

Received February 24, 2022, accepted March 14, 2022, date of publication March 16, 2022, date of current version March 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3160192

Solving the Datum Search as a Partially Observed Stochastic Game

BRANKO RISTIC^{ID1}, ALEX SKVORTSOV², SANJEEV ARULAMPALAM^{ID2}, (Senior Member, IEEE), HAYDAR DEMIRHAN^{ID3}, AND DU YONG KIM^{ID1}, (Member, IEEE)

¹School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

²Maritime Division, DST Group, Edinburgh, SA 5111, Australia

³School of Science, RMIT University, Melbourne, VIC 3000, Australia

Corresponding author: Branko Ristic (branko.ristic@rmit.edu.au)

This work was supported by the Defence Science and Technology Group under the Research Agreement “Game Theory Based Search Techniques for Evasive Underwater Targets.”

ABSTRACT The Flaming Datum (FD) problem refers to the search for a hostile Evader that is fleeing after momentarily revealing its position. If Evader’s direction is fixed but unknown, while its maximal speed is known, Koopman argued in 1980 that the best trajectory for Searcher is a spiral starting from the revealed position. The objective of our study is to verify the hypothesis of a spiral search path, by formulating the FD problem in the framework of a finite two-player zero-sum partially observed stochastic search game, where the opponent plays repeatedly a fixed but unknown pure strategy. Using a realistic sensor model, current information about the position of Evader is represented by an occupancy map, updated in the Bayesian framework. The utility is computed as the entropy reduction of the occupancy map. The game was implemented in software and solved using the maxmin method. By running repeatedly the described search game, we have found that the search pattern, although random (due to the uncertainties in sensing and mixed strategies), is indeed a spiral on every play of the game, thus confirming the hypothesis.

INDEX TERMS Search games, Bayesian estimation, entropy, machine intelligence.

I. INTRODUCTION

Search is a “hide and seek” game with a long history. The theory was pioneered by Koopman [1]–[3] primarily in the military context, followed by developments by Stone *et al.* [4]. This framework provides the optimal a priori search plan for a given detection model, target motion model and the cost of the search. Searching strategies in biological applications [5]–[8], robotics [9] and security applications [10], [11] are driven by sequentially acquired measurements and result in random search paths. A taxonomy of search problems is presented in [12], distinguishing the case where the target is static and the case where the target can move as a reactive intelligent player. The latter class of problems falls under the theoretical framework of the game theory [13].

In situations characterised by dynamic interaction between the players, where the players do not have complete information about the environment, a class of partially observable stochastic games (POSG) [14] provides a very general

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott^{ID}.

theoretical framework of study. A POSG can be viewed as an application of partially observable Markov decision processes (POMDP) [15] to game theory, in which there are multiple agents (players) with possibly conflicting goals, and their joint actions determine the state transitions and rewards. In general POSG are difficult to solve because a multiple exponential number of histories is needed to be evaluated for each agent.

In this paper we focus on a specific POSG: a two-player zero-sum (TPZS) partially observable stochastic search game, in which Evader (target) repeatedly plays a fixed pure strategy, unknown to Searcher. This framework is appropriate to solve the so-called Flaming Datum problem [16]. Consider a situation where a submarine torpedoes a ship in a convoy and, in doing so, reveals its position. Suppose the attack happened at time $t = 0$. Retaliatory forces arrive at the position where the attack has happened with a time delay and engage in an active search for the submarine until time T . The torpedoed ship marks the position where the submarine was at $t = 0$ and is referred to as the flaming datum (FD). In the meantime, the submarine dived in an attempt to leave the

search area undetected and survive. The payoff of the search is the probability that submarine will be found within the fixed search interval T . Koopman [17] studied this problem assuming the submarine travels with a fixed but unknown course, and at a known radial speed. In such circumstances (without using the game theory) Koopman established that the searcher's best trajectory is generally a spiral of some kind, starting from the FD (see also [18]). The FD problem has been studied as a TPZS game under different constraints and assumption on the behaviour and capacities of the two players [16], [19], [20].

In our formulation of the FD problem, Searcher is equipped with a sensor characterised by a realistic range-dependent probability of detection. Under the assumption that Evader's course is fixed but unknown, we cast the problem in the framework of a finite TPZS POSG. Current knowledge about the location of Evader is represented with a probabilistic occupancy map. This dynamic map is updated, after every stage of the game, by fusion of sensor data with the current map using the Bayes rule. The payoff assigned to each pair of actions by the two players is defined as the entropy reduction of the occupancy map. The described search game has been implemented in software and played repeatedly, because the outcome of each play is naturally random in the adopted framework of POSG. The objective is to verify the hypothesis that the search trajectory is indeed a spiral.

The rest of the paper is organised as follows. Sec. II describes the problem in a formal manner, followed by a game theoretic formulation and its solution in Sec. III. Numerical simulation results are presented in Sec. IV, and conclusions drawn in Sec. V.

II. A PROBLEM STATEMENT

The search area is discretised and modelled by a square grid consisting of $M \times N$ nodes. A node represents a square area (or a cell) of size $R \times R$. The Cartesian coordinates of the (m, n) th node are denoted by $\mathbf{l}_{m,n} = (x_{m,n}, y_{m,n})$. Thus the search area is fully specified with $\mathbf{L} = \{\mathbf{l}_{m,n}; m = 1, \dots, M, n = 1, \dots, N\}$.

