Learning and Uncertainty-Exploited Directional Antenna Control for Robust Aerial Networking

Mushuang Liu, Yan Wan, Songwei Li, and Frank L. Lewis

***Abstract*— Aerial communication using directional antennas (ACDA) is a promising solution to enable long-distance and broad-band unmanned aerial vehicle (UAV)-to-UAV communication. The automatic alignment of directional antennas allows transmission energy to focus in certain direction and hence significantly extends communication range and rejects interference. In this paper, we develop reinforcement learning (RL)-based on-line directional antennas control solutions for the ACDA system. The novel stochastic optimal control algorithm integrates RL, an effective uncertainty evaluation method called multivariate probabilistic collocation method (MPCM), and unscented Kalman Filter (UKF) for the nonlinear random switching dynamics. Simulation studies are conducted to illustrate and validate the proposed solutions.**

I. Introduction

Aerial communication using directional antennas (ACDA) (see Figure 1) is a promising solution to enable long-distance and broad-band unmanned aerial vehicle (UAV)-to-UAV communication [1]–[3]. Through using directional antennas that focus the transmission energy in certain direction, ACDA significantly extends communication distance and rejects interference, compared to omni-drectional antenna based solutions. The applications of such a system span remote large-area surveillance, remote infrastructure health monitoring, and the provision of on-demand emergency communication services [1], [3], [4].

A critical component of the ACDA system is the automatic alignment of directional antennas to maximize the communication performance. Each UAV in the ANDA system carries a rotational plate mounted with a directional antenna [1], which is controlled to align with the directional antenna carried by the other UAV. Robust automatic alignment of directional antennas is not easy to achieve, considering practical issues such as the limited sensing devices due to the physical constraints of UAV payload and power supplies, uncertain and varying UAV mobility, and unstable GPS and unknown communication environments.

In this paper, we develop a novel antenna control algorithm that adopts reinforcement learning (RL) for online optimal control, multivariate probabilistic collocation method (MPCM) for effective uncertainty evaluation, and unscented Kalman filter (UKF) for nonlinear state estimation. RL methods have been developed to solve optimal control problems with deterministic system dynamics [5]. Paper [6] further developed a stochastic optimal control solution that integrates MPCM and RL methods for systems modulated by uncertainties. In these papers, the uncertainties are relatively simple, as compared to the more complicated random switching random mobility models (RMMs) for the UAV dynamics considered in this paper. With respect to estimation, nonlinear system estimation methods such as Extended Kalman Filter (EKF) and UKF have been used typically for known and deterministic systems corrupted with additive noises, instead of the random switching RMMs. To address these limitations, in this paper we develop a new stochastic optimal control solution for systems that involve nonlinear random switching RMMs and limited measurements, by integrating UKF and RL with MPCM.

This work is supported by NSF grants under numbers 1730675, 1714519, 1839804, and the ONR Grant N00014-18-1-2221.

The authors are with the University of Texas at Arlington, Arlington, Texas, 76019. yan.wan@uta.edu

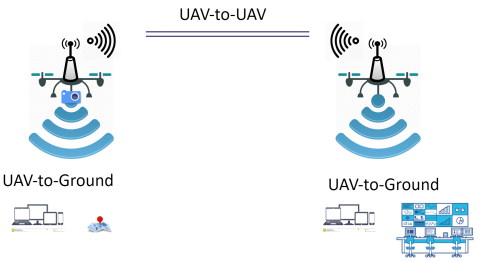


Fig. 1. Illustration of the broadband long-distance communication infrastructure using controllable UAV-carried directional antennas.

II. Modeling and Problem Formulation

In this section, we describe the ACDA system models and measurement models.

A. System Models

We consider two UAVs independently fly in a low-altitude airspace at approximately the same height to fulfill their missions such as search and rescue (see Figure 1). On each UAV, a tunable plate attached with a directional antenna is installed [1]. To establish a long-range air-to-air communication channel, the communication performance of this channel needs to be maximized.

