Bayesian Learning Model Predictive Control for Process-Aware Source Seeking

Yingke Li, *Graduate Student Member, IEEE*, Tianyi Liu, Enlu Zhou, and Fumin Zhang, *Senior Member, IEEE*

***Abstract*—Classical source seeking algorithms aim to make the robot reach the source location eventually. This letter proposes a process-aware source seeking approach which finds an informative trajectory to reach the source location. A multi-objective optimization problem is formulated based on rewards for both the search process and the terminal condition. Due to the unknown source location, solutions are found through Bayesian learning model predictive control (BLMPC). The consistency of the Bayesian estimator, as well as the convergence of the proposed algorithm are proved. The performance of the algorithm is evaluated through simulation results. The process-aware source seeking algorithm demonstrates improvements over other classical source seeking algorithms.**

***Index Terms*—Iterative learning control, optimization, predictive control for nonlinear systems, stochastic optimal control.**

I. Introduction

A MOBILE robot is often tasked with the goal of locating the source of an unknown signal field, such as magnetic force, heat, wireless signal, or chemical concentration. Source seeking algorithms may have a wide range of potential applications, including natural resource development, search & rescue, wireless communication, and nuclear threat reconnaissance [1]–[3]. Typical source seeking algorithms aim to find the global maximum of the source field, i.e., the source location. Many gradient-based [3]–[5] and model-based [1], [6]–[8] optimization techniques are utilized to solve the problem. Most algorithms aim to guarantee that the robot eventually reaches the source location, but do not explicitly take the optimality of the search process into consideration.

We formulate a process-aware source seeking problem to compute an informative trajectory to reach the source location. We are inspired by the *optimal search theory* [9] developed to search for targets that might not emit any signal. The optimal search process maximizes the probability of detecting a target subject to resource constraints. The process-aware source seeking problem combines the objectives of optimal search and source seeking as a multi-objective optimization problem, which aims to utilize the search process detection information, not just the signal gradient, to guide the process of source seeking.

The main challenge of the process-aware source seeking approach is that the location of the source is unknown, leading to great difficulty in solving the multi-objective optimization problem. We address this problem with the Bayesian learning model predictive control (BLMPC) [10], [11], where the probability distribution of the unknown source location can be learned online by a Bayesian estimator. The estimator is updated with measurements that are collected sequentially. We design a new search reward structure for the BLMPC that incorporates the search reward and the reward of reaching the source. We also propose a method to determine the planning horizon of the BLMPC to reduce the number of measurements required by the Bayesian estimator.

For offline BLMPC, bounded cumulative regret has been proved [10], [11]. While for online BLMPC, we focus on the convergence of one task execution which relies on the consistency of the Bayesian estimator. Bayesian consistency has been well studied when independent and identically distributed (i.i.d.) measurements are used [12]. We prove Bayesian consistency under non-i.i.d. (correlated and differently distributed) measurements. To the best of our knowledge, this result does not exist in the literature. We further prove that the robot trajectory is guaranteed to converge to the desired source location in finite time.

One especially challenging task in source seeking is in turbulent source field where signal information is sparse and expensive. For example, the robot often needs to stay at a certain location for a long period of time to collect enough signal strength for measurement. We perform simulations in the scenario of turbulent field where gradient-based algorithms cannot apply. We compare our proposed algorithm with another

Manuscript received March 3, 2021; revised April 29, 2021; accepted May 17, 2021. Date of publication June 3, 2021; date of current version June 29, 2021. This work was supported in part by the Office of Naval Research (ONR) under Grant N00014-19-1-2556 and Grant N00014-19-1-2266; in part by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-19-1-0283; in part by NSF under Grant CNS-1828678, Grant S&AS-1849228, and Grant GCR-1934836; and in part by the National Oceanic and Atmospheric Administration (NOAA) under Grant NA16NOS0120028. Recommended by Senior Editor M. Guay. *(Corresponding author: Fumin Zhang.)*

Yingke Li and Fumin Zhang are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30318 USA (e-mail: yli3225@gatech.edu; fumin@gatech.edu).

Tianyi Liu and Enlu Zhou are with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30318 USA (e-mail: tliu341@gatech.edu; ezhou30@gatech.edu).

Digital Object Identifier 10.1109/LCSYS.2021.3085852

gradient-free algorithm, the *expected rate algorithm* [7]. The results show that our proposed algorithm can greatly reduce the number of measurements and thus reduce the total search time when measurements are costly, which is often the case in turbulent signal field.

II. Background

Consider a robot seeking a signal source in a workspace $\mathcal{X}$ that is discretized into grid cells. Let $\theta^* \in \mathcal{X}$ be the true source location, which is unknown to the robot. An estimation of θ^* is represented by a random variable $\theta \in \mathcal{X}$ which obeys a probability distribution $\pi : \mathcal{X} \to [0, 1]$ such that $\pi(\theta)$ is the probability that the source is in cell θ.