Suppose Evader has been observed at time $t = 0$ in the square area represented by node (m^*, n^*) . Being aware that its position has been revealed, Evader will attempt to escape the search area in the quickest possible manner, by moving at its maximum speed (which is known to Searcher) at the course $\psi_0 \in [0, 2\pi)$ rad, which is unknown to Searcher. The searcher arrives at the FD coordinates $\mathbf{l}_{m^*,n^*} = (x^*, y^*)$, corresponding to node (m^*, n^*) , and performs sensing at time $\tau > 0$. Sensing, in the context of underwater surveillance, typically involves a deployment of a dipping sonar - an active sonar submerged under the sea surface [18]. The detection process is naturally imperfect: the probability of detection is decreasing with the range to the target, while there is a non-zero probability of false detections. The sensing model will be described later in this section. Based on the outcome of the sensing activity, Searcher will augment his knowledge about Evader's position and subsequently choose where to go next

to perform sensing. An implicit assumption is that the speed of Searcher U_s is much greater than the speed of Evader U_e (e.g. helicopter vs submarine). The time it takes Searcher to move to its new location is denoted Δ_1 and to carry out the sensing activity is Δ_2 , such that $\Delta_1 + \Delta_2 = \Delta$ is the update interval. The maximum time available for search, $T \gg \Delta$, is limited by the available resources. The search is successful if Evader is found in time less than or equal to T .

We will formulate the described problem as a two-player zero-sum partially observable stochastic search game. Then this game will be implemented in software and played repeatedly in order to study the characteristics of the search trajectories and estimate the probability of successful search as a function of the search time and the characteristics of the sensor.

Next we introduce the sensing model. Let Searcher at discrete-time $k = 1, 2, \dots$ be in a node with coordinates $\mathbf{p}(k) \in \mathbf{L}$. The sensing activity at k results in a set of detections $\mathbf{Z}(k) = \{\mathbf{z}_1(k), \dots, \mathbf{z}_{\mu_k}(k)\}$, where $\mathbf{z} \in \mathbf{Z}(k)$ is a vector consisting of a measured range and azimuth from $\mathbf{p}(k)$ to the perceived target location. The probability of detecting Evader, located at coordinated $\mathbf{l}_{m,n} \in \mathbf{L}$, is denoted $P_d^{(m,n)}(k)$. The probability of detection is typically a function of the distance $r^{(m,n)}(k)$ between the target position $\mathbf{l}_{m,n}$ and the searcher position $\mathbf{p}(k)$, i.e. $r^{(m,n)}(k) = \|\mathbf{p}(k) - \mathbf{l}_{m,n}\|$. The mathematical model of the probability of detection be adopted as

$$P_d^{(m,n)}(k) = \exp\left(-\frac{r^{(m,n)}(k)}{\alpha}\right), \quad (1)$$

where the sensing constant α (dimension of length) depends on the characteristics of the sensor and environment. We assume α is known and specified in terms of the grid period R . Parameter α determines the sensing area (for a given probability of detection). According to (1), a target at distance $r = 3\alpha$ will be detected with the probability of $p_o = 0.05$. Assuming 360° coverage, the sensing area \mathcal{L}_k for p_o is as a circular area of radius 3α . The sensor could also report false detections. Their spatial distribution is assumed to be uniform over \mathcal{L}_k and homogeneous over the search area. The number of false detections in \mathcal{L}_k is modelled with the Poisson distribution whose mean rate is $\lambda > 0$. The measured range and azimuth (i.e. the components of vector $\mathbf{z} \in \mathbf{Z}(k)$) are assumed to be affected by additive zero-mean Gaussian noise with the covariance matrix \mathbf{R} .

III. GAME THEORETIC FORMULATION

Because Evader is moving with a constant velocity from the FD, it effectively plays repeatedly the same (unknown) action. This fact will dramatically simplify the described search problem, because it can be expressed as a TPZS game in the normal (matrix) form of a constant size [21]. Such a game at each time k is defined with a triple $(\mathbb{A}_k, \mathbb{B}_k, \mathbf{U}_k)$, where \mathbb{A}_k and \mathbb{B}_k are the sets of admissible actions of Searcher (Player 1) and Evader (Player 2), respectively. Matrix $\mathbf{U}_k = [u_{ij}]$ is the payoff matrix whose rows are labeled

by the elements of \mathbb{A}_k , and columns with the elements of \mathbb{B}_k . The value of u_{ij} is the payoff (utility) for Player 1 if he plays $\mathbf{a}_i \in \mathbb{A}_k$ and Player 2 plays $\mathbf{b}_j \in \mathbb{B}_k$. Player 1 is referred to as the maximiser, because his goal is to act in such a manner to maximise the utility. Player 2 is referred to as the minimiser, because his goal is the opposite, to minimise the payoff for Player 1. In making a decision which action to take, each player is unaware what the opponent does.

We will consider a finite TPZS game, in which both \mathbb{A}_k and \mathbb{B}_k are finite sets. Set \mathbb{A}_k is formed as a set of nodes in vicinity of the current Searcher position coordinates $\mathbf{p}(k) \in \mathbf{L}$. We restrict motion to the search area grid, specified by \mathbf{L} , and assume the searching agent can move in a radial direction from its current coordinates. Thus, $\mathbb{A}_k = \{\mathbf{a}_i\}_{1 \leq i \leq V}$ where

$$\mathbf{a}_i = \left[\mathbf{p}(k) + \Delta_1 U_s \left(\cos \frac{2\pi i}{V}, \sin \frac{2\pi i}{V} \right) \right]_{\in \mathbf{L}}$$

where notation in the subscript “ $\in \mathbf{L}$ ” means “the nearest grid coordinates in \mathbf{L} .” Set \mathbb{B}_k is formed by the linear motion from the FD at a fixed course, that is $\mathbb{B}_k = \{\mathbf{b}_j\}_{1 \leq j \leq W}$, where

$$\mathbf{b}_j = \mathbf{l}_{m^*, n^*} + (\tau + k\Delta) U_e \left(\cos \frac{2\pi j}{W}, \sin \frac{2\pi j}{W} \right),$$

with $t_k = \tau + (k - 1)\Delta$ being the time interval from $t = 0$ to the current sensing time.

