Kalman Smoothing for Distributed Optimal Feedback Control of Unicycle Formations

Ross P. Anderson and Dejan Milutinović

Abstract. In many multi-agent control problems, the ability to compute an optimal feedback control is severely limited by the dimension of the state space. In this work, deterministic, nonholonomic agents are tasked with creating and maintaining a formation based on observations of their neighbors, and each agent in the formation independently computes its feedback control from a Hamilton-Jacobi-Bellman (HJB) equation. Since an agent does not have knowledge of its neighbors' future motion, we assume that the unknown control to be applied by neighbors can be modeled as Brownian motion. The resulting probability distribution of its neighbors' future trajectory allows the HJB equation to be written as a path integral over the distribution of optimal trajectories. We describe how the path integral approach to stochastic optimal control allows the distributed control problems to be written as independent Kalman smoothing problems over the probability distribution of the connected agents' future trajectories. Simulations show five unicycles achieving the formation of a regular pentagon.

1 Introduction

The focus of this work is on formation control, in which each agent, a robotic nonholonomic vehicle, in a team is tasked with attaining and maintaining pre-specified distances from the agents in its neighborhood. Problems of this type are beginning to demonstate their significance and potential impact in a variety of applications in both the public and private sector [3, 22, 29, 35]. Nonholonomic vehicle formations, in particular, have attracted much attention [1, 7, 8, 9, 31, 32, 33], but these studies have typically relied on stability analyses, or on ad-hoc artificial potential functions or navigation functions.

Ross P. Anderson · Dejan Milutinović
University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95060
e-mail: {anderson,dejan}@soe.ucsc.edu

D. Milutinović & J. Rosen (Eds.): Redundancy in Robot Manipulators, LNEE 57, pp. 145–166.
DOI: 10.1007/978-3-642-33971-4_9

Alternatively, the formation control problem may be defined as an optimal feedback control problem. To compute an optimal feedback control, one must solve the Hamilton-Jacobi-Bellman (HJB) equation, which is a nonlinear partial differential equation (PDE). However, the computational complexity of solving the PDE grows exponentially with the size of the team, and this severely limits the effectiveness of conventional techniques of stochastic optimal control. A number of promising approaches to address this so-called "curse of dimensionality" have been proposed, including reinforcement learning [36], neurodynamic programming [4], and approximate dynamic programming [30], just to name a few.

In this work, we approach the problem of formation control based on the path integral formulation of stochastic optimal control [17]. We explicitely take into account the fact that although agents may be capable of observing or receiving the current state of their neighbors, the future trajectories of these neighbors will seldom be known exactly, since they individually compute their control based on their own available information and observations. From this point of view, the distributed formation control problem is inherently stochastic, and not only due to unpredictable neighbors. In addition, the control is a function of an agent's noisy observations, and it must also deal with agent model uncertainties and environmental uncertainties (e.g., wind). Along these lines, this work considers the problem of controlling one agent based on observations of its neighbors *and the probability of their future motion*. This probability distribution arises from an assumption that the unknown control of an agent can be modeled as Brownian motion [15], so that based on the system kinematics, we can infer the probability of finding the relative state $\mathbf{x}$ to all neighbors in an interval $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ at a particular future time [40].

Perhaps more importantly, this probability distribution over future system trajectories can be used to statistically infer the probability distribution of the *control*, and, hence, the optimal control. When an agent's neighbors are treated as nondeterministic, one can consider that agent's optimal control to be the action that minimizes the expected value of the accumulated cost with respect to the distribution of neighbors' future trajectories (see [20] and references therein for a more precise interpretation in terms of minimization of Kullback-Leibler divergence). Along these lines, the path integral (PI) formulation of stochastic optimal control [16, 19, 18] transforms the problem of solving the HJB equation into an estimation problem on the distribution of optimal trajectories in continuous state space [39].

The path integral approach is made possible by the relation between the solutions to optimal control PDEs and the probability distribution of stochastic differential equations [27, 44] (see [24, 25, 26, 28] for an analogous approach in the open-loop control case), and it has shown great potential for systems with large state space. For example, Theodorou et al. [37, 38] have examined the the link between reinforcement learning and the path integral method for motor control and robotics, while van den Broek et al. [5, 6] and Wiegerinck et al. [41, 42] apply the path integral framework to multi-agent systems. In the latter, the agents exhibit explicitly-stochastic kinematics and cooperatively compute their control from a marginalization of the joint probability distribution of the group's system trajectory.

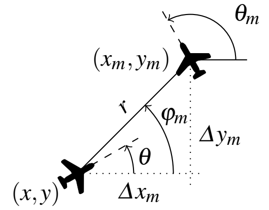

Fig. 1 Diagram of the AiF moving at direction θ and at a distance r to the neighbor m. The turning rate and acceleration of the neighbor are unknown.

In this chapter, we develop a method to apply the path integral approach to fully-distributed multi-agent systems by constructing a fast-switching process that randomly select a single neighboring agent from which the optimal control may be inferred. We also establish a connection between the optimal feedback control problem for multi-agent systems and nonlinear Kalman smoothing algorithms, which allows each agent to independently compute its control in real-time.

This chapter is organized as follows. Section 2 introduces the formation control problem as viewed by a single agent in the group and describes a way to parameterize the kinematic model so that the resulting HJB equation solution can be represented as a path integral. Section 3 reviews the path integral formulation of stochastic optimal control. Section 4 presents a duality between stochastic optimal feedback control and Kalman smoothing algorithms. Section 5 illustrates our methods with a simulated five-agent formation, and conclusions are in Section 6.

2 Control Problem Formulation

In this section we formulate the optimal feedback control problem for unicycle formations and describe a way to manipulate the model kinematics into a form that allows the HJB equation solution to be described by a path integral.

2.1 Preliminary Kinematic Model

In the problem formulation, each agent is modeled as a unicycle, which moves in the direction if its heading angle θ at a speed v:

$$dx(t) = v\cos\theta dt \tag{1}$$

$$dy(t) = v\sin\theta dt \tag{2}$$

$$d\theta(t) = \omega dt \tag{3}$$

$$dv(t) = udt, \tag{4}$$

where u is the feedback acceleration control and ω is the feedback turning rate control.