Assumption 1: The true distribution for a single source is $\delta_{\theta^*}(\theta)$ where $\delta_{\theta^*}(\theta^*) = 1$ and $\delta_{\theta^*}(\theta) = 0$ for $\theta \neq \theta^*$.

We review some relevant results in source seeking and optimal target search.

Source Seeking: The source is able to emit certain signal that can be measured by the robot. The *signal field* is represented by a scalar function $R_{\theta^*} : \mathcal{X} \to [0, \infty)$ modeling the spatial signal distribution over the workspace. Consider an arbitrary location x and the source location θ^*, $R_{\theta^*}(x)$ reaches its global maximum when $x = \theta^*$. For example, a simple model $R_{\theta^*}(x) = \exp(-\|x-\theta^*\|)$ can be used to model decreasing sensor effectiveness as a function of distance between x and θ^*.

The problem of source seeking can be formulated as to simultaneously estimate the source field and find its global maximizer:

$$\max_{x \in \mathcal{X}} R_{\theta^*}(x). \tag{1}$$

Optimal Target Search: In optimal search theory, the robot will search for targets that do not necessarily emit any signal, given the probability distribution of the target.

We assume that when the robot visits the cell where the target is located, it is able to identify the target with certain probability $\alpha \in [0, 1]$. This leads to the introduction of the detection function $\beta : \mathcal{X} \times \{0, 1, 2, \ldots\} \to [0, 1]$. Given that the target is in cell x, $\beta(x, m(x))$ is the probability of failing to detect the target on the first $m(x) - 1$ visits in cell x and succeeding on the $m(x)$th visit [9]. For example, let α be the probability of detecting the target on a single visit in cell x given that the target is in cell x. Suppose each visit has an independent probability of finding the target, then $\beta(x, m(x)) = \alpha(1 - \alpha)^{m(x)-1}$.

Definition 1: If the target location θ follows the distribution π ($\theta \sim \pi$), the probability of detection for any arbitrary cell x is defined as

$$p_{\theta\sim\pi}(x) = \pi(x)\beta(x, m(x)), \tag{2}$$

where $m(x) - 1$ is the number of visits that cell x has been visited by the robot.

A cost function $c : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is often introduced such that $c(x_i, x_{i+1})$ is the cost for the robot to move from cell x_i to cell x_{i+1}.

The problem of optimal target search can be formulated as to find the optimal trajectory $X = [x_0, \ldots, x_N]$ that maximizes the probability of detecting the target along the trajectory with a given cost constraint:

$$\max_{X} \sum_{i=1}^{N} p_{\theta\sim\pi}(x_i) \quad \text{s.t.} \quad \sum_{i=0}^{N-1} c(x_i, x_{i+1}) \leq C, \tag{3}$$

where C is the given cost constraint.

III. Problem Formulation

We formulate the **process-aware source seeking** problem as a multi-objective optimization problem. One objective, inherited from **source seeking**, is to estimate the source location and reach the source eventually. Another objective, inspired by **optimal target search**, is to utilize previous search process detection information to refine the succeeding search process.

As is common in source seeking problem, we assume that the robot is equipped with sensors to observe the environment and can estimate the source location distribution $\pi(\theta)$ through signal field measurements. Different from typical source seeking problem that ignores the measurement cost, we explicitly introduce another cost function $s : \mathcal{X} \to [0, \infty)$ to model the measurement cost such that $s(x)$ is the cost for the robot to take a measurement at cell x. An information gain function $q : \mathcal{X} \to [0, \infty)$ is also introduced such that $q(x)$ represents the reduced uncertainty on the source location distribution when taking a measurement at cell x. The indicator variable $y \in \{0, 1\}$ is used to indicate if a measurement is taken, where $y = 1$ means 'yes' and $y = 0$ means 'no'.

Formulate the robot dynamics in discrete time as

$$x_{i+1} = f(x_i, u_i), \tag{4}$$

in which $x_i, x_{i+1} \in \mathcal{X}$ are the robot locations at time i and $i + 1$, $u_i \in \mathcal{U}$ is the applied input at time i. We assume that within one step, the robot can only move to its adjacent cells or stay at the current cell. Then the set of all possible control inputs $\mathcal{U}$ contains the corresponding inputs that drive the robot to nearby cells or remain in the current cell.