The elements of a POSG in our case include the information state, the set of admissible actions for the two players and the reward function (the utility of Player 1; its negative value is the utility of Player 2 in TPZS games), assigned to each pair $(\mathbf{a}, \mathbf{b}) \in \mathbb{A}_k \times \mathbb{B}_k$. The information state of the considered POSG is the current knowledge about Evader’s location, available to Player 1 (while Player 2 plays always the same strategy) in making a decision how to act. We adopt the Bayesian probabilistic framework and adopt the information state to be a map of the search area, in which each node is assigned a probability that Evader is occupying it. This map will be referred to as the Evader occupancy map (EOM). The utility (the reward function) assigned to a pair $(\mathbf{a}, \mathbf{b}) \in \mathbb{A}_k \times \mathbb{B}_k$ is defined as the reduction of entropy in the EOM. This will be explained in more details in Sec. III-A.

Note that we formulate the search as a sequence of sensing, decision making and the movement to a new location. This effectively decomposes the overall FD search game into a sequence of sub-games played after every sensing activity. This approach, although may lack a guarantee of theoretical optimality, avoids the computational complexity of solving the overall FD search game.

A. EVADER OCCUPANCY MAP

Following [22], the presence or absence of Evader in the (m, n) th node of the search area grid at time k is modelled by a Bernoulli random variable $\delta_{m,n}(k) \in \{0, 1\}$, where by convention, $\delta_{m,n}(k) = 1$ signifies that the target is present in this node ($\delta_{m,n} = 0$ means the opposite). The EOM at

time k represents a two-dimensional array of posterior probabilities, for each grid-node of the search area, that Evader is present in that node at time k . It is computed sequentially, after processing a cumulative sequence of measurement sets $\mathbf{Z}_{1:k} := \mathbf{Z}(1), \dots, \mathbf{Z}(k)$. For the (m, n) th node this posterior probability is defined as get probability of presence acc. to measurements

$$P_{m,n}(k|k) = \Pr\{\delta_{m,n}(k) = 1 | \mathbf{Z}_{1:k}\}.$$

The EOM is then a matrix

$$\mathbf{P}_{k|k} = \begin{bmatrix} P_{1,1}(k|k) & \cdots & P_{1,N}(k|k) \\ \vdots & \ddots & \vdots \\ P_{M,1}(k|k) & \cdots & P_{M,N}(k|k) \end{bmatrix}. \quad (2)$$

At time $t = 0$, based on the full confidence in FD $\mathbf{l}_{m^*, n^*} = (x^*, y^*)$, the EOM $\mathbf{P}_{0|0}$ is specified with $P_{m^*, n^*}(0|0) = 1$ and zero otherwise.

1) UPDATING THE EOM

Let us denote the EOM, predicted to discrete-time $k = 1, 2, \dots$ from the previous time $k - 1$, as $\mathbf{P}_{k|k-1}$. Prediction from time $k - 1$ to k will be explained in Sec. III-A2. This predicted EOM is updated using the Bayes’ rule as follows. Given a detection set $\mathbf{Z}(k)$ obtained by sensing at time $t_k = \tau + (k - 1)\Delta$, if none of the detections in $\mathbf{Z}(k)$ falls into the (m, n) th cell of the EOM, the probability $P_{m,n}(k|k)$ is computed as [22]

$$P_{m,n}(k|k) = \frac{[1 - P_d^{(m,n)}(k)] P_{m,n}(k|k - 1)}{C_1} \quad (3)$$

where

$$C_1 = [1 - P_d^{(m,n)}(k)] P_{m,n}(k|k - 1) + (1 - P_{fa}) [1 - P_{m,n}(k|k - 1)] \quad (4)$$

is a normalising constant. If, on the contrary, a detection falls inside the (m, n) th cell, then the update equation is

$$P_{m,n}(k|k) = \frac{P_d^{(m,n)}(k) P_{m,n}(k|k - 1)}{C_2}, \quad (5)$$

where

$$C_2 = P_d^{(m,n)}(k) P_{m,n}(k|k - 1) + P_{fa} [1 - P_{m,n}(k|k - 1)]. \quad (6)$$

2) PREDICTION OF THE EOM VIA DIFFUSION

During the time interval $[t_{k-1}, t_k]$, for $k = 1, 2, \dots$, evader can move and accordingly, the EOM needs to reflect the uncertainty induced by this motion. Conceptually, this can be achieved through diffusion of the EOM [22], [23]. Diffusion is carried out sequentially, with short time steps of duration $\delta t < \Delta$ starting from t_{k-1} as follows:

$$P_{m,n}(t + \delta t | t) = P_{m,n}(t) \left[1 - \sum_{(i,j) \in \mathcal{N}_{m,n}} \eta_{i,j} \right] + \sum_{(i,j) \in \mathcal{N}_{m,n}} \eta_{i,j} P_{i,j}(t) \quad (7)$$