1) UAV Random Mobility Model: We here use the smooth turn (ST) mobility model ([7], [8]) to capture the uncertain intentions of UAVs executing surveillance-like missions. The random maneuvers described by a ST mobility model work as follows. At randomly selected points T_0^i, T_1^i, $T_2^i, \cdots$, where $0 = T_0^i < T_1^i < \cdots$, UAV i selects a point in the airspace along the line perpendicular to its current heading direction, and then circles around it until the UAV chooses

978-1-7281-1220-6/19/$31.00 ©2019 IEEE

another turning center. The perpendicularity guarantees the smoothess of trajectories [7]. The time duration for UAV i to maintain its current maneuver $\tau_i[T_j^i] = T_{j+1}^i - T_j^i$ follows an exponential distribution with the mean of $\tau_i[T_j^i]$. The velocity $v_i[T_j^i]$ follows a uniform distribution, and the inverse of the turning radius $\frac{1}{r_i[T_j^i]}$ follows the zero-mean Gaussian distribution with variance σ_i^2.

Denote the positions of the UAV i along x and y axes at time instant k as $x_i[k]$ and $y_i[k]$ respectively. The dynamics of UAV i (denote as $f_i(.)$) following the ST uncertain maneuvering intentions is described as

$$\begin{aligned} x_i[k+1] &= x_i[k] + v_i[k]\cos(\phi_i[k])\delta, \\ y_i[k+1] &= y_i[k] + v_i[k]\sin(\phi_i[k])\delta, \\ \phi_i[k+1] &= \phi_i[k] + \omega_i[k]\delta, \end{aligned} \tag{1}$$

where δ is the sampling period, $\phi_i[k]$ and $\omega_i[k]$ are the heading angle and angular velocity at the time instant k, and $\omega_i[k] = \frac{v_i[k]}{r_i[k]}$.

Note that the ST RMM is a random switching model composed of two types of random variables [9], [10]. Type 1 random variables includes $v_i[k]$, and $r_i[k]$. They describe the characteristics for each maneuver, and show a random switching behavior.

$$v_i[k] = \begin{cases} v_i[T_j^i], & \text{if } \exists j \in [0,1,2,..), k = T_j^i \\ v_i[k-1], & \text{if } \forall j = 0,1,2,.., k \neq T_j^i \end{cases} \tag{2}$$

$$r_i[k] = \begin{cases} r_i[T_j^i], & \text{if } \exists j \in [0,1,2,..), k = T_j^i \\ r_i[k-1], & \text{if } \forall j = 0,1,2,.., k \neq T_j^i \end{cases} \tag{3}$$

The maneuvers' random switching behavior is governed by the type 2 random variable, $\tau_i[T_j^i]$, which describes how often the switching of type 1 random variables occurs.

The two groups of uncertain maneuvers for the UAVs $(v_1[T_j^1], r_1[T_j^1], \tau_1[T_j^1])$ and $(v_2[T_j^2], r_2[T_j^2], \tau_2[T_j^2])$ are independent, as UAV mobility is application-specific, and is not constrained from the communication mission.

2) Directional Antennas Dynamics: For UAV i, the dynamics of its directional antennas is described as

$$\theta_i[k+1] = \theta_i[k] + (\omega_i^*[k] + \omega_i[k])\delta, \tag{4}$$

where θ_i is the heading angle of antennas i, and ω_i^* is the angular velocity of antennas i due to its heading control. Note that the change of θ_i is caused by both the control of antennas i, ω_i^*, and the movement of UAV i, ω_i.

B. Measurement Models

We consider two measurement signals for the ACDA system, GPS and received signal strength indicator (RSSI).

1) GPS measurement: Denoted the measured GPS signal for UAV i as $Z_{G,i}(k)$, then

$$Z_{G,i}[k] = H_G(k)X_i[k] + \varpi_{G,i}[k], \tag{5}$$

where $H_G = [1,0,0,0;0,1,0,0]$ is the measurement matrix, $X_i[k] = [x_i[k], y_i[k], \phi_i[k], \theta_i[k]]^T$ is the system state of UAV i, and $\varpi_{G,i}$ is the white Gaussian noise with zero mean and covariance $R_{G,i}$. Denote the relation between the GPS signals and system states as $h_{G,i}$, i.e., $Z_{G,i}[k] = h_{G,i}(X_i[k])$.