Consider the robot trajectory $X = [x_0, \ldots, x_N]$ controlled by the control input $U = [u_0, \ldots, u_{N-1}]$, with the measurement decision $Y = [y_0, \ldots, y_N]$. The process-aware source seeking problem is formulated as to maximize the overall detection and information rewards during the trajectory under certain cost constraint, with the terminal objective of reaching the source:

$$\begin{aligned} \max_{U,Y} \quad & \sum_{i=1}^{N} p_{\theta\sim\pi}(x_i) + \sum_{i=1}^{N} y_i q(x_i) \\ \text{s.t.} \quad & x_{i+1} = f(x_i, u_i), \\ & x_N = \arg\max_{x \in \mathcal{X}} R_{\theta^*}(x), \\ & \sum_{i=0}^{N-1} c(x_i, f(x_i, u_i)) + \sum_{i=1}^{N} y_i s(x_i) \leq C, \end{aligned} \tag{5}$$

where C is total cost constraint.

This problem contains the distributional estimation of the source location $\pi(\theta)$ which needs to be updated and solved when new measurements are collected. We consider to utilize the Bayesian learning MPC framework to find a feasible solution of it in practice.

IV. Bayesian Learning MPC

We utilize Bayesian learning to learn the source location distribution and propose to break down this long horizon optimal control problem into short horizon MPC problems by measurements. In this way, we decompose the decisions of Y and U and propose an iterative algorithm which is composed of three steps: Bayesian estimation, planning horizon decision and learning-based MPC.

A. Bayesian Estimation

There are many types of sensors that can be used to measure the source field. Here we introduce an abstract sensor model that can be adapted to different types of sensors. The sensor measurement collected at cell x is then modeled as a noisy observation $z(x)$ of the signal function $R_\theta(x)$. For example, if the target is emitting a signal, and $z(x)$ represents the number of such signals being detected within a given time Δt. We assume that z can be modeled as a random variable of Poisson distribution with $R_\theta(x)\Delta t$ as the rate parameter. The measurement model is then

$$p_\theta(z|x) = \frac{\exp(-R_\theta(x)\Delta t)(R_\theta(x)\Delta t)^z}{z!}. \tag{6}$$

Let $z_{1:k}$ be the measurements by the robot at locations $x_{1:k}$. We further assume that the sensor is memoryless, which means that the measurements are conditionally independent.

Assumption 2:

$$p(z_k|\theta, z_{1:k-1}, x_{1:k}) = p(z_k|\theta, x_k) = p_\theta(z_k|x_k). \tag{7}$$

Note that the memoryless assumption for sensor models are commonly used in the literature for target search [1], [2], [6], [7].

We define $\pi_k(\theta) = p(\theta|z_{1:k}, x_{1:k})$ as the posterior distribution of source location after k iterations. Therefore,

$$p(\theta|z_{1:k-1}, x_{1:k}) = p(\theta|z_{1:k-1}, x_{1:k-1}) = \pi_{k-1}(\theta), \tag{8}$$

followed by the fact that moving to a new location does not affect the information on the source location until a measurement is taken. Using the conditional independence assumption for the sensor model, and applying the Bayesian rule, we obtain

$$\pi_k(\theta) = \frac{\pi_{k-1}(\theta)p_\theta(z_k|x_k)}{\sum_\theta \pi_{k-1}(\theta)p_\theta(z_k|x_k)}. \tag{9}$$

The computation of the posterior distribution of the current iteration $\pi_k(\theta)$ requires the posterior distribution of the previous iteration $\pi_{k-1}(\theta)$ and the measurement observed at the current iteration k. We will use $\pi_k(\theta)$ as a surrogate for $\pi(\theta)$ to solve the problem (5).

B. Planning Horizon

At each step, the robot can choose if a measurement is taken based on the indicator variable y. In practice, the measurements in neighboring cells are often highly correlated, and frequent measurements may not be necessary. Since the long horizon problem is broken down into short horizon problems by measurements, the planning horizon for each iteration determines how frequently measurements should be taken. In each planning window, the robot takes a measurement at the first step, and proceeds without taking new measurements in the next I_k steps.

The posterior $\pi_k(\theta)$ is a valid surrogate for $\pi(\theta)$ until a new measurement is taken. Consider entropy $H(\pi_k(\theta)) = -\sum_{\theta\in\mathcal{X}} \pi_k(\theta)\log \pi_k(\theta)$, which is a measure of the average uncertainty of the estimated source location distribution $\pi_k(\theta)$. We choose $q(\cdot) \propto H(\pi_k(\theta))$ to be the information gain, $s(\cdot) = S$ to be the cost for taking a measurement. Since the information gain will benefit the following planning for I_k steps, the intuition of leveraging measurement benefits and cost leads to the following choice of planning horizon I_k:

$$I_k = \left\lfloor \frac{\kappa S}{H(\pi_k(\theta))} \right\rfloor, \tag{10}$$

where κ is a design parameter chosen according to S, and $\lfloor\cdot\rfloor$ is the floor function such that $\lfloor a \rfloor$ gives the closest integer less than or equal to a.