where $t = t_{k-1}, t_{k-1} + \delta t, t_{k-1} + 2\delta t, \dots$ until t_k . The term $\mathcal{N}_{m,n}$ in (7) represents the set of nodes which are the neighbours of the (m, n) th node on the grid, that is $\mathcal{N}_{m,n} = \{(m-1, n), (m+1, n), (m, n-1), (m, n+1), (m-1, n-1), (m+1, n+1), (m+1, n-1), (m-1, n)\}$. The first summand on the right hand side of (7) is due to the possible movement of the evader from the (m, n) th cell to one of its neighbours, while the second is modeling its possible motion into the (m, n) th cell, coming from one of its neighbours. The “spilling” factors η_{ij} are related to the maximum speed of Evader and the time interval δt . Because Evader can move in any direction, it makes sense to group the neighbouring cells contained in $\mathcal{N}_{m,n}$ into non-diagonal and diagonal neighbours, and to treat them separately. The former are the first four members of $\mathcal{N}_{m,n}$ listed above, and they are assigned an equal spilling factor η' ; the latter are the last four members of $\mathcal{N}_{m,n}$ listed above, and they are assigned an equal spilling factor η'' . Because the diagonal neighbours are further than the non-diagonal neighbours, it make sense to adopt $\eta'' < \eta'$. Furthermore, in order to make sure the first summand on the right hand side of (7) is positive, we obtain a condition $1 - 4\eta' - 4\eta'' \geq 0$, which implies $\eta' + \eta'' \leq \frac{1}{4}$.

B. UTILITY MATRIX

The last element of the POSG that remains to be specified is the reward function. The utility matrix $\mathbf{U}_k = [u_{ij}]$ is an expression of the reward assigned to Player 1 in response to a pair of actions $(\mathbf{a}, \mathbf{b}) \in \mathbb{A}_k \times \mathbb{B}_k$. First, let us define the entropy of the EOM $\mathbf{P}_{k|k}$:

$$H_{k|k} = -\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \left[P_{m,n}(k|k) \log_2 P_{m,n}(k|k) + (1 - P_{m,n}(k|k)) \log_2 (1 - P_{m,n}(k|k)) \right]. \quad (8)$$

Note that the entropy of the initial EOM $\mathbf{P}_{0|0}$, in the case of the full confidence in the FD, is $H_{0|0} = 0$. This reflects the fact that there is no uncertainty about Evader location at $t = 0$. The entropy grows from time $t = 0$ to $t = \tau$ because of diffusion of the EOM, reflecting the uncertainty in Evader’s position caused by its (unknown) motion.

Consider the time instant $t_k = \tau + (k-1)\Delta$, after the update of the EOM resulting in the posterior EOM $\mathbf{P}_{k|k}$. At this point of time Searcher (Player 1) and Evader (Player 2) play the game, that is choose a pair of actions $(\mathbf{a}_i, \mathbf{b}_j) \in \mathbb{A}_k \times \mathbb{B}_k$. Let the predicted EOM to time t_{k+1} be denoted as $\mathbf{P}_{k+1|k}$. Due to diffusion, its entropy will increase, that is $H_{k+1|k} \geq H_{k|k}$. The utility of $(\mathbf{a}_i, \mathbf{b}_j)$ for Player 1 is defined as the reduction of entropy (or the information gain) [24], as follows:

$$\begin{aligned} u_{ij} &\equiv \text{utility}(\mathbf{a}_i, \mathbf{b}_j) \\ &= H_{k+1|k} - \mathbb{E}\{H_{k+1|k+1}(\mathbf{a}_i, \mathbf{b}_j)\} \end{aligned} \quad (9)$$

Here $H_{k+1|k+1}(\mathbf{a}_i, \mathbf{b}_j)$ is the entropy of the EOM after it has been updated using the measurement set $\mathbf{Z}(k+1)$. This

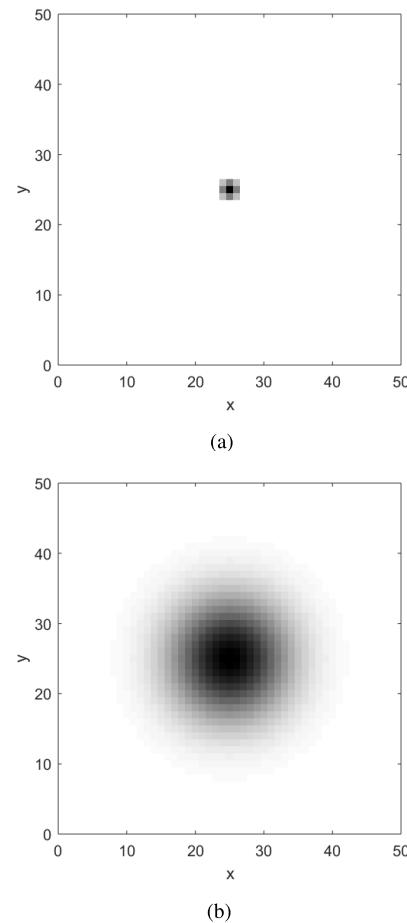


FIGURE 1. Evader occupancy map (EOM) at: (a) $t = 1$ and (b) $t = 70$. (The flaming datum at coordinates $(25, 25)$; white colour indicates low (zero) probability of occupancy).

measurement set, however, is only hypothetical, because the decision has to be made *before* the strategy profile $(\mathbf{a}_i, \mathbf{b}_j)$ has been carried out. Hence, the expectation operator \mathbb{E} with respect to the probability density function of $\mathbf{Z}(k+1)|\mathbf{Z}_{1:k}$ features in (9). In the practical computation of the utility matrix, however, the expectation operator is discarded, while the fictitious measurement set is adopted as the least informative outcome, that is $\mathbf{Z}(k+1) = \emptyset$.