2) RSSI measurement: Denote the measured RSSI signal as $Z_R[k]$, then according to Friis free space equation [11], one has

$$\begin{aligned} Z_R[k] =& P_{t|dBm}[k] + 20\log_{10}(\lambda) - 20\log_{10}(4\pi) \\ & - 20log_{10}(d[k]) + G_{l|dBi}[k] + \varpi_R[k], \end{aligned} \tag{6}$$

where $P_{t|dBm}[k]$ is the transmitted signal power, λ is the wavelength, and $d[k]$ is the distance between the two UAVs. $G_{l|dBi}[k]$ is the sum of gains at both the transmitting and receiving sides. The Ubiquiti NanoStation loco M5 directional antennas that we use in the ACDA system is modeled as

$$\begin{aligned} G_{l|dBi}[k] =& (G_{t|dBi}^{max} - G_{t|dBi}^{min}) \\ & \times \sin\frac{\pi}{2n}\frac{\sin\left(\frac{n}{2}(k_a d_a(\cos(\gamma_t[k] - \theta_t[k])) - 1) - \frac{\pi}{n}\right)}{\sin\left(\frac{1}{2}(k_a d_a(\cos(\gamma_t[k] - \theta_t[k])) - 1) - \frac{\pi}{n}\right)} \\ & + (G_{r|dBi}^{max} - G_{r|dBi}^{min}) \\ & \times \sin\frac{\pi}{2n}\frac{\sin\left(\frac{n}{2}(k_a d_a(\cos(\gamma_r[k] - \theta_r[k])) - 1) - \frac{\pi}{n}\right)}{\sin\left(\frac{1}{2}(k_a d_a(\cos(\gamma_r[k] - \theta_r[k])) - 1) - \frac{\pi}{n}\right)} \\ & + G_{t|dBi}^{min} + G_{r|dBi}^{min}, \end{aligned} \tag{7}$$

where $G_{t|dBi}^{max}$, $G_{t|dBi}^{min}$, and $G_{r|dBi}^{max}$, $G_{r|dBi}^{min}$ are the maximum and minimum gains of transmitting and receiving antennas. k_a is the wave number. n and d_a are design parameters of the directional antenna. $\theta_t[k]$ and $\theta_r[k]$ are the heading angles of the transmitting and receiving antennas at time k, respectively. $\gamma_t[k]$ and $\gamma_r[k]$ are the heading angles of the transmitting and receiving antennas corresponding to the maximal G_l at time k, respectively.

In ACDA, the two directional antennas are of the same type, and hence $G_{t|dBi}^{max} = G_{r|dBi}^{max}$, and $G_{t|dBi}^{min} = G_{r|dBi}^{min}$. In an imperfect environment, the maximum and minimum antenna gains can be environment-specific, and the desired heading angles $\gamma_t[k]$ and $\gamma_r[k]$ can be captured as $\gamma_r[k] = \arctan\frac{y_t[k]-y_r[k]}{x_t[k]-x_r[k]} + \theta_{r_{env}}$ and $\gamma_t[k] = \arctan\frac{y_r[k]-y_t[k]}{x_r[k]-x_t[k]} + \theta_{t_{env}}$ respectively. $(x_t[k], y_t[k])$ and $(x_r[k], y_r[k])$ are the positions of UAVs that carry the transmitting and receiving antennas respectively, and $\theta_{r_{env}}$ and $\theta_{t_{env}}$ are environment-specific shift angles at the receiver and transmitter sides. $\theta_{r_{env}}$ and $\theta_{t_{env}}$ are zeros in a perfect environment.

