simulated DVL beams shown in red.

## V. VIRGINIA TECH 690 AUV

Experiments were performed using the Virginia Tech 690 AUV shown in Figure 1. Specifications for the 690 are given in Table I. Significantly, the 690 vehicle is equipped with a Doppler velocity log (DVL, TELEDYNE MARINE Pathfinder) with four acoustic beams at a 30 degree janus angle. An example AUV illustrating the DVL placement on the 690 is shown in Figure 2 with DVL beams shown in red. The DVL records a measurement of the distance from the DVL sensor to the lake bed along each beam at a rate of 10Hz. The 690 vehicle is able to estimate its location using DVL-aided inertial navigation. The navigation estimate is fused with the DVL range measurements to infer the depth of the lake bed. We use all four range measurements in our implementation. Throughout this paper, vehicle state, and sensor measurements are converted into a North, East, Down (NED) coordinate frame. The origin of the NED coordinate system is shown as the green point in Figure 4.

PARAMETER	SPECIFICATION
Length	2.2 m
Displacement	42.9 kg
Diameter	17.5 cm
Endurance	24 hrs @ 2 m/s
Depth	500 meters
Communications	900MHz RF modem, Wi-Fi, acoustic communication, satellite communication
Navigation	DVL + IMU, acoustic ranging
Mission Sensor	PingDSP 3D bathymetric sonar

TABLE I: Specifications of the Virginia Tech 690 AUV.

Control and navigation of the 690 vehicle are accomplished using ROS-based autonomy software running on an

ODROID-XU4 single board computer. Path planning for these experiments is accomplished separately on an NVIDIA Jetson TX2. Line segments connecting waypoints in the paths computed onboard the Jetson are delivered to the ROS-based autonomy software on the ODROID via a TCP interface. The autonomy software running on the ODROID drives the 690 vehicle to converge to the input line segment and reports when the end of the line segment has been reached. DVL sensor measurements and vehicle navigation data are transmitted from the ODROID to the Jetson over the same interface. Notably, executing path planning on a separate computer and sending line segments over a well-defined interface, ensures that the path planning algorithm does not interfere with critical autonomy functionality.

## VI. FIELD TRIALS

The prior mean  $m(p)$  is defined using a line  $\mathcal{L}$ . Let  $d(p, \mathcal{L})$  be the minimum euclidean distance from  $p$  to  $\mathcal{L}$ . Then the prior mean at a point  $p$  is given as

$$m(p) = \max(d_{max}(1 - [d(p, \mathcal{L})/c]^2), 0). \quad (18)$$

Intuitively, the line  $\mathcal{L}$  gives the anticipated deepest set of points in the lake with the anticipated depth decreasing proportionally to the square of the distance from the line until a depth of zero is reached. The coefficients  $c$  and  $d_{max}$  are chosen to be 300 and 25 meters respectively. The specific line chosen for these trials is shown as the cyan dashed line in Figure 4. Figure 3 shows the 15 meter isobath of the prior mean function within the bounded search area. This choice of prior mean function provides a naive initial estimate of the location of the 15 meter isobath which facilitates quickly finding the true isobath location, but is not meant to represent the bathymetry accurately beyond an initial naive estimate. The parameters of the squared exponential prior covariance function in equation (5) are chosen to be  $\sigma_f = 5.0$ ,  $l_c = 40.0$  meters, and the sensor measurement noise  $\sigma_n = 2.5$  meters. The data retention threshold  $\delta_f$  and local GP diameter  $\delta_c$  presented in Section II are chosen to be  $\delta_f = 6$  meters and  $\delta_c = 40$  meters or  $0.15l_c$  and  $l_c$  respectively.

### A. FEASIBLE PATHS

Path planning for the vehicle is performed in the NED frame at a fixed depth of 2 meters. In planning, the vehicle state is given as  $s = \{x_h, p_n, p_e\}$  where  $x_h$  gives the heading in radians and  $p_n$  and  $p_e$  are the North and East components of the position of the state in meters from the NED origin. The set of next states available from any current state are defined by the following discrete dynamics

$$x_h[t+1] = x_h[t] + a \quad (19)$$

$$p_n[t+1] = p_n[t] + d_m \cos(x_h[t] + a) \quad (20)$$

$$p_e[t+1] = p_e[t] + d_m \sin(x_h[t] + a) \quad (21)$$

where  $a \in \mathcal{A}$  is an input angle in radians and  $d_m$  is the desired euclidean distance in meters between path states. Throughout these experiments,  $d_m = 30$  meters and  $\mathcal{A} = \{-90, -45, -15, 0.0, 15, 45, 90\}$  with the values of  $\mathcal{A}$  given