C. Learning-Based MPC

The MPC problem can be viewed as an approximation of the optimal control problem, which introduces a terminal reward to address the truncated remaining part. Therefore, we transfer the terminal constraint $x_N = \arg\max_{x\in\mathcal{X}} R_{\theta^*}(x)$ into the terminal reward $T_{\theta\sim\pi_k}(x) = \mathbb{E}_{\theta\sim\pi_k} R_\theta(x)$.

Whenever a new measurement is taken, we initialize a new planning window and solve an MPC problem with I_k as the planning horizon as follows:

$$\begin{aligned} \max_{U_k} J_k = &\sum_{i=1}^{I_k} p_{\theta\sim\pi_k}(x_{i|k}) + \lambda T_{\theta\sim\pi_k}(x_{I_k|k}) \\ \text{s.t.} \quad & x_{i+1|k} = f(x_{i|k}, u_{i|k}), \\ & \sum_{i=0}^{I_k-1} c(x_{i|k}, f(x_{i|k}, u_{i|k})) \le C_k \end{aligned} \tag{11}$$

where π_k is the posterior distribution of the source location after k measurements, $x_{i|k} \in \mathcal{X}$ is the i-step-ahead prediction of the robot location initialized at $x_{0|k}$, $u_{i|k} \in \mathcal{U}$ is the corresponding i-step-ahead control input, I_k is the planning horizon, C_k is the cost allocated to the kth iteration and λ is a parameter that balances the search process reward and terminal reward functions, which can often be chosen from $0.6 \sim 1.5$ in practice.

A control policy is obtained by solving the MPC problem (11). Since the number of possible choices for each $u_{i|k}$ is finite. The computation of the MPC policy U_k^* can be viewed as a path planning problem that can be solved by algorithms such as the scenario tree search. The process-aware source seeking algorithm based on Bayesian learning MPC is summarized as Algorithm 1.

V. Convergence Analysis

Under the assumption of single source, i.e., $\pi^*(\theta) = \delta_{\theta^*}(\theta)$, we can justify the convergence of the process-aware source seeking algorithm (Algorithm 1).

Algorithm 1: Process-Aware Source Seeking

Initialize prior $\pi_0(\theta)$ with uniform probability distribution over the discretized location space $\mathcal{X}$;
Initialize robot position x_1;
Initialize iterator $k = 1$;
while *source location not found* **do**
  Take a measurement z_k at the current sample location x_k;
  Compute posterior distribution $\pi_k(\theta)$ using (9);
  Decide the planning horizon l_k using (10);
  Obtain the optimal control policy by solving (11) and move to the next sample location x_{k+1};
  $k := k + 1$;
end

A. Consistency of the Bayesian Estimator

Definition 2: The Bayesian estimator is (strongly) consistent if $\pi_k(\theta)$, the estimated distribution of θ, converges to the true distribution $\delta_{\theta^*}(\theta)$, with probability 1.

To prove the consistency of the Bayesian estimator, we first prove Lemma 1 based on Assumption 3.

Assumption 3: The prior distribution $\pi_0(\theta)$ has non-zero probability at θ^*.

Lemma 1: The marginal likelihood of measurement $\hat{p}_k(z|x) = E_{\theta\sim\pi_k(\theta)}[p_\theta(z|x)] = \sum_\theta \pi_k(\theta)p_\theta(z|x)$ converges to the true measurement model, i.e., for any z, we have $\lim_{k\to\infty} \hat{p}_k(z|x) = p_{\theta^*}(z|x)$, with probability 1.

Proof: Let $\mathcal{F}_k = \sigma\{(x_s, z_s), s \le k\}$ be the $\sigma-$filtration [13] generated by the past sample locations and measurements. According to Eqn. (9), the estimated probability of the true source location satisfies the following equation,

$$\log \pi_k(\theta^*) = \log \pi_{k-1}(\theta^*) + \log \frac{p_{\theta^*}(z_k|x_k)}{\hat{p}_{k-1}(z_k|x_k)}.$$

Taking expectation on both sides, we have that

$$\begin{aligned}\mathbb{E}[\log \pi_k(\theta^*)] &= \mathbb{E}[\log \pi_{k-1}(\theta^*)] + \mathbb{E}\left[\log \frac{p_{\theta^*}(z_k|x_k)}{\hat{p}_{k-1}(z_k|x_k)}\right] \\ &= \mathbb{E}[\log \pi_{k-1}(\theta^*)] \\ &\quad + \mathbb{E}\left[\mathbb{E}\left[\log \frac{p_{\theta^*}(z_k|x_k)}{\hat{p}_{k-1}(z_k|x_k)}|x_k, \mathcal{F}_{k-1}\right]\right] \\ &= \mathbb{E}[\log \pi_{k-1}(\theta^*)] + \mathbb{E}[d_{k-1}],\end{aligned}$$

where $d_{k-1} = D_{\mathrm{KL}}(p_{\theta^*}(z|x_k)\|\hat{p}_{k-1}(z|x_k))$ is the *relative entropy* (*Kullback–Leibler divergence*) between $p_{\theta^*}(z|x_k)$ and $\hat{p}_{k-1}(z|x_k)$. Then the expectation of d_{k-1} can be represented as follows,

$$\mathbb{E}[d_{k-1}] = \mathbb{E}[\log \pi_k(\theta^*)] - \mathbb{E}[\log \pi_{k-1}(\theta^*)].$$