C. A SOFTWARE IMPLEMENTATION OF THE GAME

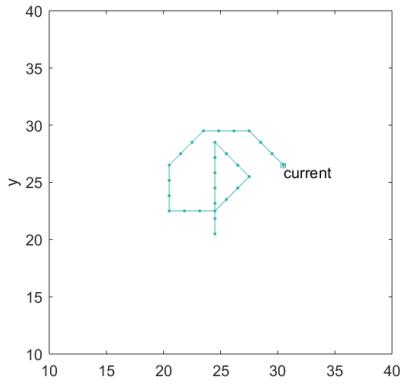
Given the utility matrix $\mathbf{U} = [u_{ij}]$, the corresponding TPZS game is solvable if there exist probability distributions p_a^* over \mathbb{A}_k and p_b^* over \mathbb{B}_k such that

$$\begin{aligned} \sum_{i=1}^V u_{ij} p_a^*(\mathbf{a}_i) &\geq v \quad \text{for } j = 1, \dots, W, \text{ and} \\ \sum_{j=1}^W u_{ij} p_b^*(\mathbf{b}_j) &\leq v \quad \text{for } i = 1, \dots, V. \end{aligned} \quad (10)$$

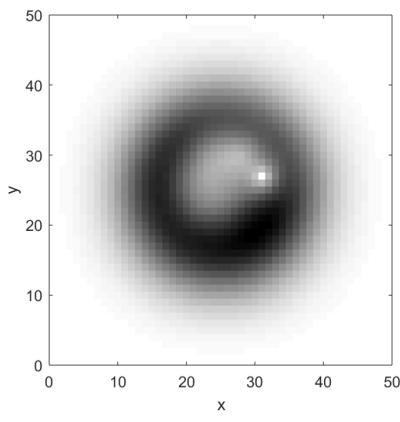
In this case p_a^* and p_b^* are said to be the optimal *mixed* strategies, while v is referred to as the *value of the game*.

TABLE 1. The parameter values used in the search game.

Parameter		value
R	grid period	1
(M, N)	grid size	$(50, 50)$
α	sensor parameter	2.5
λ	mean false detection rate	$1 \cdot 10^{-6}$
(U_s, U_e)	speed of Searcher & Evader	$(1.75, 0.075)$
$(\eta', \eta'', \delta t)$	diffusion parameters	$(1/8, 1/16, 1)$
$(\tau, T, \Delta_1, \Delta_2)$	time intervals	$(70, 300, 2, 1)$
ξ_0	reporting threshold	0.8



(a)



(b)

FIGURE 2. An example of (a) a trajectory of Searcher up to $t = 101$ and (b) the corresponding EOM at $t = 101$. Evader not found yet.

As von Neumann proved in 1928, all finite matrix games have a solution [21]. In solving the games we adopt the maxmin method, the rationale being the awareness that an intelligent competitor will always act in such a manner to create the worst possible situation for the opponent [21]. A solution of the matrix game refers to finding p_a^* , p_b^* , and v , which in practice is carried out using the linear programming (LP) method [25].

The pseudo-code of the software implementation of described TPZS partially observable stochastic search game is given in Alg. 1. According to line 4, Searcher arrives at

the FD at $t_1 = \tau$, while Evader has moved to $e(1)$ following the (unknown to Searcher) course ψ_0 . After the first sensing activity (line 6), the EOM is updated (line 7) and the peak of the EOM is tested for success (lines 8 and 9). If this peak probability is above threshold ξ_0 , the search is off; otherwise it continues in a while-loop, from lines 10 to 25. The game is solved in line 13 using LP, resulting (in general) with a mixed strategy p_a^* . An action a^* is drawn at random from p_a^* in line 14, followed by the movement of Searcher to the next position (line 16). Evader continues following its course, and moves to its new position in line 18. Finally, after the sensing activity, the EOP is updated (lines 20 and 21) and the peak of the EOM is tested for success (lines 22 to 24). If the search has been unsuccessful during the search time T , this is reported as a failure in line 26.

Algorithm 1 Pseudo-Code of the Game

```

1: Set  $\mathbf{P}_{0|0}$  based on the FD
2:  $k = 1; t_1 = \tau;$ 
3: Compute  $\mathbf{P}_{1|0}$  by diffusion of  $\mathbf{P}_{0|0}$  over the interval  $\tau$ ; see Sec.III-A2
4: Searcher position at the FD:  $\mathbf{p}(1) = \mathbf{l}_{m^*, n^*};$ 
5: Evader position:  $\mathbf{e}(1) = \mathbf{l}_{m^*, n^*} + \tau U_e(\cos \psi_0, \sin \psi_0);$ 
6: Searcher in  $\mathbf{p}(1)$  collects a measurement set  $\mathbf{Z}(1)$ 
7: Compute  $\mathbf{P}_{1|1}$  from  $\mathbf{P}_{1|0}$  using  $\mathbf{Z}(1)$ , see Sec. III-A1
8:  $(\hat{m}, \hat{n}) = \arg \max_{m,n} \mathbf{P}_{1|1}; \xi = \max_{m,n} \mathbf{P}_{1|1};$ 
9: if  $\xi > \xi_0$ , search  $\leftarrow$  OFF; Report success and return  $(\hat{m}, \hat{n})$ , else search  $\leftarrow$  ON
10: while search = ON &  $t_k \leq T$  do
11:   Compute  $\mathbb{A}_k, \mathbb{B}_k$ 
12:   Compute utility matrix  $\mathbf{U}_k$ , see Sec.III-B
13:   Solve the game  $(\mathbb{A}_k, \mathbb{B}_k, \mathbf{U}_k)$  to find  $p_a^*$ 
14:   Draw  $a^*$  from  $p_a^*$ 
15:    $k \leftarrow k + 1$ 
16:   Move Searcher to  $\mathbf{p}(k) = a^*$ 
17:   New time:  $t_k = \tau + (k - 1)\Delta$ 
18:   Evader to  $\mathbf{e}(k) = \mathbf{l}_{m^*, n^*} + t_k U_e(\cos \psi_0, \sin \psi_0)$ 
19:   Diffusion of the EOM to obtain  $\mathbf{P}_{k|k-1}$ 
20:   Searcher in  $\mathbf{p}(k)$  collects a measurement set  $\mathbf{Z}(k)$ 
21:   Update the EOM  $\mathbf{P}_{k|k}$  using  $\mathbf{Z}(k)$ 
22:    $(\hat{m}, \hat{n}) = \arg \max_{m,n} \mathbf{P}_{k|k}; \xi = \max_{m,n} \mathbf{P}_{k|k};$ 
23:   if  $\xi > \xi_0$ ,
24:     search  $\leftarrow$  OFF; Report success and return  $(\hat{m}, \hat{n})$ 
25:   end while
26: Report a failure (Evader not found)