C. Problem Formulation

We aim to design antennas' angular velocities to maximize the expected RSSI performance of ACDA over a look-ahead window. The RSSI model contains unknown environment-specific parameters ($G_{t|dB}^{max}$, $G_{t|dBm}^{min}$, and θ_{env}), and the UAV dynamics contain uncertain parameters ($(v_1[k], r_1[k], \tau_1[T_j^1], v_2[k], r_2[k], \tau_2[T_j^2])$).

Here we formulate the problem as a stochastic optimal control problem. Mathematically, considering the random switching system dynamics described in Equations (1) and (4), the optimal control policy $U[k]$ is sought to maximize the expected value function

$$V(X[k]) = E\{\sum_{l=k}^{k+N} \alpha^{l-k} Z_R[l](X[l], U[k])\}, \tag{8}$$

where $X[k]$ is the global state, $X[k] = [X_1^T[k], X_2^T[k]]^T$. $U[k]$ is the control input, $U[k] = [U_1[k], U_2[k]]^T$, $U_i[k] = [\omega_i^*[k]]$. $Z_R[l]$ is the RSSI signal at time l, and $\alpha \in (0, 1]$ is a discount factor. Note that the control is decentralized, in the sense that each antenna finds its own optimal control policy, with the assumption that the other antenna adopts its optimal control policy. In the rest of this article, we develop the control solution for one of the UAVs, and denote this UAV as the local UAV, or UAV 1, and the other UAV as the remote UAV, or UAV 2. The control solution for the other UAV is designed in the same manner.

III. Reinforcement Learning based Stochastic Optimal Control for ACDA

In this section, we develop new on-line solutions to solve the stochastic optimal control problem for the ACDA system described in Section II-C.

A. Stochastic optimal control with unknown RSSI

The stochastic optimal control solution includes two main steps: 1) state estimation, and 2) adaptive optimal controller design. GPS signal is needed in this solution to learn the environment-specific RSSI model.

1) State Estimation: The states of both local and remote UAVs need to be estimated. For the remote UAV that has random switching dynamics, the RMM-related maneuvers ($v_2[k]$, $r_2[k]$, and $\tau_2[T_j^2]$) are unknown to the local UAV, and hence the remote UAV's states ($x_2[k], y_2[k], \phi_2[k]$) can not be estimated directly using existing filtering types of methods. We design a new estimation algorithm for the nonlinear and random switching dynamics.

A critical step in the state estimation of a random switching system is to estimate the expected system state under random switching behaviors. This involves uncertainty evaluation that is typically solved by the Monte Carlo method, which is too slow to be used for on-line solutions. Here, we adopt an efficient uncertainty sampling method, called multivariate probabilistic collocation method (MPCM) [12]. MPCM permits using a very limited number of smartly selected samples to estimate the output mean for a system of input-output mapping subject to uncertain input parameters as described in the following lemma.

Lemma 1. *[12, Theorem 2] Consider a system G modulated by m independent uncertain parameters, a_i, where $i \in \{1, ..m\}$,*

$$G(a_1, ..., a_m) = \sum_{j_1=0}^{2n_1-1} \sum_{j_2=0}^{2n_2-1} ... \sum_{j_m=0}^{2n_m-1} \psi_{j_1,...,j_m} \prod_{i=1}^{m} a_i^{j_i}, \quad (9)$$

where a_i is an uncertain parameter with the degree up to $2n_i - 1$. n_i is a positive integer for any i. $\psi_{j_1,...,j_m} \in \mathbb{R}$ are the coefficients. Each uncertain parameter a_i follows an independent pdf $f_{a_i}(a_i)$. The MPCM approximates $G(a_1, ...a_m)$ with the following low-order mapping

$$G'(a_1, ..., a_m) = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} ... \sum_{j_m=0}^{n_m-1} \Omega_{j_1,...,j_m} \prod_{i=1}^{m} a_i^{j_i}, \quad (10)$$

with $E[G(a_1, ..., a_m)] = E[G'(a_1, ..., a_m)]$, where $\Omega_{j_1,...,j_m} \in \mathbb{R}$ are coefficients. MPCM reduces the number of simulations from $2^m \prod_{i=1}^{m} n_i$ to $\prod_{i=1}^{m} n_i$.