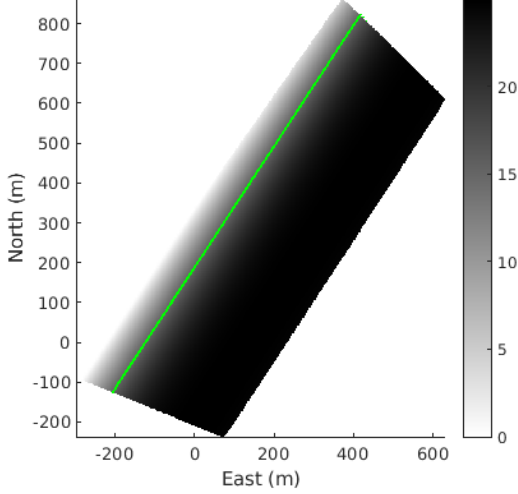


Fig. 3: Prior mean depth map with 15 meter isobath shown as a green line.



Fig. 4: Claytor lake operational area.

in degrees for convenience. Let  $p[t] = [p_n[t], p_e[t]]^\top$ . To traverse a path, the 690 vehicle tracks a line from  $p[t]$  to  $p[t+1]$  until it reaches  $p[t+1]$  at which point it attempts to track the line from  $p[t+1]$  to  $p[t+2]$ . By construction, the heading of the vehicle upon reaching a point in its path should be close to the desired heading  $x_h$ .

### B. TERMINAL REWARDS

The value of a path is found by appending a pair of lawn mower paths to the end of each short  $n$ -length path. The augmented paths are evaluated using (9), and the lawn mower path that maximizes the reward of the augmented path is selected to determine (13). Lawn mower paths are given as a sequence of straight lines parallel to the heading of the final state of the short  $n$ -length path and continuing outward on a single side at fixed intervals in the direction of the first turn

required to remain within the operational area. The value of the lawn mower path with the highest expected reward provides a lower bound on the performance of the receding horizon path.

### C. COMPUTATION OF REWARD ATTAINED

Note that equation (8) assumes that a single measurement will be recorded precisely at the location contained in each path state when, in reality, measurements are taken periodically by the DVL at a rate of 10Hz with a subset of those measurements being retained according to the methods described in Section II. In an attempt to ensure that several measurements of the lake depth are taken near each point of interest, we consider that a location has been sampled 3 seconds after the vehicle reaches the associated waypoint. The realized reduction in ambiguity is computed as the difference between the ambiguity at the point when first beginning the transition to the point and the ambiguity at the point 3 seconds after reaching the point. The choice of 3 seconds is based on the 690 vehicles average velocity of approximately 1.6 m/s.

### D. RECEDING HORIZON PATH PLANNING AND EXECUTION

In this work, we utilize a basic implementation of Monte Carlo Tree Search (MCTS) [17]–[19] with minor modifications to ensure that the lower bound requirements of Proposition 4 are satisfied. First, the reward term of the UCT is normalized using the value of the highest-reward rollout path found for the current planning epoch and is represented as  $\tilde{\mathcal{V}}_l(s_k)$ . Second, instead of back-propagating the average reward of a Monte Carlo simulation, we back-propagate the maximum  $\mathcal{V}_l(s_k)$  such that

$$\text{UCT} = \frac{\mathcal{V}_l(s_k)}{\tilde{\mathcal{V}}_l(s_k)} + C \sqrt{\frac{\ln(N)}{n}}. \quad (22)$$

Lastly, at the beginning of each planning epoch, the recomputed value of the remainder of the best path found in the previous planning epoch is compared against the value of the highest-reward lawn mower path starting from the root state. The path with the highest reward is saved as the initial highest-reward rollout path for the planning epoch such that the MCTS planner is guaranteed to satisfy (15) at each step. Furthermore, the comparison of the remainder of the path at each planning step against the best lawn mower path from the same initial state ensures that the guarantee of Proposition 4 is extended to

$$J(\tilde{\gamma}_l(s_k)) \geq B_l(s_k) \quad (23)$$

at each step  $k$  such that the theoretical lower bound for the receding horizon path produced is

$$J(\tilde{\gamma}_l(s_0)) \geq \max_{k \in \{1, \dots, l\}} [J(\tilde{\gamma}_k(s_0)) + B_l(s_k)] \quad (24)$$

as discussed in Section 7 of [20].