Each agent independently computes its respective control based on observations of its M neighbors, labeled $m = 1,\ldots,M$. To this end, we focus on the system state as viewed by one agent, which we call the agent-in-focus, or AiF for short. Define $\Delta x_m = x_m - x$ and $\Delta y_m = y_m - y$ as the Cartesian components of the distance from the AiF to the neighbor m (Fig. 1). These states evolve as:

$$d\Delta x_m(t) = -v\cos(\theta)\,dt + v_m\cos(\theta_m)\,dt \tag{5}$$
$$d\Delta y_m(t) = -v\sin(\theta)\,dt + v_m\sin(\theta_m)\,dt, \tag{6}$$

where v_m is the speed of neighbor m. Although the kinematics of the neighboring vehicles are identical, their turning rate control and acceleration control are unknown. Based on the motivation of the previous section, we assume that the turning rate and acceleration controls of neighbor m can be modeled as Wiener processes with mutually independent increments $dw_{\theta,m}$ and $dw_{v,m}$, and intensities σ_θ and σ_v, respectively:

$$d\theta_m = \sigma_\theta dw_{\theta,m} \tag{7}$$
$$dv_m = \sigma_v dw_{v,m}. \tag{8}$$

Finally, introducing the distance from the AiF to the neighbor m as $r_m = \sqrt{\Delta x_m^2 + \Delta y_m^2}$ and the angle to the neighbor m as $\varphi_m = \tan^{-1}(\Delta y_m/\Delta x_m)$, we arrive at a preliminary model for the AiF and a single neighbor m:

$$dr_m(t) = -v\cos(\varphi_m - \theta)\,dt + v_m\cos(\varphi_m - \theta_m)\,dt \tag{9}$$
$$d\varphi_m(t) = \frac{v}{r_m}\sin(\varphi_m - \theta)\,dt - \frac{v_m}{r_m}\sin(\varphi_m - \theta_m)\,dt \tag{10}$$
$$d\theta(t) = \omega dt \tag{11}$$
$$d\theta_m(t) = \sigma_\theta dw_{\theta,m} \tag{12}$$
$$dv(t) = udt \tag{13}$$
$$dv_m(t) = \sigma_v dw_{v,m}. \tag{14}$$

Note that this system, when augmented to account for M neighbors, would have a two-dimensional control $\mathbf{u} = [\omega, u]^T$, but that the stochastic process $\mathbf{w}(t) = \left[w_{\theta,1},\ldots,w_{v,M}\right]^T$ would be of dimension M.

2.2 Switching Kinematic Model

Since the AiF will be drawing from the random processes describing its neighbors kinematics as a source from which to compute its control, we wish to devise a way to connect the random motion in (12) and (14) to the controls in (11) and (13). In particular, for reasons that will become more clear in Section 3, any controlled state should be affected by just one Wiener process, and vice versa. In the model developed in the previous section, this is not the case since there is a two dimensional control $\mathbf{u}$ and M dimensional stochastic process. In order to manipulate the model into such a form, we introduce a second, faster time scale t/ε, for a small $\varepsilon > 0$, and we assume that in each infinitesimal time increment in this faster time scale, the AiF is using the relative state to just one neighbor m to compute its control, and that the choice of neighbor m switches randomly among the M neighbors at a fast rate. It will turn out [43] that under a sufficiently fast secondary time scale, the randomly-switching model to be developed in this section is equivalent to (9)-(14).

Let us define the difference in heading angle $\gamma = \theta - \theta_m$ and difference in speed $\kappa = v - v_m$. Then we can obtain (Appendix 1) the following model, one for each of M neighbors.

$$\begin{aligned} dr_m(t) = &-\left(\frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\kappa_j\right)\right)+\overline{v_m}(0)\right)\cos\left(\varphi_m-\frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\gamma_j\right)\right))-\overline{\theta_m}(0)\right)dt \\ &+\left(-\left(\kappa_m-\mathbb{E}\left(\kappa_m\right)\right)+v_m(0)\right)\cos\left(\varphi_m+\left(\gamma_m-\mathbb{E}\left(\gamma_m\right)\right)-\theta_m(0)\right)dt \end{aligned} \tag{15}$$

$$\begin{aligned} d\varphi_m(t) = &\left(\frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\kappa_j\right)\right)+\overline{v_m}(0)\right)\sin\left(\varphi_m-\frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\gamma_j\right)\right)-\overline{\theta_m}(0)\right)dt \\ &-\frac{1}{r_m}\left(-\left(\kappa_m-\mathbb{E}\left(\kappa_m\right)\right)+v_m(0)\right)\sin\left(\varphi_m+\left(\gamma_m-\mathbb{E}\left(\gamma_m\right)\right)-\theta_m(0)\right)dt \end{aligned} \tag{16}$$

$$d\gamma_m(t) = \left(M\omega dt - \sqrt{M}\sigma_\theta dw_{\theta,m}\right)\delta_{\xi(t/\varepsilon),m} \tag{17}$$

$$d\kappa_m(t) = \left(Mudt - \sqrt{M}\sigma_v dw_m\right)\delta_{\xi(t/\varepsilon),m}. \tag{18}$$

When taking into account all M neighbors of the AiF, the system state is defined through a concatenation of the model (15)-(18), one for each neighbor $m = 1,\ldots,M$.

The fast switching behavior [11] is captured by an ergodic Markov chain $\xi(t/\varepsilon)$ with a fast time scale $\varepsilon > 0$, taking on values in $\{1,2,\ldots,M\}$. We assume that this chain is independent of the Wiener process $\mathbf{w}(t)$ affecting the neighbors' heading angles and speeds in (7)-(8). The Kronecker deltas δ in (17)-(18), therefore, select the pair of "actively evolving" states among the M states in $[\gamma_1,\ldots,\gamma_M,\kappa_1,\ldots,\kappa_M]^T$. In other words, if $\xi(t/\varepsilon) = M-1$, for example, only the states γ_{M-1} and κ_{M-1} evolve as in (17) and (18), while all other relative heading angle and relative speed states have zero increment ($d\gamma_m = d\kappa_m = 0$). In this case, the AiF is using the random motion of neighbor $M-1$ to compute its control.