```

IV. NUMERICAL RESULTS**A. A SETUP WITH ILLUSTRATIVE RUNS**

The parameters used in simulations are listed in Table 1. Set \mathbb{A}_k includes the possible destinations of Searcher in directions $\{\uparrow, \nearrow, \rightarrow, \searrow, \downarrow, \swarrow, \leftarrow, \nwarrow\}$ from its current position (hence $V = 8$). Evader moves in straight line with a course ψ_0 . We set $W = 8$, that is, $\psi_0 \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$. Thus, the

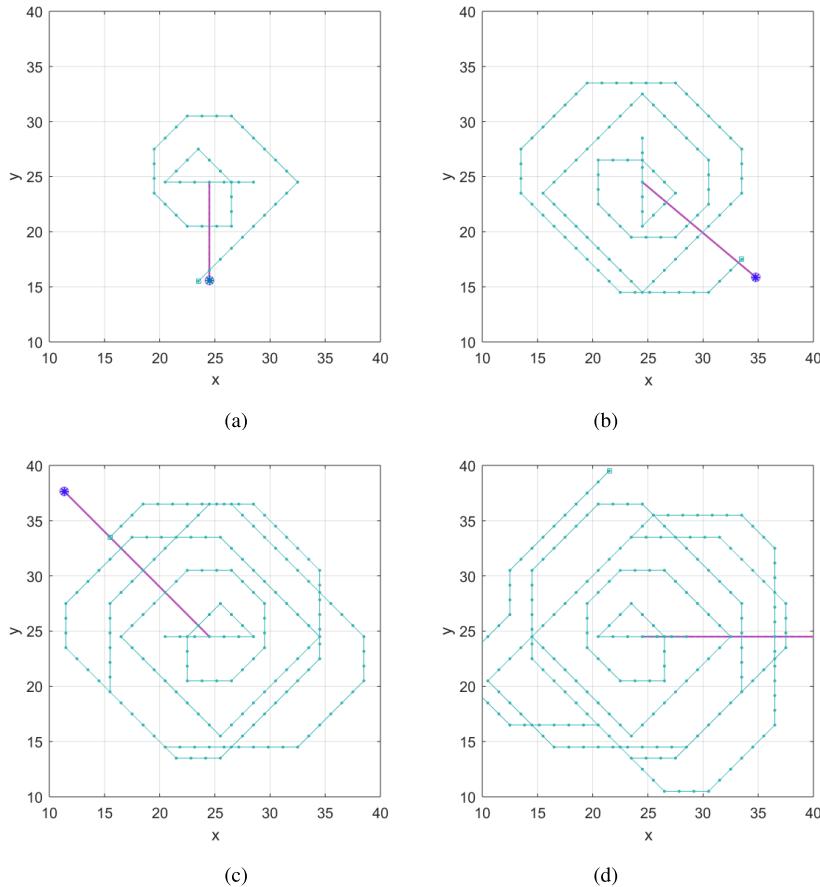


FIGURE 3. Some typical longer search trajectories. Evader always starts at time $t = 0$ from coordinates $(25, 25)$ and moves along a radial path. The search was successful in panels (a), (b) and (c), but Evader escaped in panel (d).

utility matrix at every stage of the game is 8×8 matrix. In simulations we ensure that the final position of Searcher at time t_k is on the grid, i.e. $\mathbf{p}(k) \in \mathbf{L}$. The timing parameters are $\Delta_1 = 2$ units of time (u.t.) $\Delta_2 = 1$ u.t. and $\tau = 70$ u.t.. Because of uncertainty due to the mixing strategy p_a^* (it involves randomness in decision making), and a random detection process, the trajectory of Searcher on every play of the game is random (and different), and so is the success of the search game.

Fig. 1 displays the EOM at time: (a) $t = \Delta$ and (b) $t = \tau$, with white/black areas indicating a small/large probability of occupancy. Based on Fig. 1.(a) it is obvious that the FD (i.e. Evader's position at time $t = 0$) is $(x^*, y^*) = (25, 25)$. The diffusion of the EOM via (7), reflecting the uncertainty caused by Evader's motion, can be observed in Fig. 1.(b): the black area represents the expanding-with-time circle in which Evader is confined.