For any n, taking summation over k from 1 to n on both sides, we have

$$\sum_{k=1}^{n} \mathbb{E}[d_{k-1}] = \mathbb{E}[\log \pi_n(\theta^*)] - \log \pi_0(\theta^*) \le -\log \pi_0(\theta^*) < \infty,$$

where the last inequality holds according to Assumption 3.

Therefore, take $n \to \infty$ and we get

$$\sum_{k=1}^{\infty} \mathbb{E}[d_{k-1}] \le -\log \pi_0(\theta^*) < \infty.$$

By *Markov Inequality*, we know that for any $\epsilon > 0$,

$$\sum_{k=0}^{\infty} P[d_k \ge \epsilon] \le \frac{1}{\epsilon}\sum_{k=0}^{\infty} \mathbb{E}[d_k] < \infty.$$

We can then apply *Borel-Cantelli Lemma* and show that $P(d_k \ge \epsilon, i.o.) = 0$, which further implies $\lim_{k\to\infty} d_k = 0$, with probability 1 for location-measurement sequences.

Moreover, since $d_k \ge 0$, by *Tonelli's Theorem*, we have

$$\mathbb{E}\left[\sum_{k=0}^{\infty} d_k\right] = \sum_{k=0}^{\infty} \mathbb{E}[d_k] \le -\log \pi_0(\theta^*).$$

Since $\sum_{k=0}^{\infty} d_k$ has bounded expectation, it must be finite with probability 1.

Note that the *total variation distance* between two distributions is related to the *relative entropy* by *Pinsker's Inequality*:

$$\|p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)\|_{TV} \le \sqrt{2d_{k-1}},$$

where

$$\|p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)\|_{TV} = \sup_z |p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)|.$$

Letting $k \to \infty$, by the convergence of d_k, we have

$$\lim_{k\to\infty} \int_z |p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)| dz = 0,$$

with probability 1.

According to *Dominated Convergence Theorem*, we further have

$$\int_z \lim_{k\to\infty} |p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)| dz = 0.$$

Moreover, since $|p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)| \ge 0$ and $p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)$ is continuous in z, then for any z,

$$\lim_{k\to\infty} |p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k)| = 0,$$

which means

$$\lim_{k\to\infty} \hat{p}_k(z|x_{k+1}) = p_{\theta^*}(z|x_{k+1}), \tag{12}$$

with probability 1. ■

Now we prove that the estimated distribution of θ converges to the true distribution $\delta_{\theta^*}(\theta)$ based on Assumption 4.

Assumption 4: For any (x, z), $p_\theta(z|x)$ is linearly independent, i.e., for $\forall n > 0$, if

$$c_1 p_{\theta_1}(z|x) + c_2 p_{\theta_2}(z|x) + \cdots + c_n p_{\theta_n}(z|x) = 0$$

holds for all (x, z), then $c_1 = c_2 = \cdots = c_n = 0$.

Note that this assumption requires that the measurements from different source locations can be eventually distinguished, which is usually satisfied in practice.

Theorem 1: The estimated distribution of θ converges to the true distribution $\delta_{\theta^*}(\theta)$, i.e., $\lim_{k\to\infty} \pi_k(\theta) = \delta_{\theta^*}(\theta)$, with probability 1.

Proof: Note that

$$p_{\theta^*}(z|x_k) - \hat{p}_{k-1}(z|x_k) = [1 - \pi_{k-1}(\theta^*)]p_{\theta^*}(z|x_k) + \sum_{\theta \neq \theta^*} \pi_{k-1}(\theta)p_\theta(z|x_k). \quad (13)$$

Since $\pi_k(\theta)$ is bounded, there exists a convergent sub-sequence $\{\pi_{t_1}(\theta), \pi_{t_2}(\theta), \ldots\}$ which converges to $\pi_\infty(\theta)$. Since x_k is also bounded, we could take a further sub-sequence $\{x_{\tau_1}, x_{\tau_2}, \ldots\}$ which converges to x_∞. Then take limit over Eqn. (13) along $\{\tau_1, \tau_2, \ldots\}$ and we have

$$[1 - \pi_\infty(\theta^*)]p_{\theta^*}(z|x_\infty) + \sum_{\theta \neq \theta^*} \pi_\infty(\theta)p_\theta(z|x_\infty) = 0,$$

with probability 1.