The evolution equation for M neighbors may now be written in a more general stochastic differential equation for the state vector $\mathbf{x}(t)$:

$$d\mathbf{x}(t) = f(\mathbf{x})dt + B_i\mathbf{u}dt + \Gamma_i d\mathbf{w}, \tag{19}$$

where $f(\mathbf{x})$ describes the deterministic motion in states (15)-(16), and the matrices B_i and Γ_i in the state $\xi(t/\varepsilon) = i$ and state vector $\mathbf{x}(t)$ are constructed as:

$$B_i = M \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \delta_{i1} & 0 \\ \vdots & \vdots \\ \delta_{i,M} & 0 \\ 0 & \delta_{i1} \\ \vdots & \vdots \\ 0 & \delta_{i,M} \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} r_1,\dots,r_M,\varphi_1,\dots,\varphi_M \\ \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} \gamma_1,\dots,\gamma_M \\ \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} \kappa_1,\dots,\kappa_M \end{array} \qquad \mathbf{x}(t) = \left[\begin{array}{c} r_1 \\ \vdots \\ r_m \\ \hdashline \varphi_1 \\ \vdots \\ \varphi_M \\ \hdashline \gamma_1 \\ \vdots \\ \gamma_M \\ \hdashline \kappa_1 \\ \vdots \\ \kappa_M \end{array}\right], \tag{20}$$

$$\Gamma_i = -\sqrt{M} \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \\ \sigma_\theta \mathbb{1}_i & & \mathbf{0} \\ \mathbf{0} & & \sigma_v \mathbb{1}_i \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} r_1,\dots,r_M,\varphi_1,\dots,\varphi_M \\ \left.\vphantom{0}\right\} \gamma_1,\dots,\gamma_M \\ \left.\vphantom{0}\right\} \kappa_1,\dots,\kappa_M \end{array}$$

where $\mathbb{1}_i = diag([\delta_{i,1},\dots,\delta_{i,M}])$.

The transition probabilities from state $\xi(t/\varepsilon) = i$ to $\xi(t/\varepsilon) = j$ are defined in terms of a $M \times M$ generator matrix $Q(t)/\varepsilon$

$$P(\xi(t+\Delta t) = j \mid \xi(t) = i) = q_{ij} + o(\Delta t), \;\; j \neq i, \tag{21}$$

and we choose

$$q_{ij}(t) = 1, \qquad\qquad j \neq i \tag{22}$$

$$q_{ii}(t) = -(M-1), \tag{23}$$

so that the chain has an equal probability of transitioning into any of the M states. It is well known that in the limit of $\varepsilon \to 0$, the evolution of a switching model like (19) under fast Markov switching $\xi(t/\varepsilon)$ converges weakly to an average, or homogenized, model [11, 21]. Because of the symmetry of the transition probabilities (21), the switching model (15)-(18) then converges weakly the same kinematics as the original model (9)-(12). However, unlike the original model, the switching model has the advantage that $B_iB_i^T \propto \Gamma_i\Gamma_i^T \; \forall i$, which will become important in Section 3. Therefore, we seek to control the state of $\mathbf{x}(t)$ as described by the switching kinematic model (19).

2.3 Cost Functional

In our control approach, the agents create and maintain a formation described by a vector of nominal distances. From the perspective of the AiF, these distances are $\mu = [\mu_1, \ldots, \mu_M]^T$, and it computes its respective feedback control $\mathbf{u}$ from a cost functional of the form:

$$J(\mathbf{x},t,\xi) = \min_u \mathbb{E}\left\{ \int_t^{t_f} \left(\frac{1}{2}(h(\mathbf{x}(s)) - \mu)^T A (h(\mathbf{x}(s)) - \mu) + \frac{1}{2}\mathbf{u}^T R \mathbf{u} \right) ds \right\},$$

where $h(\mathbf{x}(t)) = [r_1(t), \ldots, r_M(t)]^T$ is the vector of distances to each neighbor of the AiF, and R is the penalty on control, i.e., turning rate and acceleration control. The nominal distances μ are assumed to be constant over the planning horizon, but any change in μ could reflect the inclusion of dynamic formation changes. We define the general state cost $k(\mathbf{x})$,

$$k(\mathbf{x}) = (h(\mathbf{x}) - \mu)^T A (h(\mathbf{x}) - \mu), \tag{24}$$

to yield the cost-to-go function for the AiF

$$J_i(\mathbf{x},t) \equiv J(\mathbf{x},t,\xi = i) = \min_{\mathbf{u}} \mathbb{E}\left\{ \int_t^{t_f} \frac{1}{2} \left(k(\mathbf{x}(s)) + \mathbf{u}(\mathbf{x}(s))^T R \mathbf{u}(\mathbf{x}(s)) \right) ds \right\}. \tag{25}$$

3 Path Integral Construction

In this section the path integral representation of the switching kinematics model is derived. We begin with the (stochastic) Hamilton-Jacobi-Bellman equation for the model (19) and cost functional (25), which, for the state of the fast-switching Markov chain is $\xi(t/\varepsilon) = i$, is

$$\begin{aligned} 0 = \partial_t J_i + \min_u \Bigg\{ & (f + B_i \mathbf{u})^T \partial_{\mathbf{x}} J_i + \frac{1}{2} \mathrm{Tr}\left(\Gamma_i \Gamma_i^T \partial_{\mathbf{x}}^2 J_i \right) \\ & + \frac{1}{2} k(\mathbf{x}) + \frac{1}{2}\mathbf{u}^T R \mathbf{u} + \frac{Q(t)}{\varepsilon} J(\mathbf{x},t)(i) \Bigg\}, \qquad i = 1, \ldots, M, \end{aligned} \tag{26}$$

where

$$\frac{Q(t)}{\varepsilon} J(\mathbf{x},t)(i) = \frac{1}{\varepsilon} \sum_{j \neq i} q_{ij}(t) \left(J_j(\mathbf{x},t) - J_i(\mathbf{x},t) \right). \tag{27}$$