Fig. 2 show an example of: (a) the trajectory traversed by Searcher up to time $t = 101$ u.t.; (b) the corresponding EOM at $t = 101$ u.t. In this particular run of the search game, Evader has not been found yet, because all measurement sets $\mathbf{Z}(k)$ were empty. “No-detection” information is reflected in the EOM with the gray/white areas inside the expanding black

circle, corresponding to the trajectory traversed by Searcher. Note that white areas with time become gray and then black due to diffusion of the EOM.

Fig. 3 shows four different search trajectories obtained with the same setup. Evader was found in panels (a), (b) and (c), but managed to escape after $T = 300$ s in panel (d). Note from Fig. 3 that search trajectories, obtained by the described POSG, result in a randomised spiral search, starting from the FD, in accordance with the hypothesis in [17]. It has been consistently observed that: (1) the search trajectories are spiral; (2) only in the first few stages, the solution to the game is a mixed strategy; subsequently, the solutions are pure strategies. An avi movie of a typical run is available in Supplementary material.

B. MONTE CARLO RUNS

Next we investigate the probability of a successful search, as a function of the search time T . The results are shown in Fig. 4 for two values of the sensing parameter, namely $\alpha = 2$ and $\alpha = 2.5$. The probability of a successful search was estimated by running 5000 times the described search game (implemented in software) for each value of the search interval T . Naturally, the probability of successful

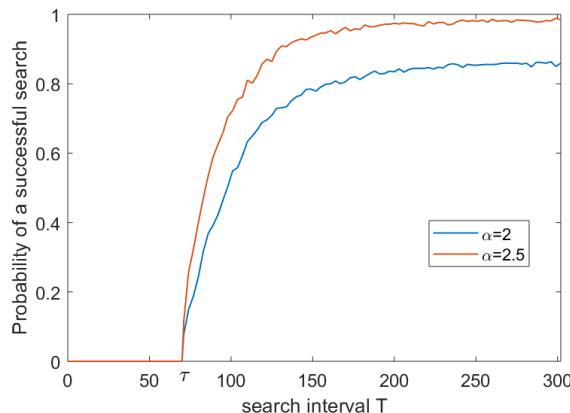


FIGURE 4. Probability of successful search as a function of the search time interval for two values of the sensing parameter α .

search is zero for $T < \tau$, and then grows quickly with $T > \tau$. However, this growth slows down as T increases, to reach a certain saturation level below probability 1. The probability of successful search, as shown in Fig. 4, depends on α : the higher value of α corresponds to the higher probability of detection and therefore, increases the probability of success.

V. CONCLUSION

The paper studied the flaming datum search under the assumption that Evader's course is fixed but unknown. It has been established in the 1980s, using analytical methods, that the searcher's best trajectory in this case is a spiral, starting from the FD. The objective of our study was to verify the hypothesis of a spiral search path, when the FD search is formulated in the framework of a finite two-player zero-sum partially observed stochastic search game. Using a realistic sensor, current knowledge about the location of Evader in this framework is represented with a probabilistic occupancy map. This map is updated after every stage of the game using the Bayes rule, while the utility is expressed as the entropy reduction of the occupancy map. The game was implemented in software and solved using the maxmin method. By running repeatedly, in simulations, the described search game, we have found that the search pattern is random on every play of the game, because of the mixed strategies and uncertainty in sensing. However, the search trajectory is indeed always a spiral, starting from the flaming datum, thus confirming the hypothesis. Mixed search strategies have been observed only in the early stages of the search game.

REFERENCES

- [1] B. O. Koopman, "The theory of search. I. kinematic bases," *Oper. Res.*, vol. 4, no. 3, pp. 324–346, Jun. 1956.
- [2] B. O. Koopman, "The theory of search. II. target detection," *Oper. Res.*, vol. 4, no. 5, pp. 503–531, Oct. 1956.
- [3] B. O. Koopman, "The theory of search: III. the optimum distribution of searching effort," *Oper. Res.*, vol. 5, no. 5, pp. 613–626, 1957.
- [4] L. D. Stone, J. O. Royset, and A. R. Washburn, *Optimal Search for Moving Targets* (International Series in Operations Research & Management Science). New York, NY, USA: Springer, 2016.