According to Assumption 4, then for any convergent sub-sequence, $1 - \pi_\infty(\theta^*) = 0, \pi_\infty(\theta) = 0 \quad \forall \theta \neq \theta^*$, which further implies

$$\lim_{k\to\infty} \pi_k(\theta) = \delta_{\theta^*}(\theta), \quad (14)$$

with probability 1. ■

Therefore, the consistency of the Bayesian estimator is proved. It is worth noting that the consistency result here does not require i.i.d. measurements that are required by the consistency results in literature such as [12].

B. Convergence of Robot Movements

Based on the consistency guarantee of Theorem 1, we can prove the convergence of our proposed process-aware source seeking algorithm.

Assumption 5: The resource is sufficient, i.e., the cost constraint in (11) will not be violated.

Definition 3: Define the reachable region $\mathcal{S}_k$ as the minimal region that contains all possible robot locations for the trajectory at a certain stage k, i.e., $X_k = [x_{0|k}, \ldots, x_{I_k|k}] \subset \mathcal{S}_k$, for all possible X_k.

We first prove Lemma 2 and Lemma 3 based on Assumption 5.

Lemma 2: There exists a finite stage K, such that the source location θ^* is in the reachable region $\mathcal{S}_K$, with probability 1.

Proof: For simplicity, we will only consider the scenario when $\lim_{k\to\infty} \pi_k = \delta_{\theta^*}$ holds, which happens with probability 1 according to Theorem 1.

We denote the minimal gap as

$$\epsilon = \min_{x \neq x' \in \mathcal{X}} |T_{\theta\sim\delta_{\theta^*}}(x) - T_{\theta\sim\delta_{\theta^*}}(x')| > 0, \quad (15)$$

given the state space $\mathcal{X}$ is finite.

We prove the lemma by contradiction. Assume that for any finite stage k, the true source location θ^* is not in the reachable region $\mathcal{S}_k$, i.e., $x_{i|k} \neq \theta^*$ for all $i = 1, \ldots, I_k$. Since $\lim_{k\to\infty} \pi_k = \delta_{\theta^*}$, for any $\epsilon > 0$ and $M > 0$, there must exist $K > 0$, such that for $k > K$, we have

$$|\pi_k(x) - \delta_{\theta^*}(x)| \leq \lambda \frac{\epsilon}{6M}, \quad |T_{\theta\sim\pi_k}(x) - T_{\theta\sim\delta_{\theta^*}}(x)| \leq \frac{\epsilon}{6}.$$

Then we have $|\pi_k(x)| < \lambda \frac{\epsilon}{6M}$ for $x \neq \theta^*$, since the state space $\mathcal{X}$ is finite, we can always find finite M such that

$$\sum_{x \neq \theta^*} \pi_k(x) \leq \lambda \frac{\epsilon}{6}.$$

Therefore, the objective $J_k(X_k)$ is close enough to $\lambda T_{\theta\sim\delta_{\theta^*}}(x_{I_k|k})$, that is

$$\begin{aligned}
&|J_k(X_k) - \lambda T_{\theta\sim\delta_{\theta^*}}(x_{I_k|k})| \\
&= \left| \sum_{i=1}^{I_k} p_{\theta\sim\pi_k}(x_{i|k}) + \lambda T_{\theta\sim\pi_k}(x_{I_k|k}) - \lambda T_{\theta\sim\delta_{\theta^*}}(x_{I_k|k}) \right| \\
&\leq \left| \sum_{i=1}^{I_k} p_{\theta\sim\pi_k}(x_{i|k}) \right| + \left| \lambda T_{\theta\sim\pi_k}(x_{I_k|k}) - \lambda T_{\theta\sim\delta_{\theta^*}}(x_{I_k|k}) \right| \\
&\leq \lambda \frac{\epsilon}{6} + \lambda \frac{\epsilon}{6} = \lambda \frac{\epsilon}{3}.
\end{aligned}$$

Therefore, for any $X, X' \subset S_k$, if $T_{\theta\sim\delta_{\theta^*}}(x_{I_k}) < T_{\theta\sim\delta_{\theta^*}}(x'_{I_k})$,

$$\begin{aligned}
J_k(X) - J_k(X') &= J_k(X) - \lambda T_{\theta\sim\delta_{\theta^*}}(x) \\
&\quad + \lambda T_{\theta\sim\delta_{\theta^*}}(x) - \lambda T_{\theta\sim\delta_{\theta^*}}(x') \\
&\quad + \lambda T_{\theta\sim\delta_{\theta^*}}(x') - J_k(X') \\
&< \lambda \frac{\epsilon}{3} - \lambda\epsilon + \lambda \frac{\epsilon}{3} < 0.
\end{aligned}$$