We have chosen zero terminal cost (at $t = t_f$) for this system of PDEs,

$$J_i(\mathbf{x},t_f) = \phi(\mathbf{x}) = 0, \qquad \forall i,\ \forall \mathbf{x}, \tag{28}$$

and added reflective boundary conditions to constrain the speed of agents to remain between $v_{LB} \leq v \leq v_{UB}$:

$$\partial_{\mathbf{x}} J_i(\mathbf{x},t) \cdot \hat{n} = 0, \qquad \forall i, \ \forall t, \quad \mathbf{x} \in \mathscr{V} \tag{29}$$

$$\begin{aligned} \mathscr{V} = \Big\{ \mathbf{x} \, : \, & \frac{1}{M} \sum_{j=1}^{M} \left(\mathbb{E}\left(\kappa_j\right)\right) + \overline{v_m}(0) = v_{LB} \bigcup \frac{1}{M} \sum_{j=1}^{M} \left(\mathbb{E}\left(\kappa_j\right)\right) + \overline{v_m}(0) = v_{UB} \\ & \bigcup \ -\left(\kappa_m - \mathbb{E}\left(\kappa_m\right)\right) + v_m(0) = v_{LB} \\ & \bigcup \ -\left(\kappa_m - \mathbb{E}\left(\kappa_m\right)\right) + v_m(0) = v_{UB} \Big\}, \end{aligned} \tag{30}$$

at the domain normals $\hat{n}$.

The HJB equation is typically solved numerically (see [23], for example), which is impossible for a problem of this size. However, we can exploit the structure of the formation control problem to formulate the HJB PDE solution as a solution to an equivalent estimation problem through a path integral representation.

The optimal control $\mathbf{u}(\mathbf{x},t,i)$ that minimizes (26) is

$$\mathbf{u}(\mathbf{x},t,i) = -R^{-1} B_i^T \partial_{\mathbf{x}} J_i(\mathbf{x},t), \tag{31}$$

which, when substituted back into the HJB equation, yields:

$$\begin{aligned} 0 = {} & \partial_t J_i + f^T \partial_{\mathbf{x}} J_i - \frac{1}{2} \left(\partial_{\mathbf{x}} J_i\right)^T B_i R^{-1} B_i^T \partial_{\mathbf{x}} J_i \\ & + \frac{1}{2} \mathrm{Tr}\left(\Gamma_i \Gamma_i^T \partial_{\mathbf{x}}^2 J_i\right) + \frac{1}{2} k(\mathbf{x}(t)) + \frac{Q(t)}{\varepsilon} J(\mathbf{x},t)(i). \end{aligned} \tag{32}$$

A logarithmic transformation [10] is applied for each state i:

$$J_i(\mathbf{x},t) = -\lambda \log \Psi_i(\mathbf{x},t), \tag{33}$$

yielding a new PDE

$$\begin{aligned} \frac{1}{\Psi_i} \partial_t \Psi_i = {} & \frac{1}{2\lambda} k(\mathbf{x}) - \frac{f^T}{\Psi_i} \partial_{\mathbf{x}} \Psi_i - \frac{1}{2\Psi_i} \mathrm{Tr}\left(\Gamma_i \Gamma_i^T \partial_{\mathbf{x}}^2 \Psi_i\right) - \frac{Q(t)}{\varepsilon} \log \Psi(\mathbf{x},t)(i) \\ & - \frac{\lambda}{2\Psi_i^2} \left(\partial_{\mathbf{x}} \Psi_i\right)^T B_i R^{-1} B_i^T \partial_{\mathbf{x}} \Psi_i + \frac{1}{2} \frac{1}{\Psi_i^2} \left(\partial_{\mathbf{x}} \Psi_i\right)^T \Gamma_i \Gamma_i^T \partial_{\mathbf{x}} \Psi_i. \end{aligned} \tag{34}$$

In the relative model (17)-(18), it can be seen that the states $\gamma_m(t)$ and $\kappa_m(t)$ collectively describe the evolution of the difference between the AiF control and the unknown control of a neighbor. This suggests that we might be able to compute a control that can, in some sense, compensate for the uncertainty associated with a neighbor's control. Moreover, this implies that a large disturbance in the relative states (i.e., (17)-(18)) likely requires a greater control input, and conversely, that non-actuated states must be noiseless. Because of this, we assume that the noise in the controlled components is inversely proportional to the control penalty, or

$$\Gamma_i \Gamma_i^T = \lambda B_i R^{-1} B_i^T, \qquad \forall i. \tag{35}$$

This selects the value of the control penalty that we shall use in the sequel as

$$R = diag\left(\lambda \sigma_\theta^{-2}, \lambda \sigma_v^{-2}\right). \tag{36}$$

From (36), the quadratic terms on the second line of (34) cancel, and the remaining PDE for Ψ_i is

$$\partial_t \Psi_i = \frac{1}{2\lambda} k(x) \Psi_i - f^T \partial_{\mathbf{x}} \Psi_i - \frac{1}{2} \mathrm{Tr}\left(\Gamma_i \Gamma_i^T \partial_{\mathbf{x}}^2 \Psi_i\right) - \Psi_i \frac{Q(t)}{\varepsilon} \log \Psi(\mathbf{x}, t)(i). \tag{37}$$

Note that this cancellation is only possible in the switching model.

Next, it is shown in Appendix 2 that a first order asymptotic approximation to (37), denoted $\Psi_0(\mathbf{x}, t)$, is independent of the state i of the chain $\xi(t/\varepsilon)$, and that this approximation satisfies the following linear PDE:

$$\begin{aligned}
\partial_t \Psi_0(\mathbf{x}, t) &= -f^T \partial_x \Psi_0(\mathbf{x}, t) - \frac{1}{2} \mathrm{Tr}\left(\Sigma \partial_{\mathbf{x}}^2 \Psi_0\right) + \frac{1}{2\lambda} k(\mathbf{x}) \Psi_0(\mathbf{x}, t) && (38)\\
&= -\left(f^T \partial_{\mathbf{x}} + \frac{1}{2} \mathrm{Tr}\left(\Sigma \partial_{\mathbf{x}}^2\right) - \frac{1}{2\lambda} k(\mathbf{x})\right) \Psi_0 && (39)\\
&= -H \Psi_0(\mathbf{x}, t), && (40)
\end{aligned}$$

where

$$\Sigma = \frac{1}{M} \sum_{j=1}^{M} \Gamma_j \Gamma_j^T \tag{41}$$

is the average of the covariance of the stochastic disturbances in the switching model (15)-(18). The boundary conditions become

$$\begin{aligned}
\Psi_0(\mathbf{x}, t_f) &= \exp(0) = 1, \qquad \forall \mathbf{x} && (42)\\
\partial_{\mathbf{x}} \Psi_0(\mathbf{x}, t) \cdot \hat{n} &= 0, \qquad \forall t,\ \mathbf{x} \in \mathscr{V}. && (43)
\end{aligned}$$

This could be numerically solved backward in time from the terminal condition. However, the Feynman-Kac equations [27, 44] connect certain linear differential operators, H included, to adjoint operators that describe the evolution of a *forward* diffusion process beginning from the current state $\widetilde{\mathbf{x}}(t_0) = \widetilde{\mathbf{x}}_0 = \mathbf{x}$ and ending at $\widetilde{\mathbf{x}}_N = \widetilde{\mathbf{x}}(t_N) = \widetilde{\mathbf{x}}(t_f)$.