Denote the switching behavior of the remote UAV at time k as $s[k]$. $s[k] = 1$ or 0 represent the current maneuver switches at time k or not. With the two possible switching behaviors, the expected conditional current state can be derived as

$$\begin{aligned} &E(X_2[k]|X_2[k-1]) \\ =&E(X_2[k]|X_2[k-1], s[k-1]=0)P(s[k-1]=0) \\ &+ E(X_2[k]|X_2[k-1], s[k-1]=1)P(s[k-1]=1) \end{aligned} \quad (11)$$

Here we integrate MPCM and UKF for a 5-step state estimation procedure to estimate the remove UAV's state from the measurement $Z_{G,2}[k]$.

Step 1: Initialize. Select initial conditions $\hat{X}_2[0]$ and $P[0]$ to initialize the system.

Step 2: Select MPCM points. $n_1 n_2$ MPCM simulation point pairs are selected according to the pdfs of $\mathcal{V}_{j_1}[T_j^2]$ and $\mathcal{R}_{j_2}[T_j^2]$ and the MPCM procedure [12, Section II], where n_1 and n_2 are the selected degree of $\mathcal{V}_{j_1}[T_j^2]$ and $\mathcal{R}_{j_2}[T_j^2]$ respectively. Denote the selected MPCM point pairs as $(\mathcal{V}_{j_1}[T_j^2], \mathcal{R}_{j_2}[T_j^2])$, where $j_1 \in \{0, ..., n_1 - 1\}$ and $j_2 \in \{0, ..., n_2 - 1\}$).

Step 3: Estimate the state when $\mathbf{s[k-1]=0}$. When $s[k-1] = 0$, the remote UAV does not change its maneuver, and hence the conditional expected current state $E(X_2[k]|\hat{X}_2[k-1], s[k-1] = 0, Z_{G,2}[k])$ can be estimated using UKF. Denote the expected state and covariance found by UKF as $E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k], s[k-1] = 0)$ $E(P[k]|P[k-1], Z_{G,2}[k], s[k-1] = 0)$ respectively. Please refer to [13] for the detailed UKF procedure.

Step 4: Estimate the state when $\mathbf{s[k-1]=1}$. When $s[k-1] = 1$, the remote UAV changes its maneuver according to the random variables $v_2[T_j^2]$ and $r_2[T_j^2]$. With the MPCM points selected in Step 2, the expected state $E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k], s[k-1] = 1)$ and covariance $E(P[k]|P[k-1], Z_{G,2}[k], s[k-1] = 1)$ can be estimated using the following two sub-steps.

(a). Estimate system state at each selected MPCM point. The system state is estimated at each selected MPCM point $(\mathcal{V}_{j_1}[T_j^2], \mathcal{R}_{j_2}[T_j^2])$ by conducting the UKF procedures shown in *Step 3*. Denote the estimated state from UKF at each MPCM point as $\hat{X}_{j_1 j_2}[k|k-1, Z_{G,2}[k], s[k-1] = 1]$ with the covariance $P_{j_1 j_2}[k|k-1, Z_{G,2}[k], s[k-1] = 1]$.

(b). Estimate the expected state using MPCM. Define $G_X(v_2[T_j^2], r_2[T_j^2])$ and $G_P(v_2[T_j^2], r_2[T_j^2])$ as the relationship between the system state and covariance with the uncertain parameters respectively. With the selected MPCM points and the derived system states and covariance at these points, reduced polynomial mappings from uncertain parameters to the system state and covariance (denoted as $G'_X(v_2[T_j^2], r_2[T_j^2])$ and $(G'_P(v_2[T_j^2], r_2[T_j^2])$ respectively) can be obtained according to Lemma 1.