- [5] S. E. Halford, "How do site-specific DNA-binding proteins find their targets?" *Nucleic Acids Res.*, vol. 32, no. 10, pp. 3040–3052, Jun. 2004.
- [6] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley, "Lévy flight search patterns of wandering albatrosses," *Nature*, vol. 381, no. 6581, pp. 413–415, 1996.
- [7] P. Fauchald and T. Tveraa, "Using first-passage time in the analysis of area-restricted search and habitat selection," *Ecology*, vol. 84, no. 2, pp. 282–288, Feb. 2003.
- [8] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'Infotaxis' as a strategy for searching without gradients," *Nature*, vol. 445, no. 7126, pp. 406–409, 2007.
- [9] J. Burgués and S. Marco, "Environmental chemical sensing using small drones: A review," *Sci. Total Environ.*, vol. 748, Dec. 2020, Art. no. 141172.
- [10] B. Ristic, A. Skvortsov, and A. Gunatilaka, "A study of cognitive strategies for an autonomous search," *Inf. Fusion*, vol. 28, pp. 1–9, Mar. 2016.
- [11] 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. 1–15, Nov. 2018.
- [12] T. H. Chung, G. A. Hollinger, and V. Isler, "Search and pursuit-evasion in mobile robotics," *Auton. Robots*, vol. 31, no. 4, p. 299, 2011.
- [13] K. Leyton-Brown and Y. Shoham, "Essentials of game theory: A concise multidisciplinary introduction," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 2, no. 1, pp. 1–88, Jan. 2008.
- [14] K. Horák and B. Bošanský, "Solving partially observable stochastic games with public observations," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 2029–2036.
- [15] A. O. Hero, C. Kreucher, and D. Blatt, "Information theoretic approaches to sensor management," in *Foundations and Applications of Sensor Management*, A. Hero, D. Castanon, D. Cochran, and K. Kastella, Eds. New York, NY, USA: Springer, 2008, ch. 3.
- [16] A. Washburn and R. Hohzaki, "The diesel submarine flaming datum problem," *Mil. Operations Res.*, vol. 6, no. 4, pp. 19–30, Sep. 2001.
- [17] B. Koopman, *Search and Screening*. New York, NY, USA: Pergamon, 1980.
- [18] R. B. Yoash, M. P. Atkinson, and M. Kress, "Where to dip? Search pattern for an antisubmarine helicopter using a dipping sensor," *Mil. Oper. Res.*, vol. 23, no. 2, pp. 19–40, 2018.
- [19] R. Hohzaki and A. R. Washburn, "An approximation for a continuous datum search game with energy constraint," *J. Oper. Res. Soc. Jpn.*, vol. 46, no. 3, pp. 306–318, 2003.
- [20] T. Matsuo and R. Hohzaki, "A search game with incomplete information about target's energy," *Scientiae Mathematicae Japonicae*, vol. 80, no. 1, pp. 57–76, 2017.
- [21] A. Washburn, *Two-Person Zero-Sum Games*, 4th ed. New York, NY, USA: Springer, 2014.
- [22] D. W. Krout, W. L. J. Fox, and M. A. El-Sharkawi, "Probability of target presence for multistatic sonar ping sequencing," *IEEE J. Ocean. Eng.*, vol. 34, no. 4, pp. 603–609, Oct. 2009.
- [23] W. Yue, Y. Xi, and X. Guan, "A new searching approach using improved multi-ant colony scheme for multi-UAVs in unknown environments," *IEEE Access*, vol. 7, pp. 161094–161102, 2019.
- [24] T. M. Cover, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.
- [25] P. R. Thie and G. E. Keough, *An Introduction to Linear Programming and Game Theory*, 3rd ed. Hoboken, NJ, USA: Wiley, 2008.



BRANKO RISTIC received the Ph.D. degree from the Queensland University of Technology, in 1995. He held various research/engineering positions in former Yugoslavia and Australia, until, he joined the Defence Science and Technology Organization, in 1996. Since 2015, he has been a Principal Research Fellow with the School of Engineering, RMIT University. He coauthored two books, *Beyond the Kalman Filter: Particle Filters for Tracking Applications* (Artech House, 2004), *Particle Filters for Random Set Models* (Springer, 2013), and over 100 journal articles. His research interests include information fusion, Bayesian estimation, target tracking, Monte Carlo methods, search techniques, sensor control, and reasoning/decision making under uncertainty.



ALEX SKVORTSOV received the Ph.D. degree in theoretical physics from the Moscow University of Applied Physics and Technology. He has significant research and development experience in defence sponsored projects (data fusion, sensor networks, flow noise, and new vibro-elastic materials) on which he worked in academia and industry. He has been working with the Defence Science and Technology Group, since 2005. Since 2014, he has been the Group Leader and the Principal Scientist in acoustic signature control. His research interest includes finding novel solutions for acoustic signature management.



HAYDAR DEMIRHAN is currently a Senior Lecturer with the School of Science, Mathematical Sciences Discipline, RMIT University. He has 60 research publications in high-quality scientific outlets. His research interests include Bayesian learning and its application to AI, computer simulation experiments, fuzzy inference, and cryptographic randomness.



SANJEEV ARULAMPALAM (Senior Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from the University of Melbourne, VIC, Australia, in 1998. In March 2000, he was a recipient of the Anglo-Australian Postdoctoral Research Fellowship, Awarded by the Royal Academy of Engineering, London, U.K. This postdoctoral research was carried out in the U.K., both at the Defence Evaluation and Research Agency and Cambridge University, where he worked on particle filters for nonlinear tracking problems. Recently, he was a Chief Defence Scientist Fellow, investigating multimodal information fusion for undersea warfare to support research with the Undersea Command and Control Branch of Maritime Division, Defence Science & Technology Group (DSTG), Australia. He is currently a Senior Scientist with the Maritime Division of DSTG and an Adjunct Professor with the University of Adelaide. His research interests include estimation theory, target tracking, and sequential Monte Carlo methods. He has coauthored the book *Beyond the Kalman Filter: Particle Filters for Tracking Applications* (Artech House, 2004). He was a recipient of the 2018 IEEE Signal Processing Society's Donald G. Fink Award for one of his papers for its substantial impact over several years.



DU YONG KIM (Member, IEEE) received the B.E. degree in electrical and electronics engineering from Ajou University, South Korea, in 2005, and the M.S. and Ph.D. degrees from the Gwangju Institute of Science and Technology, South Korea, in 2006 and 2011, respectively. He was a Postdoctoral Researcher, he worked on statistical signal processing and image processing with the Gwangju Institute of Science and Technology, from 2011 to 2012, the University of Western Australia, from 2012 to 2014, and Curtin University, from 2014 to 2018. He is currently working as a Vice-Chancellor's Research Fellow with the School of Engineering, RMIT University. His research interests include Bayesian filtering theory and its applications to machine learning, computer vision, sensor networks, and automatic control.

• • •