Defining a step to be a line segment connecting consecutive states in a path. The planning horizon is 5 steps, the execution horizon is a single step, and the total mission length  $l$  is

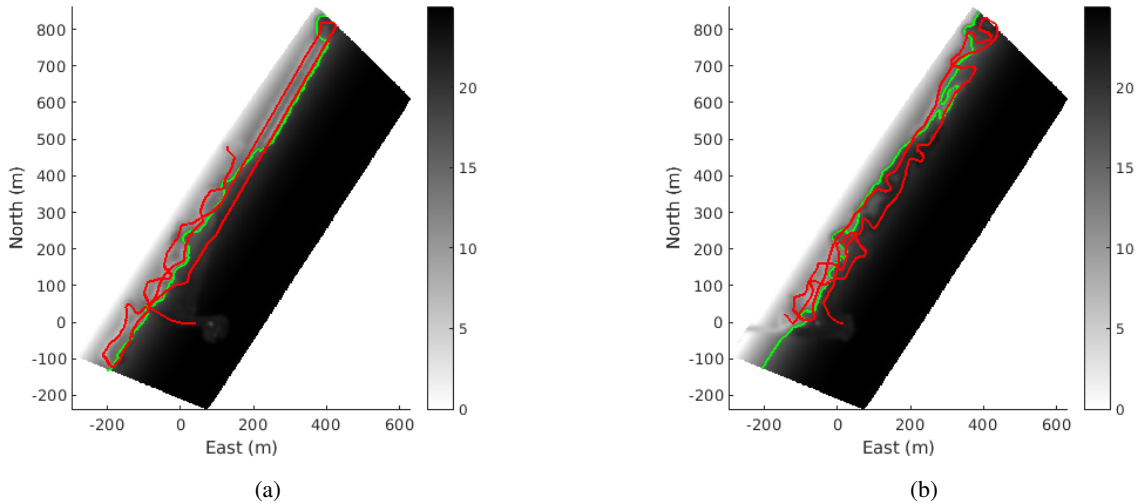


Fig. 5: The posterior predictive depth map within the bounded search region given the measurements acquired while executing the receding horizon path shown in red produced using terminal rewards (a) and the receding horizon path produced without the addition of terminal rewards (b). The 15 meter isobath is shown in green.

100 steps. For comparison, one path was planned using the proposed receding horizon method with terminal rewards and another receding horizon was planned without the addition of terminal rewards. The receding horizon path produced using the proposed method is shown as the solid red line in Figure 5a. The receding horizon path produced without the use of terminal rewards is similarly shown in Figure 5b.

## VII. RESULTS AND CONCLUSIONS

The results of planning and executing receding horizon paths are shown in Figures 5 and 6. Figure 5 shows the

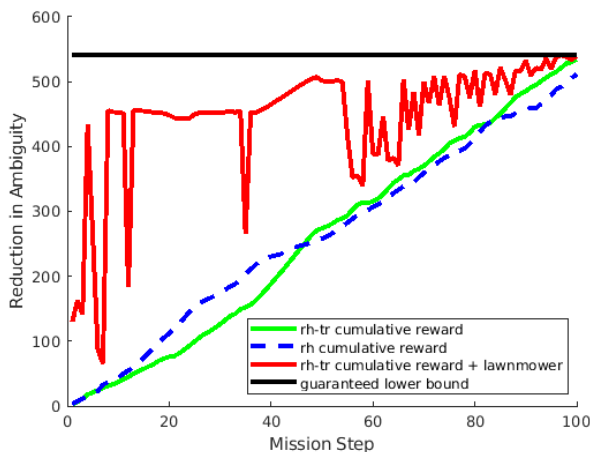


Fig. 6: Cumulative rewards for the RH path and RH path constructed with terminal rewards are shown in dashed blue and solid green respectively. The cumulative reward for the RH path constructed with terminal rewards plus the lower bound on the value-to-go is shown as the solid red line. The guaranteed lower bound is shown as the solid black line.

posterior depth maps produced using equations (3) and (4) and the data sets acquired while traversing the receding horizon paths according to the methods discussed in Section II. Figure 5 also shows the 15 meter isobath in green. The cumulative reward for the receding horizon path produced without the use of terminal rewards is shown as the dashed blue line in Figure 6 with the cumulative reward for the receding horizon path produced with terminal rewards shown as the solid green line. The solid black line shows the value of the lower bound given in equation 24 which is the maximum value of the solid red line. The red line gives the sum of the current cumulative reward attained by the receding horizon path up to step  $k$  plus the lower bound on the value-to-go. The red line is not monotone because it depends on the value of a naive lawnmower path starting from the agent's current state and not the value of the augmented short horizon path. Also, it is important to note that the red line includes a deterministic component due to the anticipated reward, while the blue and green lines represent only the stochastic actual reward. Thus, the green line terminating slightly below the black line may be comfortably attributed to the stochastic nature of the reward.