Thus, the approximation will not change the order of the possible terminal states. In other words, we have the same optimal solution set of the following two optimization problems,

$$[\arg\max_{X\subset\mathcal{S}_k} J_k(X)]_{I_k} = \arg\max_{x\in\mathcal{S}_k} T_{\theta\sim\delta_{\theta^*}}(x).$$

Note that the update of our algorithm is given by

$$\begin{aligned}
x_{k+1} &= [\arg\max_{X\subset\mathcal{S}_k} J_k(X)]_{I_k} = \arg\max_{x\in\mathcal{S}_k} T_{\theta\sim\delta_{\theta^*}}(x) \\
&= \arg\max_{x\in\mathcal{S}_k} \mathbb{E}_{\theta\sim\delta_{\theta^*}} R_\theta(x) \\
&= \arg\max_{x\in\mathcal{S}_k} R_{\theta^*}(x).
\end{aligned}$$

This implies the monotonicity of $R_{\theta^*}(x_k)$, i.e., $R_{\theta^*}(x_{k+1}) \geq R_{\theta^*}(x_k), \forall k > 0$.

Since $\mathcal{X}$ is finite, $\{R_{\theta^*}(x_k)\}_k$ must be a bounded monotonic sequence and thus converges to some value R^*. Hence, there must exist $K > 0$, such that for $k > K$, we have

$$|R_{\theta\sim\delta_{\theta^*}}(x_k) - R_{\theta\sim\delta_{\theta^*}}(x_{k+1})| < \epsilon,$$

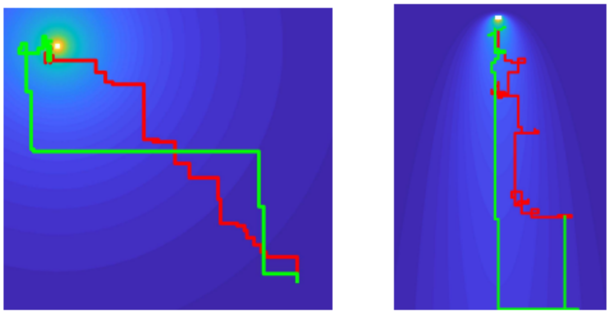
where ϵ is the minimal gap defined in (15).

Note that at each update, if the algorithm moves to a new location, $R_{\theta\sim\delta_{\theta^*}}(x_k)$ will change for at least ϵ. This observation implies that when $k > K$, we have $x_k = x^* \in \mathcal{X}$, where x^* is a local maximizer.

Since θ^* is the unique global maximizer, we have $\lim_{k\to\infty} x_k = \theta^*$, with probability 1. This contradicts the assumption that the true source location θ^* is not in the reachable region $\mathcal{S}_k$. Therefore, there exists a finite stage K, such that the true source location θ^* is in the reachable region $\mathcal{S}_K$, as long as $\lim_{k\to\infty} \pi_k = \delta_{\theta^*}$ holds, which happens with probability 1. ■

Lemma 3: For a large enough stage K, if the true source location θ^* is in the reachable region $\mathcal{S}_K$, the robot is guaranteed to reach the source location and stay there, with probability 1.

Proof: Recall that $J_k = \sum_{i=1}^{I_k} p_{\theta\sim\pi_k}(x_{i|k}) + \lambda T_{\theta\sim\pi_k}(x_{I_k|k})$. We know that when k is large, both $p_{\theta\sim\pi_k}(x)$ and $T_{\theta\sim\pi_k}(x)$ is maximized at θ^*. Therefore, the reward J_k is maximized



(a) Without wind. (b) With wind.

Fig. 1. Trajectories of our proposed process-aware algorithm (red) and the expected rate algorithm (green) in different scenarios. The colormap represents the mean rate of hits of the field.

TABLE I
COMPARISON BETWEEN PROCESS-AWARE ALGORITHM WITH EXPECTED RATE ALGORITHM. RESULTS ARE AVERAGED OVER 20 STOCHASTIC SIMULATIONS

Scenario	Algorithm	Trajectory Length	Measurements	Total Search Time
Without Wind	Process-aware	254.7	76.4	636.7
	Expected Rate	261.5	261.5	1569
With Wind	Process-aware	275.7	65.1	601.2
	Expected Rate	252.7	252.7	1516.2

when X_k contains as many θ^* as possible. This requires that the robot takes the shortest path to reach the source location and stay there. Therefore, the robot is guaranteed to reach the source location and stay there with probability 1. ■

Now we prove the convergence of our proposed algorithm.

Theorem 2: The robot movements controlled by the solution of the MPC problem (11) converges to the true source location θ^* in finite time, with probability 1.