In expected value, the solution to (40) is

$$\Psi_0(\widetilde{\mathbf{x}}_0, t_0) = \mathbb{E}_{p(\chi|\widetilde{\mathbf{x}}_0)} \left\{ \exp\left(-\frac{1}{2\lambda} \int_{t_0}^{t_N} k(\mathbf{x}(s)) ds \right) \right\}, \tag{44}$$

where $\widetilde{\mathbf{x}}(t)$ satisfies the path integral-associated, uncontrolled dynamics (cf. (19))

$$d\widetilde{\mathbf{x}}(t) = f(\widetilde{\mathbf{x}}(t))dt + \sqrt{\Sigma}d\mathbf{w} \tag{45}$$

$$\widetilde{\mathbf{x}}(t_0) = \mathbf{x}. \tag{46}$$

Note that since the process $\widetilde{\mathbf{x}}(t)$ is uncontrolled, the expected values of the relative states γ_m and κ_m reduce to $\theta(0) - \theta_m(0)$ and $v(0) - v_m(0)$, respectively. This simplifies the passive components of the uncontrolled model, i.e., $f(\cdot)$ in (15)-(16), to the following form, one for each neighbor m:

$$dr_m(t) = -\kappa(0)\cos(\varphi_m - \theta(0))\,dt - (\kappa_m - v_m(0))\cos(\varphi_m + \gamma_m - \theta(0))\,dt \tag{47}$$

$$d\varphi_m(t) = \kappa(0)\sin(\varphi_m - \theta(0))\,dt + \frac{1}{r_m}(\kappa_m - v_m(0))\sin(\varphi_m + \gamma_m - \theta(0))\,dt \tag{48}$$

The expectation in (44) is taken with respect to the distribution $p(\chi|\widetilde{\mathbf{x}}_0)$ of sample paths χ that begin at $\widetilde{\mathbf{x}}_0 = \mathbf{x}$ and evolve as (45). By discretizing the interval $[t_0, t_N]$ into N intervals of equal length Δt, $t_0 < t_1 < \ldots < t_N$, we can write a sample of the discretized trajectory χ_N^i as

$$\chi_N^i = \left(\widetilde{\mathbf{x}}_1^i, \ldots, \widetilde{\mathbf{x}}_N^i\right),$$

which is sampled from

$$\chi_N^i \sim p(\chi_N|\widetilde{\mathbf{x}}_0) = p(\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0).$$

Under this discretization in time, the solution (44) can be written as

$$\Psi_0(\widetilde{\mathbf{x}}_0, t_0) = \lim_{\Delta t \to 0} \int d\chi_N p(\chi_N|\widetilde{\mathbf{x}}_0) \exp\left[-\frac{\Delta t}{2\lambda}\sum_{k=1}^{N} k(\widetilde{\mathbf{x}}_k)\right], \tag{49}$$

where $d\chi_N = \prod_{k=1}^{N} d\widetilde{\mathbf{x}}_k$ and where $p(\chi_N|\widetilde{\mathbf{x}}_0)$ is the probability of a discretized sample path, conditioned on the starting state $\widetilde{\mathbf{x}}_0$, given by

$$p(\chi_N|\widetilde{\mathbf{x}}_0) = \prod_{k=0}^{N-1} p(\widetilde{\mathbf{x}}_{k+1}|\widetilde{\mathbf{x}}_k). \tag{50}$$

Since, in the uncontrolled process (45), the noise is Gaussian with zero mean and covariance Σ, the transition probabilities may be written as

$$p(\widetilde{\mathbf{x}}_{k+1}|\widetilde{\mathbf{x}}_k) = \frac{1}{\sqrt{2\pi|\Sigma|\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\widetilde{\mathbf{x}}_{k+1} - \widetilde{\mathbf{x}}_k - f(\widetilde{\mathbf{x}}_k)\Delta t)^T \Sigma^{-1} (\widetilde{\mathbf{x}}_{k+1} - \widetilde{\mathbf{x}}_k - f(\widetilde{\mathbf{x}}_k)\Delta t)\right). \tag{51}$$

We can then write the probability of a complete trajectory as

$$p(\chi_N|\widetilde{\mathbf{x}}_0) \propto \exp\left(-\frac{\Delta t}{2}\sum_{k=0}^{N-1}\left(\frac{\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k}{\Delta t}-f(\widetilde{\mathbf{x}}_k)\right)^T \Sigma^{-1}\left(\frac{\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k}{\Delta t}-f(\widetilde{\mathbf{x}}_k)\right)\right). \tag{52}$$

The path integral representation of $\Psi_0(\widetilde{\mathbf{x}}_0,t_0)$ is obtained from equations (49-52), and can be written as an exponential of an "action" [14] $S(\chi_N|\widetilde{\mathbf{x}}_0)$ along the time-discretized sample trajectory $(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N)$:

$$\Psi_0(\widetilde{\mathbf{x}}_0,t_0) = \frac{1}{|2\pi\Sigma\Delta t|^{N/2}}\lim_{\Delta t\to 0}\int d\chi_N \exp\left(-S(\chi_N|\widetilde{\mathbf{x}}_0)\right) \tag{53}$$

$$\begin{aligned} S(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0) &= \sum_{k=1}^{N}\frac{\Delta t}{2\lambda}k(\widetilde{\mathbf{x}}_k) \\ &+\sum_{k=0}^{N-1}\frac{1}{2\Delta t}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k-\Delta t f(\widetilde{\mathbf{x}}_k)\right)^T \\ &\times\Sigma^{-1}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k-\Delta t f(\widetilde{\mathbf{x}}_k)\right). \end{aligned} \tag{54}$$