Notably, Figure 6 shows that both receding horizon paths performed significantly better than the value of the initial lower bound  $B_l(s_0)$  (the initial value of the red line) with the receding horizon path produced using the proposed methods performing slightly better than the receding horizon path that did not include terminal rewards. While there is no guarantee that paths produced using the proposed methods will perform better or worse than receding horizon paths produced using other methods, these results do not indicate any degradation in performance for including terminal rewards defined by naive heuristic paths. Furthermore, the results of these experiments appear to confirm the results of Proposition 4.

## REFERENCES

- [1] S. McCammon, G. Marcon dos Santos, M. Frantz, T. Welch, G. Best, R. K. Shearman, J. D. Nash, J. A. Barth, J. A. Adams, and G. A. Hollinger, "Ocean front detection and tracking using a team of heterogeneous marine vehicles," *Journal of Field Robotics*, vol. 38, no. 6, pp. 854–881, 2021.
- [2] S. Frolov, B. Garau, and J. Bellingham, "Can we do better than the grid survey: Optimal synoptic surveys in presence of variable uncertainty and decorrelation scales," *Journal of Geophysical Research: Oceans*, vol. 119, no. 8, pp. 5071–5090, 2014.
- [3] Y. Zhang, J. G. Bellingham, J. P. Ryan, B. Kieft, and M. J. Stanway, "Autonomous four-dimensional mapping and tracking of a coastal upwelling front by an autonomous underwater vehicle," *Journal of Field Robotics*, vol. 33, no. 1, pp. 67–81, 2016.
- [4] M. Dunbabin and A. Grinham, "Quantifying spatiotemporal greenhouse gas emissions using autonomous surface vehicles," *Journal of field robotics*, vol. 34, no. 1, pp. 151–169, 2017.
- [5] A. Singh, "Nonmyopic adaptive informative path planning for multiple robots," 2009.
- [6] B. Schlotfeldt, D. Thakur, N. Atanasov, V. Kumar, and G. J. Pappas, "Anytime planning for decentralized multirobot active information gathering," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1025–1032, 2018.
- [7] H. Yetkin, J. McMahan, N. Topin, A. Wolek, Z. Waters, and D. J. Stilwell, "Online planning for autonomous underwater vehicles performing information gathering tasks in large subsea environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6354–6361, IEEE, 2019.
- [8] Y. Kantaros, B. Schlotfeldt, N. Atanasov, and G. J. Pappas, "Asymptotically optimal planning for non-myopic multi-robot information gathering," *Proceedings of Robotics: Science and Systems (RSS), Freiburg, Germany*, pp. 22–26, 2019.
- [9] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Information acquisition with sensing robots: Algorithms and error bounds," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6447–6454, IEEE, 2014.
- [10] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Decentralized active information acquisition: Theory and application to multi-robot slam," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4775–4782, IEEE, 2015.
- [11] J. Wakulicz, K. M. B. Lee, C. Yoo, T. Vidal-Calleja, and R. Fitch, "Informative planning for worst-case error minimisation in sparse gaussian process regression," *arXiv preprint arXiv:2203.03828*, 2022.
- [12] B. Biggs, D. J. Stilwell, H. Yetkin, and J. McMahan, "Performance guarantees for receding horizon search with terminal cost," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6362–6368, IEEE, 2019.
- [13] B. Biggs, D. J. Stilwell, and J. McMahan, "Extended performance guarantees for receding horizon search with terminal cost," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6741–6748, IEEE, 2020.
- [14] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
- [15] A. Gotovos, "Active learning for level set estimation," Master's thesis, Eidgenössische Technische Hochschule Zürich, Department of Computer Science, 2013.
- [16] B. Bryan, R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman, "Active learning for identifying function threshold boundaries," *Advances in neural information processing systems*, vol. 18, 2005.
- [17] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in games*, vol. 4, no. 1, pp. 1–43, 2012.
- [18] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," in *European conference on machine learning*, pp. 282–293, Springer, 2006.
- [19] L. Kocsis, C. Szepesvári, and J. Willemsen, "Improved monte-carlo search," *Univ. Tartu, Estonia, Tech. Rep.*, vol. 1, 2006.
- [20] B. Biggs, J. McMahan, P. Baldoni, and D. J. Stilwell, "Multi-agent receding horizon search with terminal cost," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9086–9092, IEEE, 2021.