Concurrent Active Learning in Autonomous Airborne Source Search: Dual Control for Exploration and Exploitation

Zhongguo Li, *Member, IEEE*, Wen-Hua Chen, *Fellow, IEEE*, and Jun Yang, *Senior Member, IEEE*

***Abstract*—A concurrent learning framework is developed for source search in an unknown environment using autonomous platforms equipped with onboard sensors. Distinct from the existing solutions that require significant computational power for Bayesian estimation and path planning, the proposed solution is computationally affordable for onboard processors. A new concept of concurrent learning using multiple parallel estimators is proposed to learn the operational environment and quantify estimation uncertainty. The search agent is empowered with the dual capability of exploiting current-estimated parameters to track the source and probing the environment to reduce the impacts of uncertainty, namely Concurrent Learning based Dual Control for Exploration and Exploitation (CL-DCEE). In this setting, the control action not only minimizes the tracking error between future agent's position and estimated source location, but also the uncertainty of predicted estimation. More importantly, the rigorous proven properties, such as the convergence of CL-DCEE algorithm, are established under mild assumptions on noises, and the impact of noises on the search performance is examined. Simulation results are provided to validate the effectiveness of the proposed CL-DCEE algorithm. Compared with the information-theoretic approach, CL-DCEE not only guarantees convergence, but produces better search performance and consumes much less computational time.**

***Index Terms*—Autonomous search, dual control, exploration and exploitation, path planning, source search and estimation.**

I. Introduction

Identifying the sources of airborne release (including chemical, biological, radiological, and nuclear materials) is one of the most important tasks in disaster management and environment protection [1]. In the early literature, source term estimation (STE) is mainly supported by onsite measurement using static sensor networks that are deployed beforehand in some specific areas of potential risks [1], [2]. This type of strategy is very costly, and only feasible for high-risk industry, e.g., nuclear power plants [3]. Recently, significant research efforts have been dedicated to the development of dynamic estimation of airborne release assisted by mobile platforms, for example, autonomous ground robots [4], [5], [6] and unmanned aerial vehicles (UAVs) [7]. Compared with conventional static methods, autonomous search is much more flexible and cost effective in emergent accident management.

There are various methods dealing with this problem [1], [5]. Among them, informative path planning (IPP) becomes increasingly popular, e.g., Infotaxis [6] and Entrotaxis [8]. Vergassola et al. [6] proposed an informative search approach, referred as Infotaxis, by which the agent moves to the next position that is expected to minimize uncertainties of the posterior distribution. Hutchinson et al. [8] developed the Entrotaxis algorithm that steers the agent to search over the most uncertain area in the next movement. More recently, some advanced versions of the abovementioned algorithms have been developed aiming to improve their robustness, search speed, and accuracy in more complex search environment, including Infotaxis II [9] and Entrotaxis-jump [10]. Essentially, information-theoretic approaches aim to reduce uncertainties of estimated source location and unknown environment parameters. Therefore, the reward function is defined according to the information gain using some informative measures, for example, entropy, Kullback–Leibler divergence, variance, and Fisher information matrix [11], [12].

Apart from information-theoretic methods, another main branch for source seeking is the optimization approach. Stochastic extremum seeking is employed to direct a nonholonomic unicycle toward the maximum of an unknown signal field [13]. Simultaneous perturbation stochastic approximation approach is implemented for source search [14], [15], [16], which can be traced back to the early work in [17]. In particular, both centralized and distributed coordination algorithms are developed for model-based and model-free source seeking using network-connected mobile robots in [16]. An adaptive gradient climbing method is designed for cooperative mobile sensors to seek the optimizer of environmental field in [18]. It is worth mentioning that convergence guarantees of those algorithms have been well studied by leveraging advanced control and stochastic approximation techniques. In those learning-based control approaches, the unknown parameters of the environment are *passively* updated.

Recently, Chen et al. [5] have reformulated the autonomous search problem from a control-theoretic perspective, referred as Dual Control for Exploration and Exploitation (DCEE). The ultimate goal of autonomous search is to design a control strategy that can navigate the agent to an *unknown release* in an *unknown environment*, which is a well-posed goal-oriented control problem. Distinct from traditional control settings where operational systems are manipulated by following predefined references or setpoints, the autonomous search problem does not have such a given reference or path that can directly lead the agent to the source. Instead, the agent is required to *explore* the operational environment to learn the source parameters, and at the same time *exploit* its belief to move towards the source. This novel dual control framework achieves a natural balance between the two objectives, and has demonstrated superior performance in real experiments compared with model predictive control (MPC) and IPP. Although DCEE offers a conceptually promising framework in autonomous search, currently it still suffers from two drawbacks: computational burden and no rigorous analysis of its properties, such as stability and convergence. IPP and

Manuscript received 24 July 2022; revised 14 October 2022; accepted 6 November 2022. Date of publication 14 November 2022; date of current version 26 April 2023. This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Established Career Fellowship "Goal-Oriented Control Systems: Disturbance, Uncertainty and Constraints" under Grant EP/T005734/1. Recommended by Senior Editor Tetsuya Iwasaki and Guest Editors George J. Pappas, Anuradha M. Annaswamy, Manfred Morari, Claire J. Tomlin, Rene Vidal, and Melanie N. Zeilinger. *(Corresponding author: Zhongguo Li.)*

The authors are with the Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, U.K. (e-mail: z.z.li@lboro.ac.uk; w.chen@lboro.ac.uk; j.yang3@lboro.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3221907.

Digital Object Identifier 10.1109/TAC.2022.3221907

DCEE approaches demand massive computational burden imposed by nonlinear particle filters and optimization-based path planning [8], [19]. More specifically, the Bayesian inference engine is involved in the optimization loop for IPP since the influence of the control action on the predicted posterior of the estimated source and environment parameters is evaluated at each iteration. In IPP and DCEE, the agent's movement (path planning) and source estimation (environment acquisition) are strongly coupled: the agent takes actions according to the current estimation of source parameters and the estimators update their knowledge by using the concentration collected at the agent's new position determined by path planning. This coupling, together with noisy measurement, environment turbulence, complicated particle filtering, and optimization involved in the implementation of the search strategy, makes the rigorous analysis of theoretic properties of these search strategies quite challenging.

Inspired by the concept of DCEE, we propose a concurrent learning-based DCEE scheme with multiple estimators that encompasses dual effects: driving the agent to the believed location by exploiting current estimation, and reducing uncertainties by exploring the unknown operational environment, which is referred as Concurrent Learning based Dual Control for Exploration and Exploitation (CL-DCEE) for the sake of simplicity. The underlying principle of the concurrent learning scheme advocated in this work is distinct from the classic dual control in handling the two intricate coupling elements of the *system* and the *environment*. Existing dual control approaches impose a probing effect on the *system* itself, for example, state estimation in stochastic control [20] and parameter estimation in adaptive control [21]. On the other hand, the dual effect introduced in our formulation is used to explore the operational *environment*, as our objective is to acquire a better understanding of the unknown environment such that the agent is able to approach the true source location.