Proof: From Lemma 2 and Lemma 3, we know that there exists a finite stage K such that the robot reaches the source location and stays there. Therefore, the solution to the MPC problem will generate a robot trajectory that converges to the source location. ■

VI. SIMULATION RESULTS

We consider a source of chemical plume, which generates plume particles in the 2D space. The field function is represented by the rate of hits, which is defined as the average number of particles per unit time measured by the sensor at a certain location. The rate of hits for a chemical plume source can be given as:

$$R_\theta(x) = \frac{R_s}{\log\frac{\gamma}{a}} \exp\Big(-\frac{\langle \theta - x, V\rangle}{2D}\Big) K_0\Big(\frac{\|\theta - x\|_2}{\gamma}\Big), \quad (16)$$

where R_s is the rate at which the plume source releases the plume particles in the environment, $\gamma = \sqrt{D\tau/(1 + \frac{\|V\|^2\tau}{4D})}$ is the average distance travelled by a plume particle in its life time, a is the size of the sensor detecting plume particles, V is the average wind velocity, D is the diffusivity of the plume particles and K_0 is the Bessel function of zeroth order.

Performance of the proposed algorithm is assessed by numerical simulations. We perform the simulation for a model of chemical plume where detectable particles are emitted at rate $R = 1$, have a lifetime $\tau = 2500$, and propagate with diffusivity $D = 1$. $V = [0, 0]$ in the absence of wind and $V = [0, 1]$ in the presence of wind. We assume the sensor size $a = 1$, the robot takes $\Delta t = 5s$ to take a measurement of the signal and $1s$ to move to the adjacent cells. For the algorithm, we choose $\alpha = 0.5$, $\lambda = 0.8$, $\kappa = 10$, $S = 5$ and assume the resource is sufficient, i.e., the cost constraint will not be violated.

We compare our proposed algorithm with another gradient-free algorithm, the *expected rate algorithm* [7]. Some instances of the trajectories are displayed in Fig. 1. The average trajectory length, number of measurements and total search time are demonstrated in Table I. As illustrated, our proposed algorithm can greatly reduce the number of measurements and thus reduce the total search time when measurements are costly, which is often the case in turbulent signal field.

VII. CONCLUSION AND FUTURE WORK

In this letter, we presented the process-aware source seeking algorithm, whose advantages were demonstrated by the analysis and simulation results. Our proposed method can contribute to multi-objective optimization problems by providing a general approach to jointly optimize the process and the goal. Future work may focus on extending this algorithm to the case of multiple sources and multiple robots.

REFERENCES

[1] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'Infotaxis' as a strategy for searching without gradients," *Nature*, vol. 445, pp. 406–409, Jan. 2007.

[2] J.-B. Masson, M. B. Bechet, and M. Vergassola, "Chasing information to search in random environments," *J. Phys. A Math. Theor.*, vol. 42, no. 43, 2009, Art. no. 434009.

[3] Z. Li, K. You, and S. Song, "AUV based source seeking with estimated gradients," *J. Syst. Sci. Complexity*, vol. 31, no. 1, pp. 262–275, 2018.

[4] E. Ramirez-Llanos and S. Martinez, "Stochastic source seeking for mobile robots in obstacle environments via the SPSA method," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1732–1739, Apr. 2019.

[5] J. Lin, S. Song, K. You, and M. Krstic, "Stochastic source seeking with forward and angular velocity regulation," *Automatica*, vol. 83, pp. 378–386, Sep. 2017.

[6] H. Hajieghrary, A. F. Tomas, and M. A. Hsieh, "An information theoretic source seeking strategy for plume tracking in 3D turbulent fields," in *Proc. IEEE Int. Symp. Safety Security Rescue Robot.*, 2016, pp. 1–8.

[7] V. Mishra and F. Zhang, "A stochastic optimization framework for source seeking with infotaxis-like algorithms," in *Proc. IEEE 55th Conf. Decis. Control*, 2016, pp. 6845–6850.

[8] J. R. Bourne, E. R. Pardyjak, and K. K. Leang, "Coordinated Bayesian-based bioinspired plume source term estimation and source seeking for mobile robots," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 967–986, Aug. 2019.

[9] L. D. Stone, J. O. Royset, and A. R. Washburn, *Optimal Search for Moving Targets*. Cham, Switzerland: Springer, 2016.

[10] K. P. Wabersich and M. N. Zeilinger, "Bayesian model predictive control: Efficient model exploration and regret bounds using posterior sampling," 2020. [Online]. Available: arXiv:2005.11744.

[11] K. P. Wabersich and M. N. Zeilinger, "Performance and safety of Bayesian model predictive control: Scalable model-based RL with guarantees," 2020. [Online]. Available: arXiv:2006.03483.

[12] S. M. Huttegger, "Bayesian convergence to the truth and the metaphysics of possible worlds," *Philos. Sci.*, vol. 82, no. 4, pp. 587–601, 2015.

[13] A. Klenke, *Probability Theory: A Comprehensive Course*. Berlin, Germany: Springer, 2014.