From (31), (33), the optimal control is given by

$$\begin{aligned} \mathbf{u}(\widetilde{\mathbf{x}}_0,t_0,i) &= \lim_{\Delta t\to 0}\lambda R^{-1}B_i^T\partial_{\widetilde{\mathbf{x}}_0}\log\Psi_0 \\ &= \lim_{\Delta t\to 0}\int d\chi_N P(\chi_N|\widetilde{\mathbf{x}}_0)\mathbf{u}_L(\chi_N|\widetilde{\mathbf{x}}_0,i) \\ &= \lim_{\Delta t\to 0}\mathbb{E}_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\left\{\mathbf{u}_L(\chi_N|\widetilde{\mathbf{x}}_0,i)\right\} \\ &= \mathbb{E}_{P(\chi|\widetilde{\mathbf{x}}_0)}\left\{\mathbf{u}_L(\chi|\widetilde{\mathbf{x}}_0,i)\right\} \end{aligned} \tag{55}$$

where $\lim_{\Delta t\to 0}P(\chi_N|\widetilde{\mathbf{x}}_0) = P(\chi|\widetilde{\mathbf{x}}_0)$ is the probability of an *optimal* trajectory:

$$P(\chi_N|\widetilde{\mathbf{x}}_0) \propto e^{-S(\chi_N|\widetilde{\mathbf{x}}_0)}, \tag{56}$$

and the local controls $\mathbf{u}_L(\chi_N|\widetilde{\mathbf{x}}_0)$ are

$$\mathbf{u}_L(\chi_N|\widetilde{\mathbf{x}}_0,i) = \frac{1}{M}B_i\frac{\widetilde{\mathbf{x}}_1-\widetilde{\mathbf{x}}_0}{\Delta t}, \tag{57}$$

where the B_i selects a pair (γ_m,κ_m) based on the corresponding value of $\xi(t/\varepsilon)$, as in (20). Then (55) is

$$\mathbf{u}(\mathbf{x},i) = \frac{1}{M}B_i\frac{\mathbb{E}_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}-\mathbf{x}}{\Delta t}. \tag{58}$$

In the formulation, after computing $\mathbf{u}(\mathbf{x}(t),t) = \mathbf{u}(\widetilde{\mathbf{x}}_0,t_0)$, each agent executes only the first increment of that control, at which point the optimal control is recomputed for the next time horizon $\left[t_0,t_f\right] = \left[t,t+t_f\right]$. In other words, after computing $\mathbb{E}_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}$, the AiF applies the control $\mathbf{u}$ as chosen by the B_i in (58) to its

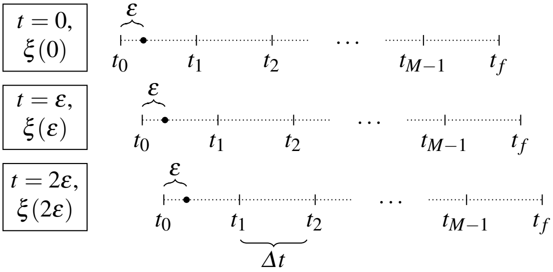
$t = 0$, $\xi(0)$
$t = \varepsilon$, $\xi(\varepsilon)$
$t = 2\varepsilon$, $\xi(2\varepsilon)$
ε
t_0
t_1
t_2
t_{M-1}
t_f
Δt

Fig. 2 Receding-horizon timing. At $t = 0$, the AiF computes $\mathbb{E}_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}$, and randomly chooses a value for ξ. The control applied is the ξ^{th} component of (58). In the next time step ε, the process repeats.

heading angle and speed. Then the problem repeats with a random selection of ξ (see Fig. 2).

The result of this section is a PDF (49) of the system trajectories, including their costs, that is marginalized over each infinitesimal temporal increment of the process under consideration. The optimal control (58) applied by the AiF in state $\mathbf{x}$ is estimated from this trajectory PDF once the joint probability $P(\chi_N|\widetilde{\mathbf{x}}_0)$ has been computed, a nontrivial task to be discussed in the following section.

4 Computing the Control with Kalman Smoothers

In this section we present our approach that estimates the hidden state of a nonlinear stochastic process, which corresponds to the maximally-likely trajectory under state and control costs, using appropriately-chosen noisy measurements.

In general, the marginalization (49) is difficult to evaluate. If one were able to sample K trajectories from the distribution $P(\chi_N|\widetilde{\mathbf{x}}_0)$, approximation of the posterior distribution, i.e., the optimal control (55), would only require a quick calculation:

$$\mathbf{u}(\mathbf{x}) = \frac{1}{K}\sum_{i=1}^{K}\mathbf{u}_L^{(i)}(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0),$$

Along these lines, previous works based on the path integral approach to stochastic optimal control use Markov Chain Monte Carlo (MCMC) techniques [12] to sample from $P(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0)$. Although MCMC techniques can be used to generate samples of the maximally-likely trajectory, we find them to be slow in practice due to the high dimension of this problem ($\chi_N \in \mathbb{R}^{6NM}$).

Moreover, is is not necessary to sample the entire distribution $P(\chi_N|\widetilde{\mathbf{x}}_0)$ since only the value of $\widehat{\mathbf{x}}_1 \equiv \mathbb{E}_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}$ is needed. Note that this estimate is the first infinitesimal increment of the optimal trajectory χ_N^*:

$$\begin{aligned}\chi_N^* &= \underset{\chi_N}{\operatorname{argmax}} \{P(\chi_N|\mu, \widetilde{\mathbf{x}}_0)\} \\ &\propto \underset{\chi_N}{\operatorname{argmax}} \{P(\mu|\chi_N) P(\chi_N|\widetilde{\mathbf{x}}_0)\}.\end{aligned}$$

In this work, we treat the temporal discretization of the optimal trajectory χ_N^* as the hidden state of a stochastic process, where measurements are related to the system goal μ (24). The optimal control can then be computed from the optimal estimate $\widehat{\mathbf{x}}_1$ of the expected value $_{P(\chi_N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}$ given the process and measurements over a fixed interval $t_1, \ldots, t_N$. Therefore, we define the following nonlinear smoothing problem.