Two approaches are proposed to address the two challenges of computational burden and proven properties. Instead of implementing computationally demanding particle filtering, an efficient multiestimator scheme is proposed for source and environment learning. The number of estimators used in CL-DCEE is much smaller than that of particles required for Bayesian filters in information-theoretic algorithms. These estimators run in parallel from a set of randomly started initial estimates. There are several fundamental incentives promoting us to employ multiple concurrent estimators. First, compared with employing a single estimator, this multiestimator approach provides a means to quantify uncertainty associated with source estimators, which is of great importance to empower the search agent with *dual capability of exploration and exploitation*. Second, it significantly improves the performance and robustness over a single estimator. The performance of a single estimator (such as an observer or learning machine) is often severely influenced by the initialization and setting of the individual estimator. To the best of our knowledge, there are few results on multiestimator-assisted control algorithms. Devising multiple parallel estimators for the source parameters is conducive to eliminating undesirable behavior caused by random initialization of an individual, and also, it allows us to take advantage of a priori probability density function (PDF) of source parameters. To further reduce the computational load, effective gradient-based optimization algorithms are utilized to replace the complicated path planning process in the existing methods. It is shown that by combining these two techniques, we are able to reduce the computational load by 100 times while significantly increasing the admissible control set. Most importantly, we establish theoretical guarantee of convergence of the CL-DCEE algorithm.

In summary, the key contributions are enumerated as follows.

1) A concurrent active learning algorithm with multiple environment estimators is developed, which achieves a balanced tradeoff between *exploitation* of believed source location and *exploration* of uncertain environment, that is, simultaneously navigating the agent to the source and reducing the impacts of uncertainties associated with the acquired environment knowledge.
2) The convergence of the proposed autonomous search algorithm, CL-DCEE, is rigorously established under sensor and control noises, by leveraging a memory-based stochastic approximation and a gradient-based path planning strategy.
3) The proposed CL-DCEE provides a *computationally efficient* solution for autonomous search of airborne source release. Simulation results are provided to demonstrate the performance of the proposed method in comparison with information-theoretic approaches. Our solution shows superior performance with significant reduction in computational time.

The rest of this article is organized as follows. In Section II, we formulate the autonomous search problem and develop feasible value functions for the path planning and estimation. In Section III, CL-DCEE algorithm is proposed by deploying multiple environment estimators. Section IV provides simulation results and detailed discussions in comparison with existing approaches. Finally, Section V concludes this article.

II. Problem Formulation

A. Agent Modelling

The searching agent is considered as a fully autonomous vehicle, for example, a mobile robot or a UAV, which is equipped with chemical/biological sensors. We assume that the agent has been devised with a low-level controller that can steer the agent to the desired position directed by high-level decision-making process. Therefore, the dynamics of the search agent can be simplified as follows:

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k + \boldsymbol{u}_k + w_k \tag{1}$$

where $\boldsymbol{p}_k = [p_{k,x}, p_{k,y}, p_{k,z}]^\mathsf{T} \in \Omega \subseteq \mathbb{R}^3$ denotes the position of the searching agent at current step k with Ω being a convex and compact searching space, $\boldsymbol{u}_k \in \mathcal{U} \subseteq \mathbb{R}^3$ is the control action with $\mathcal{U}$ being the admissible set of actions, and w_k is the control error. It is worth mentioning that the admissible set $\mathcal{U}$ can be continuous, which is distinct from the existing results in [4] and [5], where the search is restricted to certain directions with a fixed step size.

Atmospheric transport and dispersion model (ATDM), governing the spatial-temporal diffusion of the pollutant materials, is utilized to predict the concentration in space and time, given the parameters of a release. We denote true source parameters as $\boldsymbol{\Theta}_s = [\boldsymbol{s}^\mathsf{T}, q_r]^\mathsf{T} \in \mathbb{R}^4$ with $\boldsymbol{s} = [s_x, s_y, s_z]^\mathsf{T} \in \mathbb{R}^3$ being the position of the source and $q_r \in \mathbb{R}^+$ representing a positive release rate. The dispersion model is given by

$$\begin{aligned}\mathcal{M}(\boldsymbol{p}_k, \boldsymbol{\Theta}_s) = \frac{q_r}{4\pi\zeta_{s1}\|\boldsymbol{p}_k - \boldsymbol{s}\|} \exp\left[\frac{-\|\boldsymbol{p}_k - \boldsymbol{s}\|}{\zeta}\right] \\ \times \exp\left[\frac{-(p_{k,x} - s_x)u_s \cos\rho_s}{2\zeta_{s1}}\right] \\ \times \exp\left[\frac{-(p_{k,y} - s_y)u_s \sin\rho_s}{2\zeta_{s1}}\right]\end{aligned} \tag{2}$$

where environmental parameters include the wind speed u_s, wind direction ρ_s, diffusivity ζ_{s1}, the particle lifetime ζ_{s2}, and a composite coefficient $\zeta = \sqrt{\frac{\zeta_{s1}\zeta_{s2}}{1+(u_s^2\zeta_{s2})/(4\zeta_{s1})}}$.

Information collection in autonomous search of an airborne release is mainly from onboard chemical/biological sensors. As the search agent moves to a new position, concentration measurement will be taken. The agent is required to remain at current position for a short period

to obtain a reliable reading, referred as the sampling time. The sensor reading can be modeled as follows:

$$z(\boldsymbol{p}_k) = \mathcal{M}(\boldsymbol{p}_k, \boldsymbol{\Theta}_s) + v_k \tag{3}$$

where $\mathcal{M}$ is the true chemical concentration, and v_k represent additive Gaussian noises imposed on the sensor readings.

B. Objective Function Construction

The dispersion model in (2) is referred as an isotropic plume model [6]. There are many other commonly used dispersion models, such as Gaussian plume [22] and computational fluid dynamics [23]. Nevertheless, the model can be understood as a concentration function that possesses the highest value at the release center and decreases monotonically as the increase of the distance to the center (in terms of expectation). Thus, it can be used to formulate an optimization objective for the autonomous agent, by taking the position $\boldsymbol{p}_k$ as the optimization variable. Following the convention in optimization theory, the objective function is defined as follows:

$$g(\boldsymbol{p}_k, \boldsymbol{\Theta}_k) = (\mathcal{M}_0 - \mathcal{M}(\boldsymbol{p}_k, \boldsymbol{\Theta}_k))^2 \tag{4}$$

where $\mathcal{M}_0$ is a predefined upper bound of the concentration measurement. It is clear that the optimal solution of (4) is $\boldsymbol{p}_k^* = \boldsymbol{s}$, where $\mathcal{M}(\boldsymbol{p}_k^*, \boldsymbol{\Theta}_s)$ is maximized.

To estimate the source location $\boldsymbol{s}$ and release rate q_r based on available measurements, we may define an additional value function taking the source term as the decision variable. Least square methods can serve for this purpose, given by

$$f(\boldsymbol{\Theta}_i, \boldsymbol{p}_i) = [\mathcal{M}(\boldsymbol{p}_i, \boldsymbol{\Theta}_i) - z(\boldsymbol{p}_i)]^2 \tag{5}$$

where $z(\boldsymbol{p}_i)$ denotes the measured concentration at agent position $\boldsymbol{p}_i \forall i = 1, \ldots, k$. The least square function relies on the agent positions and corresponding sensor readings, which are subject to random noises, and thereby stochastic gradient descent method will be introduced later in this article to estimate the source parameters.