With the polynomial mappings, the expected state and covariance are $E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k], s[k-1]=1) = E[G'_X(v_2[T_j^2], r_2[T_j^2])]$, and $E(P[k]|P[k-1], Z_{G,2}[k], s[k-1]=1) = E[G'_P(v_2[T_j^2], r_2[T_j^2])]$, according to MPCM's mean calculation procedure [12].

Step 5: Estimate the expected system state. The estimated state and covariance are derived as

$$\begin{aligned}
&E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k])\\
=&P_2 E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k], s[k-1]=1])\\
&+(1-P_2)E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k], s[k-1]=0]),\\
&E(P[k]|P[k-1], Z_{G,2}[k])\\
=&P_2 E(P[k]|P[k-1], Z_{G,2}[k], s[k-1]=1])\\
&+(1-P_2)E(P[k]|P[k-1], Z_{G,2}[k], s[k-1]=0]).
\end{aligned}$$

As such, the estimate of $X_2[k]$ is $\hat{X}_2[k] = E(X_2[k]|\hat{X}_2[k-1], Z_{G,2}[k])$, and the expected error covariance is $P[k] = E(P[k]|P[k-1], Z_{G,2}[k])$.

2) Adaptive optimal control: An on-line adaptive optimal controller is designed to maximize the expected value function (8) with the estimated system state.

As the uncertain parameters are independent from the states, the value function can be further re-written as

$$\begin{aligned}
V(X[k]) =& E[\sum_{l=k}^{k+N} \alpha^{l-k} Z_R[l](X[l], U[k])]\\
=& E[Z_R[k](X[k], U[k]) + \sum_{l=k+1}^{k+N} \alpha^{l-k} Z_R[l](X[l], U[k])].
\end{aligned} \tag{11}$$

The above equation can be solved forward-in-time using RL [5]. In particular, we use the policy iteration (PI) method to find the optimal control policy by iteratively conducting the two steps: policy evaluation and policy improvement. The policy evaluation step is designed to solve the value function $V(X[k])$ using Equation (11), given the current control policy. The policy improvement step is designed to find the best control policy to minimize the value function [5]. The two steps are conducted iteratively until convergence.

Policy Evaluation

$$\begin{aligned}
V_{j+1}(X[k]) =& E[Z_R[k](X[k], U[k])\\
&+ \sum_{l=k+1}^{k+N} \alpha^{l-k} Z_{j,R}[l](X[l], U[k])]
\end{aligned} \tag{12}$$

Policy Improvement

$$\begin{aligned}
U_{j+1}(X[k]) =& \arg\max_{U_j[k]} E[Z_R[k](X[k], U[k])\\
&+ \sum_{l=k+1}^{k+N} \alpha^{l-k} Z_{j+1,R}[l](X[l], U[k])]
\end{aligned} \tag{13}$$

where j is the iteration step index, and $Z_{j,R}[l](X[l], U[k])$ is the RSSI model with parameters learned in the jth iteration step.

Note that Equation (12) involves three unknown parameters for the environment-specific RSSI model ($G_{t|dB}^{max}$, $G_{t|dBm}^{min}$, and θ_{env}), which need to be learned. In particular, for each iteration j, three time steps (k, $k+1$ and $k+2$) are needed to come up with three equations to iteratively solve for the three parameters. To calculate the value function $V_{j+1}(X[k])$ at each time step, $E[\sum_{l=k+1}^{k+N} \alpha^{l-k} Z_{j,R}[l](X[l], U[k])]$ is approximated by the output mean of a system mapping using MPCM, $G_V(v_2[T_j^2], r_2[T_j^2]) = E[\sum_{l=k+1}^{k+N} \alpha^{l-k} Z_{j,R}[l](X[l], U[k])]$. MPCM can accurately calculate the expected value function using a limited number of sampling points. The convergence and optimality of the proposed control algorithm are discussed in [10].