Nonlinear Smoothing Problem:
Given measurements $\mathbf{y}_k = \mathbf{y}(t_k)$ for $t_k = t_1, \ldots, t_N = t_f$, where $t_{k+1} - t_k = \Delta t$, compute the estimate $\widehat{\mathbf{x}}_{1:N}$ of the hidden state $\widetilde{\mathbf{x}}_{1:N}$ from the nonlinear state-space model:

$$\widetilde{\mathbf{x}}_{k+1} = \widetilde{\mathbf{x}}_k + \Delta t f(\widetilde{\mathbf{x}}_k) + \varepsilon_k \tag{59}$$

$$\mathbf{y}_k = h(\widetilde{\mathbf{x}}_k) + \eta_k, \tag{60}$$

where $f(\cdot)$ and $h(\cdot)$ are as in Section 2, and ε_k and η_k are independent multivariate Gaussian random variables with zero mean and covariances:

$$\mathbb{E}\left(\varepsilon_k \varepsilon_k^T\right) = \Delta t \Sigma \tag{61}$$

$$\mathbb{E}\left(\eta_k \eta_k^T\right) = \frac{\lambda}{\Delta t} A^{-1}. \tag{62}$$

The smoothing is initialized from $\widetilde{\mathbf{x}}_0 = \mathbf{x}$, the current state of the system as viewed by the AiF. Measurements $\mathbf{y}_k$ are always exactly $\mathbf{y}_k = \mu$. If μ is expected to change over time (e.g., a dynamic formation), then $\mathbf{y}_k = \mu_k$.

Note that only the estimate $\widehat{\mathbf{x}}_1$ is needed to compute the control. The measurement noise in the estimation problem (62) is related to the instantaneous state costs (24), and the process noise (61) is related to the instantaneous control costs (25).

To show that the estimation of $\hat{\mathbf{x}}_1$ can be computed based on the nonlinear smoothing, we write the probability of an estimated hidden state $\widehat{\mathbf{x}}_k$ in the filtering algorithm predication/update steps [13], which is proportional to the measurement likelihood $p(\mathbf{y}_k|\widehat{\mathbf{x}}_k)$ and the predicted state $p(\widehat{\mathbf{x}}_k|\widehat{\mathbf{x}}_{k-1})$:

$$p(\widehat{\mathbf{x}}_k|\widehat{\mathbf{x}}_{k-1}) \propto p(\mathbf{y}_k|\widehat{\mathbf{x}}_k) p(\widehat{\mathbf{x}}_k|\widehat{\mathbf{x}}_{k-1}),$$

where

$$\begin{aligned}p(\mathbf{y}_k|\widehat{\mathbf{x}}_k) \equiv p(\mu_k|\widehat{\mathbf{x}}_k) &= \mathscr{N}\left(h(\widehat{\mathbf{x}}_k), \eta_k \eta_k^T\right) \\ &\propto \exp\left\{-\frac{\Delta t}{2\lambda}(h(\widehat{\mathbf{x}}_k) - \mu)^T A (h(\widehat{\mathbf{x}}_k) - \mu)\right\} \\ &= \exp\left\{-\frac{\Delta t}{2\lambda} k(\widehat{\mathbf{x}}_k)\right\}\end{aligned} \tag{63}$$

and

$$
\begin{aligned}
p(\widehat{\mathbf{x}}_k|\widehat{\mathbf{x}}_{k-1}) &= \mathcal{N}\left(\widehat{\mathbf{x}}_{k-1}+\Delta t f(\widehat{\mathbf{x}}_{k-1}), \Delta t \Sigma\right) \\
&\propto \exp\left\{-\frac{1}{2\Delta t}\left(\widehat{\mathbf{x}}_k-\widehat{\mathbf{x}}_{k-1}-\Delta t f(\widehat{\mathbf{x}}_{k-1})\right)^T\right. \\
&\qquad \left.\times \Sigma^{-1}\left(\widehat{\mathbf{x}}_k-\widehat{\mathbf{x}}_{k-1}-\Delta t f(\widehat{\mathbf{x}}_{k-1})\right)\right\}. \qquad (64)
\end{aligned}
$$

Comparing the right hand sides of (63-64) with (54), it can be seen that the problm of estimating $\hat{\mathbf{x}}_1$ is equivalent to the estimation of $\hat{\mathbf{x}}_1$ in the smoothing problem.

The most-likely trajectory originating from state $\widehat{\mathbf{x}}_0$, that is, the hidden states $\widehat{\mathbf{x}}_k$, $k = 1,\ldots,N$, can be found by filtering and then smoothing the process given the measurements μ_k using a nonlinear smoother, such as an Extended Kalman RTS Smoother (EKF-RTS) or Unscented Kalman RTS Smoother (UKF-RTS) [34]. A nonlinear Kalman smoothing algorithm assumes that the increments given by (63) and (64) are to some extent Gaussian, but the algorithm is sufficiently fast to be applied in *real-time* by each unicycle in a potentially large group with an even larger state space, motivating its use in this work.

The control to be applied in the current state $\mathbf{x}$ is given by (58), using the $\widehat{\mathbf{x}}_1$ estimated by the smoother. After this increment, the process repeats. When the smoothing is complete and agents have applied their computed control, each agent must then observe the actual states of its neighbors so that the next iteration begins with the correct initial condition. In practice, the controller/smoother must be capable of efficiently filtering and smoothing over the horizon $\left[t_0, t_f\right]$. The computational complexity of such a smoother is analyzed in [34].

The effect of scaling parameter λ becomes clear in in the dual estimation formulation. For $\lambda \gg 1$, the measurement noise is large, and the smoother will place more weight on its predictions. Consequently, the passive components of the system $f(\cdot)$ will dominate, and less control will be applied. Similarly, for $\lambda \ll 1$, the smoother will trust the measurements, and a greater amount of control will be applied. The net effect is that λ decides the fraction of the process noise in the original control problem that is propagated into the estimation problem.