III. Concurrent Learning for Dual Control With Exploration and Exploitation

A. Framework of Concurrent Learning With Dual Effects

The dual control framework for autonomous source search and estimation was first introduced by Chen et al. [5] recently. The goal is to drive the agent towards the believed position of a release and in the meanwhile reduce uncertainty associated with the estimation of the target position. Generally speaking, uncertainty is often measured in a stochastic sense for PDF of a variable. In [8], particle filters are utilized to estimate the source parameters and the uncertainty associated with the estimated source target. However, it requires a large number of particles to support the Bayesian inference engine, which incurs heavy computational burden. Quantifying the uncertainty is of great importance, as demonstrated in our previous works [4], [5], [8]. To alleviate this problem, we, thus, introduce a set of N source term estimators, and they are initialized according to the prior knowledge of the source parameters. It is worth emphasizing that the number of estimators N is much smaller than that of particles in Bayesian filters as shown later.

The concentration information collected up to time step k is denoted by $\boldsymbol{\mathcal{Z}}_k := \{z(\boldsymbol{p}_1), z(\boldsymbol{p}_2), \ldots, z(\boldsymbol{p}_k)\}$. Let $\boldsymbol{\Theta}_k^i$ be the STE of the ith estimator at the kth measurement, and $\bar{\boldsymbol{\Theta}}_k := \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Theta}_k^i$ as the nominal estimation, i.e., the mean, of the source parameters. The posterior distribution of source estimation can be represented by $\rho_{k|k} := p(\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}_k)$ at time k. When the search agent moves to a new position directed by the control input $\boldsymbol{u}_k$, the hypothetical posterior distribution of source estimation will be updated as $\hat{\rho}_{k+1|k} := p(\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}_{k+1|k})$, where $\boldsymbol{\mathcal{Z}}_{k+1|k} = \{\boldsymbol{\mathcal{Z}}_k, \hat{z}_{k+1|k}\}$, and consequently, the future belief of concentration can be regarded as a random variable conditional on $\boldsymbol{u}_k$, denoted as $\hat{z}_{k+1|k} \sim p(\hat{z}_{k+1|k}|\boldsymbol{u}_k)$. As a result, the control input $\boldsymbol{u}_k$ will not only affect the future concentration measurement at agent's new position, but also affect the belief of future measurement distribution.

Motivated by the abovementioned discussion, the control input $\boldsymbol{u}_k$ should be designed to navigate the agent to the position where the *predicted* posterior of the measurement $\hat{z}_{k+1|k}$ is close to the predefined threshold $\mathcal{M}_0$, as defined in $g(\boldsymbol{p}_k, \boldsymbol{\Theta}_k)$. Therefore, the conditional cost function can be formulated as follows:

$$\min_{\boldsymbol{u}_k \in \mathcal{U}} J(\boldsymbol{u}_k) = \min_{\boldsymbol{u}_k \in \mathcal{U}} \mathbb{E}_{\boldsymbol{\Theta}} \left[\mathbb{E}_{\hat{z}_{k+1|k}}[(\mathcal{M}_0 - \hat{z}_{k+1|k})^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}]\right] \tag{6a}$$

$$\text{subject to} \qquad \boldsymbol{p}_{k+1|k} = \boldsymbol{p}_k + \boldsymbol{u}_k + w_k. \tag{6b}$$

The physical interpretation is based on all the available information, including priors and available measurements, and we would like the robot moving to a place where the *predicted* maximum concentration is located. This mechanism is behind Chemotaxis, a widely adopted search strategy in nature from bacteria to human being [24]. We show that the control action $\boldsymbol{u}_k$, obtained from the optimization problem in (6), *implicitly* carries dual effects. We define $\bar{z}_{k+1|k}$ as the nominal predicted concentration of the future virtual measurements, i.e., the mean of $p(\hat{z}_{k+1|k}|\boldsymbol{u}_k)$, written as follows:

$$\bar{z}_{k+1|k} := \mathbb{E}[\hat{z}_{k+1|k}|\boldsymbol{\mathcal{Z}}_{k+1|k}]. \tag{7}$$

Based on the definition of $\bar{z}_{k+1|k}$, we can further define $\tilde{z}_{k+1|k} = \hat{z}_{k+1|k} - \bar{z}_{k+1|k}$. Therefore, the objective function can be reformulated as follows:

$$J(\boldsymbol{u}_k) = \mathbb{E}_{\boldsymbol{\Theta}, \hat{z}_{k+1|k}}[(\mathcal{M}_0 - \bar{z}_{k+1|k} - \tilde{z}_{k+1|k})^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}]. \tag{8}$$

Expanding (8) leads to

$$\begin{aligned} J(\mathbf{u}_k) &= \mathbb{E}\left[(\mathcal{M}_0 - \bar{z}_{k+1|k})^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}\right] + \mathbb{E}\left[\tilde{z}_{k+1|k}^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}\right] \\ &\quad - 2\mathbb{E}\left[\tilde{z}_{k+1|k}(\mathcal{M}_0 - \bar{z}_{k+1|k}) | \boldsymbol{\mathcal{Z}}_{k+1|k}\right] \\ &= \mathbb{E}\left[(\mathcal{M}_0 - \bar{z}_{k+1|k})^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}\right] + \mathbb{E}\left[\tilde{z}_{k+1|k}^2 | \boldsymbol{\mathcal{Z}}_{k+1|k}\right]. \end{aligned} \tag{9}$$

In the case of N estimators, we have $\bar{z}_{k+1|k} = \frac{1}{N}\sum_{i=1}^N \hat{z}_{k+1|k}^i$, with $\hat{z}_{k+1|k}^i$ being the predicted measurement at agent's future position $\boldsymbol{p}_{k+1|k}$ based on the ith source estimator $\boldsymbol{\Theta}_k^i$. Then, the optimization problem for CL-DCEE can be formulated as follows:

$$\min_{\boldsymbol{u}_k \in \mathcal{U}} J(\boldsymbol{u}_k) = \min_{\boldsymbol{u}_k \in \mathcal{U}} \left[(\mathcal{M}_0 - \bar{z}_{k+1|k})^2 + \mathcal{P}_{k+1|k}\right] \tag{10a}$$

$$\mathcal{P}_{k+1|k} := \frac{1}{N}\sum_{i=1}^N (\hat{z}_{k+1|k}^i - \bar{z}_{k+1|k})^2 \tag{10b}$$

$$\boldsymbol{p}_{k+1|k} = \boldsymbol{p}_k + \boldsymbol{u}_k + w_k. \tag{10c}$$

Remark 1: According to the definition of $\mathcal{P}_{k+1|k}$ in (10b), it is clear that $\mathcal{P}_{k+1|k}$ is the *predicted* variance of $\hat{z}_{k+1|k}^i \forall i = 1, \ldots, N$, given that each estimator has a uniform weight of $1/N$. The value function in (10a) consists of two parts: the first part exploits current information by navigating the agent toward the believed position of higher concentration, and the second part aims to gather more information by reducing the variance of future virtual measurements. Recently, how to balance between exploration and exploitation has aroused extensive discussions and arguments in many areas, in particular artificial intelligence, optimization, and decision making [5], [20]. In some cases,

artificial weights are introduced on purpose to impose both effects [20]. From the abovementioned formulation process of our framework, a *natural* balance between the two effects is derived from a physically meaningful cost function. Accordingly, our framework eliminates the requirement for choosing tradeoff weights.