B. Using the learned RSSI model in both GPS-available and GPS-denied environments

With the learned RSSI model, the optimal solution can then be obtained in both GPS-available and GPS-denied environments. In a GPS-denied environment, the RSSI is the only measurement. In this case, the optimal control solution can be found following a similar procedure as shown in Section III-A, by replacing $Z_{G,2}[k]$ and $h_{G,2}$ with $Z_R[k]$ and h_R. In the GPS-available environment, GPS and RSSI measurements can be fused to estimate the system states [14] to improve the reliability. The details are omitted here due to the limited space.

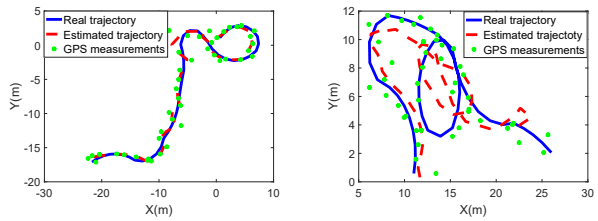
IV. Simulation Studies

Simulation studies are conducted to validate the results and algorithms. Two UAVs move in a 2-D airspace following the ST RMM independently. Two directional antennas of the same type are mounted on the two UAVs respectively. The total simulation time is $T = 45s$, with the sampling period $\delta = 1s$.

We first simulate the case when the GPS is available but the RSSI model is unknown. Gaussian noise is added to the GPS measurements. Estimation for UAV 1 is based on UKF with known maneuver ($v_1[k]$ and $r_1[k]$), while the estimation for UAV 2 is based on the integration of UKF and MPCM as described in Section III-A. Figures 2(a) and 2(b) show the trajectories of UAV 1 and UAV 2 respectively. It can be seen from the figures that 1) the estimated trajectories for UAVs 1 and 2 are both close to their real trajectories, indicating that the proposed state estimation algorithm performs well in both known and unknown maneuver cases; 2) compared with UAV 2, the estimated trajectory of UAV 1 is closer to its real trajectory as expected, indicating that the state estimation algorithm with known maneuver guarantees a better performance.

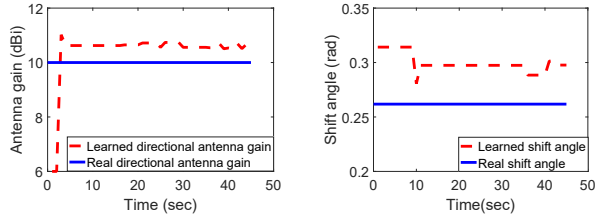
With the estimated states, we simulate the RL-based stochastic optimal control algorithm. Figures 3(a) and 3(b) show the learned environment-specific antennas' maximum gain ($G_{t|dBi}^{max}$) and the shift angle caused by the environment ($\theta_{t_{env}}$) respectively. As shown in the figures, the learned parameters are very close to their true values, which indicates the effectiveness of the learning algorithm.

With the learned RSSI model, we simulate the proposed stochastic optimal control algorithm in GPS-denied and GPS-



(a) (b)

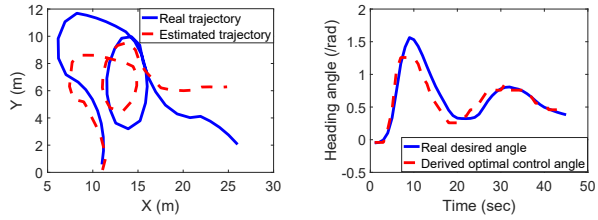
Fig. 2. (a) Trajectories of UAV 1, and (b) Trajectories of UAV 2. The blue solid curves are real trajectories, red dotted curves are estimated trajectories, and green dots are GPS measurements.



(a) (b)

Fig. 3. Learned environment-specific (a) maximum directional antenna gain ($G^{max}_{t|dBm}$), and (b) shift angle (θ_{env}) in the RSSI model.

available environments respectively. Figures 4 and 5 show the estimation and control performances in the GPS-denied and GPS-available cases respectively. It can be seen from the figures that: 1) the estimated trajectories and derived heading angles are very close to their true trajectories and real optimal heading angles in both cases, indicating that the proposed solutions work well in both GPS-available and GPS-denied environments; 2) the angle errors in the GPS-available case are much smaller than those in the GPS-denied case, indicating that the fusion of the GPS and RSSI promises a better performance.