To obtain the variance $\mathcal{P}_{k+1|k}$, we resort to the classical principle of predicting variance estimation in extended Kalman filters [25], which can be formulated as follows:

$$\mathcal{P}_{k+1|k} = \mathcal{P}_{k|k}\boldsymbol{F}_{k+1}^{\mathsf{T}}\boldsymbol{F}_{k+1}, \;\; F_{k+1}^{i} = \frac{\partial \mathcal{M}(\boldsymbol{p}_k, \boldsymbol{\Theta}_k^i)}{\partial \boldsymbol{p}} \tag{11}$$

where $\mathcal{P}_{k|k} = \frac{1}{N}\sum_{i=1}^{N}(z_k^i - \bar{z}_k)^2$ denotes current variance of estimated measurement, $z_k^i = \mathcal{M}(\boldsymbol{p}_k, \boldsymbol{\Theta}_k^i)$, $\bar{z}_k = \frac{1}{N}\sum_{i=1}^{N} z_k^i$, and $\boldsymbol{F}_{k+1} = \text{col}[F_{k+1}^1, \ldots, F_{k+1}^N]$. It should be noted that the mean and variance of those variables are calculated by aggregating all elements in the ensemble.

Now, we can present the gradient-based optimization algorithm for the STE and path planning. For notational convenience, we will use

$$y(\boldsymbol{p}_k, \boldsymbol{\Theta}_k) = (\mathcal{M}_0 - \bar{z}_{k+1|k})^2 \tag{12}$$

to denote the first term in the dual objective (10a). Inspired by the memory-based regression parameter estimation methods [26], the N source estimators can be updated according to

$$\boldsymbol{\Theta}_{k+1}^i = \boldsymbol{\Theta}_k^i - \sum_{t=k-q+1}^{k} \eta_t \tilde{\nabla}_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_t) \quad \forall i = 1, 2, \ldots, N \tag{13}$$

where q is a positive integer denoting the number of past measurement used at the kth iteration, and η_t is a constant step size to be designed. The approximated gradients of the least square function (5) can be written as

$$\tilde{\nabla}_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_k) = \nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_k) + \mu_k \tag{14}$$

where μ_k denotes the gradient noises, which can be regarded as a source of perturbation to the true gradient caused by the sensory noises. The path planning is given by

$$\begin{aligned} \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + \boldsymbol{u}_k + w_k \\ \boldsymbol{u}_k &= -\delta_k[\nabla_{\boldsymbol{p}} y(\boldsymbol{p}_k, \boldsymbol{\Theta}_k) + \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}] \end{aligned} \tag{15}$$

where δ_k is constant step size to be designed, and $\boldsymbol{\Theta}_k$ represents the collection of all N estimators. Note that $\nabla_{\boldsymbol{p}} y(\boldsymbol{p}_k, \boldsymbol{\Theta}_k)$ and $\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}$ are pure predictions without measurement noises. Basically, algorithms (13) and (15) use gradient descent method to ensure that the agent moves toward the believed position of a release, and the source estimators converge to the true parameters that minimize the least square function in (5).

For convergence analysis, some basic assumptions on the gradient and control noises are introduced in the following.

Assumption 1: The noise μ_k in (14) satisfies the following properties:

$$\mathbb{E}[\mu_k] = 0, \;\; \mathbb{E}[\|\mu_k\|^2] \leq \varrho^2 \tag{16}$$

where ϱ is a positive constant. The position control error has similar properties

$$\mathbb{E}[w_k] = 0, \;\; \mathbb{E}[\|w_k\|^2] \leq \rho^2 \tag{17}$$

where ρ is a positive constant.

Remark 2: There is an important difference between the traditional gradient-based search methods, such as in Chemotaxis and the proposed CL-DCEE algorithm in this study. In the early works (see e.g., [27] and [28]), mobile robots are equipped with sensors that directly collect the local gradients of concentration and utilize the *measured* gradients to plan their next movement. Clearly, this type of search suffers severely from sensor errors and turbulent fluctuations, since the next movement is purely determined by instantaneous gradient measurements. In our framework, the search agent measures local concentration value and uses all the available information, including priors and available measurements, to learn the source parameters. Based on the acquired knowledge of source, the search agent uses *model-evaluated* gradients to plan its next movement. This learning process lasts over the entire period of search, and therefore, an instant sample, subject to noise and turbulence, will not cause considerable interruption to the path planning.

B. Convergence Analysis

In this section, we will show that the path planning algorithm (15) in conjunction with multiple source estimators (13) will lead the agent to a small neighborhood of the source location $\boldsymbol{s}$.

Theorem 1: Under Assumption 1, all N source estimators in (13) converge to a neighborhood of the true position of the release $\boldsymbol{s}$ from a random initialization set if the learning rate η_t of each estimator is chosen such that

$$\Gamma_k^i = \left\| I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}_k^i(t) \right\|^2 \tag{18}$$

satisfies $0 < \Gamma_k^i < 1$, where $\boldsymbol{\mathcal{T}}_k^i(t) := \int_0^1 \nabla_{\boldsymbol{\Theta}}^2 f(\boldsymbol{\Theta}_s + \tau\tilde{\boldsymbol{\Theta}}_k^i, \boldsymbol{p}_t)d\tau$. Moreover, the expected mean square errors (mse), $\mathbb{E}\|\boldsymbol{\Theta}_k^i - \boldsymbol{\Theta}_s\|^2 \forall i = 1, \ldots, N$, converge at a geometric rate to a bounded neighborhood of zero, given by

$$\lim_{k\to\infty} \mathbb{E}\|\boldsymbol{\Theta}_k^i - \boldsymbol{\Theta}_s\|^2 \leq \frac{\sup_{j\in[1,\infty)}(\sum_{t=j-q+1}^{j} \eta_t^2 \varrho^2)}{1 - \sup_{j\in[1,\infty)}(\Gamma_j^i)}. \tag{19}$$

Proof: It follows from (13) and (14) that

$$\boldsymbol{\Theta}_{k+1}^i = \boldsymbol{\Theta}_k^i - \sum_{t=k-q+1}^{k} \eta_t[\nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_t) + \mu_t]. \tag{20}$$

Now, let $\tilde{\boldsymbol{\Theta}}_k^i = \boldsymbol{\Theta}_k^i - \boldsymbol{\Theta}_s$ denote the error of the agent's estimation relative to source parameters. Then, substituting $\tilde{\boldsymbol{\Theta}}_k$ into (20) results in the error dynamics as

$$\tilde{\boldsymbol{\Theta}}_{k+1}^i = \tilde{\boldsymbol{\Theta}}_k^i - \sum_{t=k-q+1}^{k} \eta_t[\nabla_{\boldsymbol{\Theta}} f\left(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_t\right) + \mu_t]. \tag{21}$$