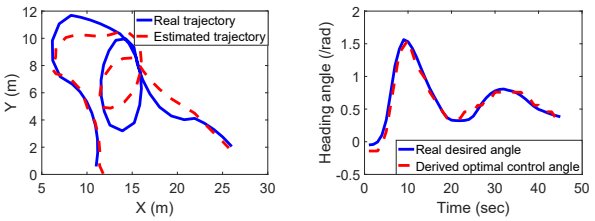


(a) (b)

Fig. 4. Performances of the developed (a) estimation algorithm, and (b) controller algorithm, in a GPS-denied environment.

V. Conclusions and Future Works

In this paper, we developed an RL-based on-line directional antenna control solution for the ACDA system. In particular, to capture the uncertain intentions of UAVs, we adopted a UAV ST RMM. With this nonlinear random switching mobility model, a new state estimation algorithm that integrates MPCM and UKF was developed. To account for an unstable GPS environment and provide online optimal



(a) (b)

Fig. 5. Performances of the developed (a) estimation algorithm, and (b) controller algorithm, in a GPS-available environment.

control solution, we developed a novel stochastic optimal controller by integrating RL and MPCM. The algorithm also features the learning of communication RSSI models and the use of RSSI to inform controller design.

References

[1] J. Chen, J. Xie, Y. Gu, S. Li, S. Fu, Y. Wan, and K. Lu, "Long-range and broadband aerial communication using directional antennas (acda): design and implementation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 793–10 805, 2017.

[2] S. Li, C. He, M. Liu, Y. Wan, Y. Gu, J. Xie, S. Fu, and K. Lu, "The design and implementation of aerial communication using directional antennas: Learning control in unknown communication environments," *IET Control Theory & Applications*, 2019.

[3] Y. Wan and S. Fu, "Communicating in remote areas or disaster situations using unmanned aerial vehicles," *Homeland Security Today Magazine*, pp. 32–35, 2015.

[4] S. Li, Y. Wan, S. Fu, M. Liu, and H. F. Wu, "Design and implementation of a remote uav-based mobile health monitoring system," in *Proceedings of SPIE on Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure*, Portland, OR, 2017.

[5] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32–50, 2009.

[6] J. Xie, Y. Wan, K. Mills, J. J. Filliben, and F. L. Lewis, "A scalable sampling method to high-dimensional uncertainties for optimal and reinforcement learning-based controls," *IEEE Control Systems Letters*, vol. 1, no. 1, pp. 98–103, 2017.

[7] Y. Wan, K. Namuduri, Y. Zhou, and S. Fu, "A smooth-turn mobility model for airborne networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3359–3370, 2013.

[8] J. Xie, Y. Wan, K. Namuduri, S. Fu, G. L. Peterson, and J. F. Raquet, "Estimation and validation of the 3d smooth-turn mobility model for airborne networks," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, San Diego, CA, 2013.

[9] T. Li, Y. Wan, M. Liu, and F. L. Lewis, "Estimation of random mobility models using the expectation-maximization method," in *Proceedings of IEEE 14th International Conference on Control and Automation (ICCA)*, Anchorage, AK, 2018.

[10] M. Liu, Y. Wan, and F. L. Lewis, "Adaptive optimal decision in multi-agent random switching systems," *IEEE Control Systems Letters*, vol. 4, pp. 265–270, 2019.

[11] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. PTR New Jersey, 1996.

[12] Y. Zhou, Y. Wan, S. Roy, C. Taylor, C. Wanke, D. Ramamurthy, and J. Xie, "Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 10, pp. 1347–1363, 2014.

[13] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[14] J. Yan, Y. Wan, S. Fu, X. J., S. Li, and K. Lu, "Rssi-based decentralized control for robust long-distance aerial networks using directional antennas," *IET Control Theory and Applications*, vol. 11, no. 11, pp. 1838–1847, 2016.