To relate the gradient term with $\tilde{\boldsymbol{\Theta}}_k^i$, we resort to the mean value theorem [29]. For a twice-differentiable function $H(x) : \mathbb{R}^m \to \mathbb{R}$, the following relation holds, for any $a, b \in \mathbb{R}^m$:

$$\nabla_x H(b) = \nabla_x H(a) + \left[\int_0^1 \nabla_x^2 H[a + \tau(b-a)]d\tau\right](b-a). \tag{22}$$

Therefore, applying the abovementioned theorem leads to

$$\begin{aligned} \nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_k^i, \boldsymbol{p}_t) &= \nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_s, \boldsymbol{p}_t) \\ &\quad + \left[\int_0^1 \nabla_{\boldsymbol{\Theta}}^2 f(\boldsymbol{\Theta}_s + \tau\tilde{\boldsymbol{\Theta}}_k^i, \boldsymbol{p}_t)d\tau\right]\tilde{\boldsymbol{\Theta}}_k^i. \end{aligned} \tag{23}$$

Let us denote $\boldsymbol{\mathcal{T}}^i_k(t) := \int_0^1 \nabla^2_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_s + \tau\tilde{\boldsymbol{\Theta}}^i_k, \boldsymbol{p}_t)d\tau$. Consequently, we have

$$\tilde{\boldsymbol{\Theta}}^i_{k+1} = \Big(I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big)\tilde{\boldsymbol{\Theta}}^i_k - \sum_{t=k-q+1}^{k} \eta_t \mu_t \tag{24}$$

where $\nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}_s, \boldsymbol{p}_t) = \mathbf{0}$ has been used. Taking the square of the Euclidean norm of the error dynamics (24) gives

$$\begin{aligned}\|\tilde{\boldsymbol{\Theta}}^i_{k+1}\|^2 &= \Big\|\Big(I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big)\tilde{\boldsymbol{\Theta}}^i_k - \sum_{t=k-q+1}^{k} \eta_t \mu_t\Big\|^2 \\ &= \Big\|\Big(I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big)\tilde{\boldsymbol{\Theta}}^i_k\Big\|^2 + \sum_{t=k-q+1}^{k} \eta_t^2 \|\mu_t\|^2 \\ &\quad - 2\Big[(I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t))\tilde{\boldsymbol{\Theta}}^i_k\Big]^{\mathsf{T}} \sum_{t=k-q+1}^{k} \eta_t \mu_t.\end{aligned} \tag{25}$$

Let $\mathbb{Q}^i_k := \mathbb{E}\,\|\tilde{\boldsymbol{\Theta}}^i_k\|^2$ denote the expected mse of the variable $\tilde{\boldsymbol{\Theta}}_k$. Then, taking the expectation of (25) results in

$$\begin{aligned}\mathbb{Q}^i_{k+1} &\le \Big\|I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big\|^2 \mathbb{Q}^i_k + \sum_{t=k-q+1}^{k} \eta_t^2 \varrho^2 \\ &\quad - 2\,\mathbb{E}\Big[(I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t))\tilde{\boldsymbol{\Theta}}^i_k\Big]^{\mathsf{T}} \sum_{t=k-q+1}^{k} \eta_t \mu_t \\ &= \Big\|I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big\|^2 \mathbb{Q}^i_k + \sum_{t=k-q+1}^{k} \eta_t^2 \varrho^2\end{aligned} \tag{26}$$

where conditions of the gradient noise in Assumption 1 have been utilized, i.e., μ_t is a white noise independent of $\boldsymbol{\Theta}^i_k$ with bounded variance. To guarantee the convergence of the estimators, it is required that

$$\Gamma^i_k = \Big\|I_4 - \sum_{t=k-q+1}^{k} \eta_t \boldsymbol{\mathcal{T}}^i_k(t)\Big\|^2 \tag{27}$$

within unit circle. Then, we have

$$\lim_{k\to\infty} \mathbb{Q}^i_k \le \frac{\sup_{j\in[1,\infty)}(\sum_{t=j-q+1}^{j} \eta_t^2 \varrho^2)}{1 - \sup_{j\in[1,\infty)}(\Gamma^i_j)} \tag{28}$$

where $\lim_{k\to\infty}(\prod_{j=1}^{k} \Gamma^i_j)\mathbb{Q}^i_0 = 0$ has been applied. In view of (26), it can be concluded that the estimator mse converges to a small neighborhood of zero at a geometric rate, given by $\mathcal{O}(\sup_{j\in[1,\infty)} \Gamma^i_j)$. ■

Remark 3: To ensure that Γ^i_k in (18) is within unit circle, it is sufficient to require $\sum_{t=k-q+1}^{k} \boldsymbol{\mathcal{T}}^i_k(t) > 0$ for a positive integer q under small learning rate $\eta_t > 0$. This is a commonly used condition of persistent excitation in adaptive control and system identification [26], [30]. In essence, at each iteration k, not only current data sample, $(\boldsymbol{p}_k, z(\boldsymbol{p}_k))$ is utilized for updating the environment parameter $\boldsymbol{\Theta}_{k+1}$ but also past $q-1$ step measurements $(\boldsymbol{p}_j, z(\boldsymbol{p}_j))$ for $j \in \{k-q+1, \ldots, k-1\}$ are used. This type of technique is motivated by the memory regressor extension, see [26], to relax the requirement of persistent excitation. The proposed parameter adaption algorithm in (13) encompasses two special cases commonly used in existing literature: stochastic gradient approximation ($q = 1$) and full batch approximation ($q = k$). It is worth noting that increasing the iteration length q can enhance the robustness and accuracy of the adaption algorithm as the excitation effect will be more significant, but may also incur additional computational load [31].

Remark 4: Different from existing filtering techniques, such as extended Kalman filter and Gaussian mixture filter, which usually rely on process models and stochastic properties of process noises to quantify the level of estimation uncertainty, the proposed concurrent learning method uses a hybrid approach that combines both model-based and model-free techniques. The model-based parallel estimators essentially yield a distribution of the estimation at each iteration. A model-free approach is used to calculate the mean and variance of the estimation based on the distribution of the estimations yielded by these parallel estimators. Recently, this hybrid model-based and model-free approach has been proven to be very successful and promising via extensive simulation and experimental studies [32], [33] in machine learning community. It takes the advantage of the model-based approaches in sampling efficiency, but alleviates its inherited model biased error using a model-free ensemble. However, there is no rigorous result for the ensemble approach in machine learning community despite its widely perceived success. Inspired by its success in machine learning, we propose a hybrid parameter estimation approach consisting of N parallel gradient-based estimators and an ensemble process. This approach not only significantly increases the robustness of the parameter estimation particularly in the presence of intermittent sensor measurement, but also provides a reliable way to quantify the level of uncertainty of the current estimation, which is important in realizing the DCEE concept.

Theorem 1 shows the convergence of the estimators, i.e., the estimator will eventually converge using feasible path planning methods, but the optimality is not guaranteed. Convergence of source estimation can be achieved as long as the agent keeps collecting information that fulfils the conditions specified in Theorem 1. In a real search problem, the search environment is complex and there is limited time/sampling budget, and therefore, the search agent has to actively plan its path to quickly approach the source.

Although the path planning and environment acquisition are coupled, it has been shown that under Assumption 1 source estimators can converge to true parameters when measurement samples $k \to \infty$. This important property allows us to employ the well-known separation principle for the convergence analysis of the overall algorithm. Such an analytical principle has been widely used to establish the stability of disturbance observer-based control [34], where design of the controller is separated from design of the observer. In addition, we will further analyze the composite search performance (steady-state performance) in relation to the noise characters.

Theorem 2: Consider a dispersion described by ATDM (2) and the measurement errors and disturbances satisfy Assumption 1. Let η_t satisfy the condition specified in Theorem 1. If the step size δ_k is designed such that

$$0 < 2\|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2 < 1 \tag{29}$$

where $\boldsymbol{\mathcal{L}}_k := \int_0^1 \nabla^2_{\boldsymbol{p}} y(\boldsymbol{s} + \tau\tilde{\boldsymbol{p}}_k, \boldsymbol{\Theta}_k)d\tau$, then the search agent converges to a bounded neighborhood of the source location using the proposed CL-DCEE. Moreover, the steady-state mse bound between agent and true source is given by

$$\lim_{k\to\infty} \mathbb{E}\,\|\boldsymbol{p}_k - \boldsymbol{s}\|^2 \le \frac{\bar{\nu}^2 + \varrho^2}{1 - \sup_{j\in[1,\infty)}(2\|I_3 - \delta_j \boldsymbol{\mathcal{L}}_j\|^2)} \tag{30}$$

where $\bar{\nu} > 0$ denotes the upper bound of the gradient norm of the estimators' variance $\|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|$.

Proof: According to the path update law (15), we have

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k - \delta_k[\nabla_{\boldsymbol{p}} y(\boldsymbol{p}_k, \boldsymbol{\Theta}_k) + \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}] + w_k. \tag{31}$$

Denote $\tilde{\boldsymbol{p}}_k = \boldsymbol{p}_k - \boldsymbol{s}$ as the error of the agent's position relative to the source position. Consequently, the error dynamics of $\tilde{\boldsymbol{p}}_k$ can be

written as

$$\tilde{\boldsymbol{p}}_{k+1} = \tilde{\boldsymbol{p}}_k - \delta_k \nabla_{\boldsymbol{p}} y(\boldsymbol{p}_k, \boldsymbol{\Theta}_k) - \delta_k \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k} + w_k. \tag{32}$$

Following a similar argument as in Theorem 1, we have

$$\tilde{\boldsymbol{p}}_{k+1} = (I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k - \delta_k \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k} + w_k \tag{33}$$

where

$$\boldsymbol{\mathcal{L}}_k := \int_0^1 \nabla^2_{\boldsymbol{p}} y(\boldsymbol{s} + \tau \tilde{\boldsymbol{p}}_k, \boldsymbol{\Theta}_k) d\tau. \tag{34}$$

Then, taking the square of the Euclidean norm for both sides of the error dynamics (33) leads to

$$\begin{aligned}
\|\tilde{\boldsymbol{p}}_{k+1}\|^2 &= \|(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k - \delta_k \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k} + w_k\|^2 \\
&= \|(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k\|^2 + \delta_k^2 \|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2 + \|w_k\|^2 \\
&\quad + 2[(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} w_k - 2\delta_k \nabla_{\boldsymbol{p}}^{\mathsf{T}} \mathcal{P}_{k+1|k} w_k \\
&\quad - 2\delta_k [(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}.
\end{aligned} \tag{35}$$

Let $\mathbb{P}_k := \mathbb{E}\,\|\tilde{\boldsymbol{p}}_k\|^2$ denote the expected mse between the agent's position and the source location. Taking the expectation of (35) and further applying the noise conditions in Assumption 1 and (17), we have

$$\begin{aligned}
\mathbb{P}_{k+1} &\le \|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2 \mathbb{P}_k + \mathbb{E}[\delta_k^2 \|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2] + \rho^2 \\
&\quad + \mathbb{E}[2[(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} w_k] - \mathbb{E}[2\delta_k \nabla_{\boldsymbol{p}}^{\mathsf{T}} \mathcal{P}_{k+1|k} w_k] \\
&\quad + \mathbb{E}[-2\delta_k [(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}] \\
&\le 2\|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2 \mathbb{P}_k + \rho^2 + 2\delta_k^2 \|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2
\end{aligned} \tag{36}$$

where the following three relationships have been applied to derive the second inequality:

$$\begin{aligned}
&\mathbb{E}[2[(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} w_k] = 0 \\
&\mathbb{E}[2\delta_k \nabla_{\boldsymbol{p}}^{\mathsf{T}} \mathcal{P}_{k+1|k} w_k] = 0 \\
&\mathbb{E}[-2\delta_k [(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k]^{\mathsf{T}} \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}] \\
&\quad \le \|\delta_k \nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2 + \mathbb{E}\,\|(I_3 - \delta_k \boldsymbol{\mathcal{L}}_k)\tilde{\boldsymbol{p}}_k\|^2 \\
&\quad = \delta_k^2 \|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2 + \|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2 \mathbb{P}_k.
\end{aligned} \tag{37}$$

In view of the definition of $\mathcal{P}_{k+1|k}$, it is known that the last term in (36), $2\delta_k^2 \|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2$, is a measure of the variance of the estimator error that is upper bounded by

$$\mathbb{E}\,\|\tilde{\boldsymbol{\Theta}}_k^i\|^2 \le \max\left\{\|\tilde{\boldsymbol{\Theta}}_0^i\|^2, \frac{\sup_{j\in[1,\infty)}(\sum_{t=j-q+1}^{j} \eta_t^2 \varrho^2)}{1 - \sup_{j\in[1,\infty)}(\Gamma_j^i)}\right\} \tag{38}$$

where $\|\tilde{\boldsymbol{\Theta}}_0\|^2$ is the initial estimation error of the estimators. Note that the ATDM is a smooth function with respect to the source estimators $\boldsymbol{\Theta}_k^i$, and thus, $\boldsymbol{F}_{k+1}$ is bounded for bounded $\boldsymbol{\Theta}_k^i$, as in (38). Therefore, we can always find an upper bound $\bar{\nu}^2 > 0$ such that $\|\nabla_{\boldsymbol{p}} \mathcal{P}_{k+1|k}\|^2 \le \bar{\nu}^2$. Thus,

$$\mathbb{P}_{k+1} \le (2\|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2)\mathbb{P}_k + \bar{\nu}^2 + \varrho^2. \tag{39}$$

If we choose δ_k such that $(2\|I_3 - \delta_k \boldsymbol{\mathcal{L}}_k\|^2)$ is within unit circle, then the convergence of (39) is guaranteed.

Now, we analyze the steady-state search performance. It follows from (39) that:

$$\lim_{k\to\infty} \mathbb{E}\,\|\boldsymbol{p}_k - \boldsymbol{s}\|^2 \le \frac{\bar{\nu}^2 + \varrho^2}{1 - \sup_{j\in[1,\infty)}(2\|I_3 - \delta_j \boldsymbol{\mathcal{L}}_j\|^2)} \tag{40}$$

where $\lim_{k\to\infty} \prod_{j=1}^{k}(2\|I_3 - \delta_j \boldsymbol{\mathcal{L}}_j\|^2)\mathbb{P}_0 = 0$ has been applied. Similarly, it can be obtained from (39) that the agent converges to a bounded mse in (40) at a geometric rate, given by $\mathcal{O}(\sup_{j\in[1,\infty)}(2\|I_3 - \delta_j \boldsymbol{\mathcal{L}}_j\|^2))$. ■

Remark 5: There is a significant difference between the existing dual control formulation and our framework. Previous studies introduce the exploration effect on the *system* for the purposes of state or parameter estimation [20], [21], whereas in our work, the probing effect is used to explore the *environment* (in this case, learn the source location and release rate). This crucial distinction allows us to learn the unknown environment by reducing estimation uncertainty. Compared with our previous work in [5], there are several distinctions in this work.

1) The formulation in (10) is a concentration-driven optimization problem, whereas [5] uses a position-driven mechanism. From the traditional search strategies point of view, one relates to Chemotaxis while another to Infotaxis [1].
2) DCEE [5] uses particle filters for the STE and also for posterior estimation, which is quite computationally expensive. We moved away from this framework to reduce computational burden to make autonomous algorithms easily implemented on mobile sensor platforms that normally have limited computational resources.
3) The feasible action set Ω can be continuous, whereas only a limited number of feasible actions can be chosen in [5].
4) In this work, we provide a complete theoretical analysis of the modified dual control algorithm using gradient descent. On the other hand, there is no theoretical analysis of the convergence property of the DCEE in [5].

Remark 6: If we remove the second term $\mathcal{P}_{k+1|k}$ in the path-planning objective in (10a), then our algorithm reduces to the pure exploitation strategy, which solely relies on the current estimators of the source parameters. It should be emphasized that the learning process of pure exploitation is *passive* or accidental, since source parameters are updated when the agent makes full use of current belief. The probing effect is included in the value function, by which the agent can *actively* or deliberately learn the environment. In this sense, our CL-DCEE framework is closely related to active learning in MPC [20]. Generally speaking, dual control of exploration and exploitation in an uncertain environment belongs to a much wider class of machine learning problems, in particular, reinforcement learning [35], [36].

IV. Simulation Study

In this section, simulation results will be provided to validate the effectiveness of the proposed algorithms. Since Entrotaxis [8] has demonstrated better performance compared with other existing methods, we will use Entrotaxis as a benchmark for the simulation study. It is worth noting that those informative path planning approaches require a significant amount of computational power due to the implementation of the nonlinear Bayesian filtering and the sampling search-based path planning structure. Detailed settings of the simulation environment can be found in the extended version of this work [37].

Each algorithm has been repeated 200 times with the same configurations. The obtained mse of the CL-DCEE and Entrotaxis algorithms are shown in Fig. 1. MSE evaluates the performance of the source estimators, calculated by $\mathbb{E}(\bar{\boldsymbol{s}}_k - \boldsymbol{s})^2$. It is clear that all algorithms can gradually achieve acceptable estimation of the source position within limited budgets. Uniform distribution of the source location has been applied in the initialization process, as it is assumed that there is no prior information regarding source position. As a result, the initial guess of the source position is around the center of the search space. In general, Entrotaxis requires a large number of measurements to update its particle filter, which leads to a slow acquisition rate of source estimation. On the other hand, our proposed algorithms allow quick update of the source estimators by using instantaneous measurements. CL-DCEE

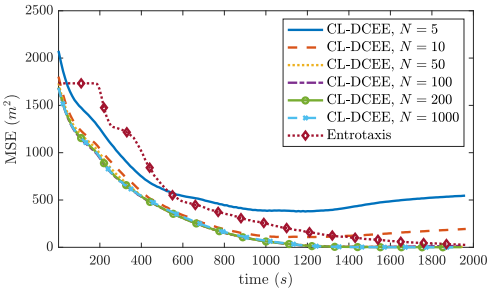

Fig. 1. MSE between the estimated and true source positions.

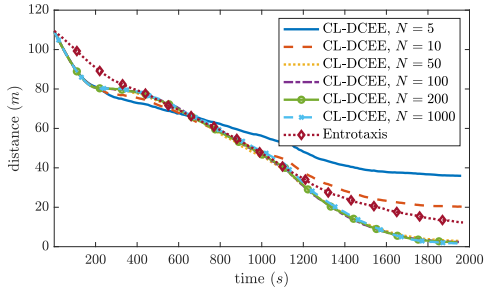

Fig. 2. Distance between agent's position and the true source.

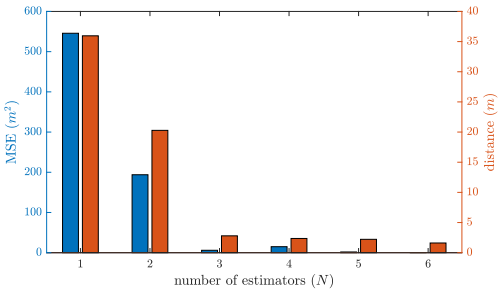

Fig. 3. Performance of CL-DCEE with different number of estimators.

algorithm converges to bounded mse at approximately 1000 s. This property is helpful in conducting emergent identification of the source parameters.

Apart from the estimation accuracy, it is also desired that the search agent can move to the source position, so as to closely monitor the status of the release or take further remedy actions. In Fig. 2, the distance between the agent and the source is displayed. A noticeable phenomenon is that the agent's position using Entrotaxis is quite far from the source position. The proposed CL-DCEE can keep the agent in the neighborhood of the true source, and the steady-state distances are around 35.96, 20.28, 2.80, 2.36, 2.22, and 1.61 m, for $N = 5, 10, 50, 100, 200$, and 1000, respectively.

To show the influence of the number of estimators, we have presented the average performance using different values of N, as shown in Fig. 3. Initially, increasing N can significantly enhance the performance in terms of estimators' mse and the agent's distance to the source (N ranging from 5 to 50). For $N \geq 50$, increasing N is no longer able to provide much performance improvement (N ranging from 50 to 1000). Therefore, the proposed CL-DCEE framework does not require a large number of estimators, and tens of them will be sufficient for autonomous search problem. It also implies that the number of estimators for the ensemble approach should be properly selected to balance estimation performance and computational complexity.

TABLE I
TIME CONSUMED BY RUNNING DIFFERENT ALGORITHMS FOR 200 TRIALS

	CL-DCEE						Entrotaxis
Estimators/Particles	5	10	50	100	200	1,000	10,000
Time (s)	21.8	23.1	24.3	26.3	30.5	56.8	2940.4

An important advantage of the proposed method is the computational efficiency. For clear comparison, we have summarized time consumed by different algorithms, as given in Table I. The simulations are carried out using MATLAB with a processor of 2.8 GHz Quad-Core Intel Core i7. It can be seen that our algorithm is much faster than Entrotaxis. It only consumes less than 1% of the time for Entrotaxis ($N \leq 100$). As a result, CL-DCEE also occupies much less memory storage since the number of estimators is much smaller. This is a very important and advantageous feature because processors used on mobile platforms are usually lower price portable chips that cannot offer intensive computational power or large memory. Detailed discussions and more simulation examples can be found in the extended version of this work [37].

V. CONCLUSION

A computationally efficient solution has been developed for autonomous search of an airborne release with proven properties like convergence. A new learning framework, inspired by DCEE, has been formulated to solve this goal-oriented control problem in an unknown environment with an unknown target. Gradient-based optimization algorithms have been proposed to estimate the source parameters, and to plan next movement by formulating suitable value functions. Theoretical guarantee for convergence and steady-state performance are analyzed under measurement noises and uncertain turbulence. From the simulation and experimental studies, the effectiveness of the proposed solution has been validated. It has been demonstrated that our algorithm achieves superior performance comparing with IPP, and it also consumes much less computation time.

REFERENCES

[1] M. Hutchinson, H. Oh, and W.-H. Chen, "A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors," *Inf. Fusion*, vol. 36, pp. 130–148, 2017.
[2] K. Shankar Rao, "Source estimation methods for atmospheric dispersion," *Atmospheric Environ.*, vol. 41, no. 33, pp. 6964–6973, 2007.
[3] I. Tsitsimpelis, C. J. Taylor, B. Lennox, and M. J. Joyce, "A review of ground-based robotic systems for the characterization of nuclear environments," *Prog. Nucl. Energy*, vol. 111, pp. 109–124, 2019.
[4] M. Hutchinson, C. Liu, and W.-H. Chen, "Information-based search for an atmospheric release using a mobile robot: Algorithm and experiments," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2388–2402, Nov. 2019.
[5] W.-H. Chen, C. Rhodes, and C. Liu, "Dual control for exploitation and exploration (DCEE) in autonomous search," *Automatica*, vol. 133, 2021, Art. no. 109851.
[6] M. Vergassola, E. Villermaux, and B. I. Shraiman, "Infotaxis as a strategy for searching without gradients," *Nature*, vol. 445, pp. 406–409, 2007.
[7] J. Yang, C. Liu, M. Coombes, Y. Yan, and W.-H. Chen, "Optimal path following for small fixed-wing UAVs under wind disturbances," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 3, pp. 996–1008, May 2021.
[8] M. Hutchinson, H. Oh, and W.-H. Chen, "Entrotaxis as a strategy for autonomous search and source reconstruction in turbulent conditions," *Inf. Fusion*, vol. 42, pp. 179–189, 2018.
[9] B. Ristic, A. Skvortsov, and A. Gunatilaka, "A study of cognitive strategies for an autonomous search," *Inf. Fusion*, vol. 28, pp. 1–9, 2016.

[10] Y. Zhao, B. Chen, Z. Zhu, F. Chen, Y. Wang, and D. Ma, "Entrotaxis-jump as a hybrid search algorithm for seeking an unknown emission source in a large-scale area with road network constraint," *Expert Syst. Appl.*, vol. 157, 2020, Art. no. 113484.

[11] R. Khodayi-mehr, W. Aquino, and M. M. Zavlanos, "Model-based active source identification in complex environments," *IEEE Trans. Robot.*, vol. 35, no. 3, pp. 633–652, Jun. 2019.

[12] Y. Sun, F. Gomez, and J. Schmidhuber, "Planning to be surprised: Optimal Bayesian exploration in dynamic environments," in *Proc. Int. Conf. Artif. Gen. Intell.*, 2011, pp. 41–51.

[13] S.-J. Liu and M. Krstic, "Stochastic source seeking for nonholonomic unicycle," *Automatica*, vol. 46, no. 9, pp. 1443–1453, 2010.

[14] 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.

[15] S.-i. Azuma, M. S. Sakar, and G. J. Pappas, "Stochastic source seeking by mobile robots," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2308–2321, Sep. 2012.

[16] N. A. Atanasov, J. Le Ny, and G. J. Pappas, "Distributed algorithms for stochastic source seeking with mobile robot networks," *J. Dyn. Syst., Meas. Control*, vol. 137, no. 3, 2014, Art. no. 031004.

[17] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Autom. Control*, vol. 37, no. 3, pp. 332–341, Mar. 1992.

[18] P. Ogren, E. Fiorelli, and N. E. Leonard, "Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment," *IEEE Trans. Autom. Control*, vol. 49, no. 8, pp. 1292–1302, Aug. 2004.

[19] J.-G. Li, Q.-H. Meng, Y. Wang, and M. Zeng, "Odor source localization using a mobile robot in outdoor airflow environments with a particle filter algorithm," *Auton. Robots*, vol. 30, no. 3, pp. 281–292, 2011.

[20] A. Mesbah, "Stochastic model predictive control with active uncertainty learning: A survey on dual control," *Annu. Rev. Control*, vol. 45, pp. 107–117, 2018.

[21] N. M. Filatov and H. Unbehauen, "Survey of adaptive dual control methods," *IEEE Proc.-Control Theory Appl.*, vol. 147, no. 1, pp. 118–128, Jan. 2000.

[22] Y. Wang, H. Huang, L. Huang, and B. Ristic, "Evaluation of Bayesian source estimation methods with prairie grass observations and Gaussian plume model: A comparison of likelihood functions and distance measures," *Atmospheric Environ.*, vol. 152, pp. 519–530, 2017.

[23] G. C. Efthimiou, I. V. Kovalets, A. Venetsanos, S. Andronopoulos, C. D. Argyropoulos, and K. Kakosimos, "An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment," *Atmospheric Environ.*, vol. 170, pp. 118–129, 2017.

[24] J. B. Stock and M. Baker, "Chemotaxis," in *Encyclopedia of Microbiology*. New York, NY, USA: Elsevier Inc., 2009, pp. 71–78.

[25] G. Welch and G. Bishop, "An introduction to the Kalman filter," Univ. North Carolina Chapel Hill, Chapel Hill, NC, USA, TR 95-041, 1995.

[26] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. A unified framework and some new schemes," *Annu. Rev. Control*, vol. 50, pp. 278–293, 2020.

[27] A. Dhariwal, G. S. Sukhatme, and A. A. Requicha, "Bacterium-inspired robots for environmental monitoring," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2004, pp. 1436–1443.

[28] R. A. Russell, D. Thiel, R. Deveza, and A. Mackay-Sim, "A robotic system to locate hazardous chemical leaks," in *Proc. IEEE Int. Conf. Robot. and Autom*. 1995, pp. 556–561.

[29] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-hill, 1976.

[30] M. Guay and T. Zhang, "Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties," *Automatica*, vol. 39, no. 7, pp. 1283–1293, 2003.

[31] F. Ding and T. Chen, "Performance analysis of multi-innovation gradient type identification methods," *Automatica*, vol. 43, no. 1, pp. 1–14, 2007.

[32] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4754–4765.

[33] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6402–6413.

[34] W.-H. Chen, "Disturbance observer based control for nonlinear systems," *IEEE/ASME Trans. Mechatronics*, vol. 9, no. 4, pp. 706–710, Dec. 2004.

[35] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2737–2752, Jul. 2018.

[36] H. Jeong, B. Schlotfeldt, H. Hassani, M. Morari, D. D. Lee, and G. J. Pappas, "Learning q-network for active information acquisition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 6822–6827.

[37] Z. Li, W.-H. Chen, and J. Yang, "Concurrent active learning in autonomous airborne source search: Dual control for exploration and exploitation," 2021 *arXiv:2108.08062*.