Recall that the spatial boundary condition (43) constrains the speeds of the agents within upper and lower limits, i.e. to remain outside the set $\mathcal{V}$. In the context of the smoothing problem, this requires that the probability of a filter prediction, measurement update, or smoothing update to be zero if the estimate enters the boundary $\mathcal{V}$. To deal with such a problem, the smoothing algorithm should be capable of handling inequality constraints. Several algorithms of this type exist (see [2], for example), but in order to keep computation time at a minimum, we instead employ a more straightforward approach. After each prediction step, if the current estimate $\widehat{\mathbf{x}}_k$ is in violation of the constraints, the estimate is projected in a least-squares sense to lie inside the contraint boundaries using Matlab's `lsqlin`. The same method is also applied if the estimate violates the speed constraints during any of the update or smoothing steps.

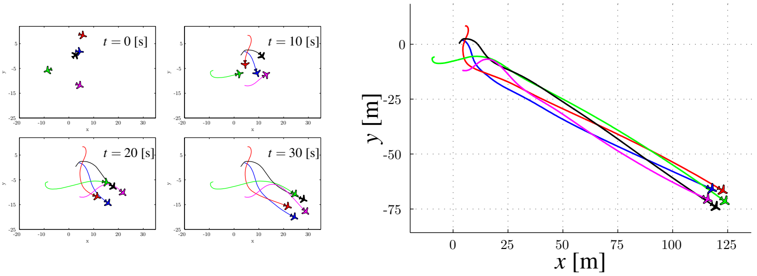

Fig. 3 Five agents, starting from random initial positions and a common speed $v = 2.5$ [m/s], must achieve a regular pentagon formation by an individually-optimal choice of acceleration and turning rate, without any active communication

5 Results

In this section, we apply the methods to a formation control problem in which five agents achieve the formation of a regular pentagon. Each agent is individually estimating the hidden optimal trajectory based on the relative kinematics of all of its neighbors. The instantaneous state cost (24) penalizes the mean squared distance from the unicycle to all of its $M = 4$ neighbors in excess of the side length of the pentagon (5 [m]) or the diagonal of the pentagon, depending on the relative configuration of the pentagon encoded in μ.

The system and control algorithm parameters were chosen as $\lambda = 1$, $\sigma_\theta = \sigma_v = 0.1$, $t_f = 30$ s, $A = 0.1 I_{4\times 4}$, $v_{LB} = 1$ [m/s], $v_{UB} = 3$ [m/s], and $\Delta t = \Delta \varepsilon = 0.1$ s. The control was computed from the result of a Discrete-time Unscented Kalman Rauch-Tung-Striebel Smoother [34]. Fig. 3 shows the trajectories of all agents, while the the inter-agent distances can be seen in Fig. 4. With an initial speed of 2.5 [m/s], the agents never hit their limiting speeds v_{LB} or v_{UB}. Once the pentagon has formed, the agents' heading angles are not equal, and the formation rotates. Without

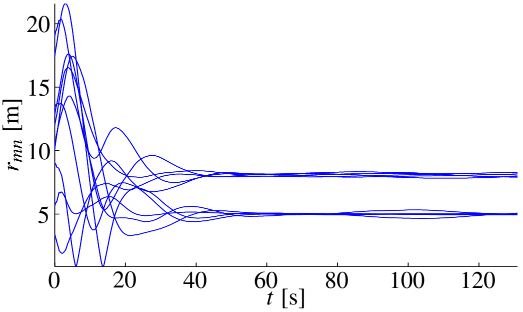

Fig. 4 Inter-agent distance r_{mn} as a function of time. The two radii correspond to the objective pentagon side length (5 [m]) and the pentagon's diagonal ($\frac{5}{2}(1+\sqrt{5}) \approx 8.1$ [m]). The pentagon continues to rotate after forming.

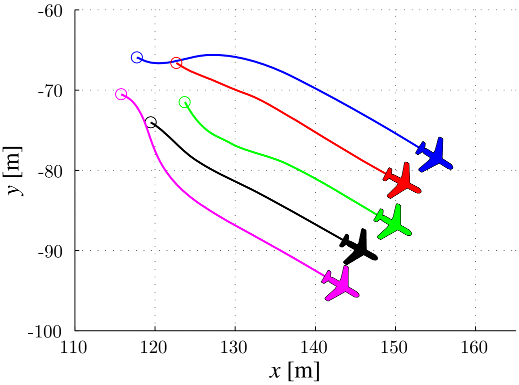

Fig. 5 Transition from a pentagon to a line. The initial condition is the final frame of Fig. 3.

a goal of alignment among the agents, each agent is anticipating the pentagon to rotate (in expected value), and it computes its control so as to maintain its nominal distances in the rotating pentagon.

A dynamic formation was then created by modifying the nominal distances μ during simulation. In Fig. 5, after the pentagon had formed, the formation distances were redefined so that the formation morphed into a line.

6 Discussion

This work considers the problem of unicycle formation control in a distributed optimal feedback control setting. Since this gives rise to a system with a high dimensional state space, we exploit the stochasticity inherent in distributed multi-agent control problems and apply the path integral approach in order to compute the control. The uncertainty in turning rates and accelerations of an agent's neighbors are modeled as stochastic processes, and a fast switching kinematic model links this stochasticity to an agent's control, allowing the optimal control problem to be framed as an estimation problem.

Each agent computes its optimal control in real-time by applying a nonlinear Kalman smoothing algorithm. The measurement noise and process noise of the smoothing problem are created using the structure of the cost function and stochastic kinematics. Aside from mutual observations among agents, the formation is created and maintained without any communication among them.

A number of other goals, e.g., alignment of heading angle, are possible through a simple change in the cost function. More complex aspects of multi-agent formation control, such as collision avoidance and dynamic communication networks, for example, could be handled by a robust smoothing algorithm and will be explored in future research.

Acknowledgements. This work was supported by NSF GRFP 0809125.

References

1. Anderson, B., Fidan, B., Yu, C., Walle, D.: UAV formation control: theory and application. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control, pp. 15–34. Springer (2008)
2. Bell, B.M., Burke, J.V., Pillonetto, G.: An inequality constrained nonlinear Kalman-Bucy smoother by interior point likelihood maximization. Automatica 45(1), 25–33 (2009)
3. Bellingham, J.G., Rajan, K.: Robotics in remote and hostile environments. Science 318(5853), 1098–1102 (2007)
4. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-dynamic programming. Athena Scientific, Belmont (1996)
5. van den Broek, B., Wiegerinck, W., Kappen, B.: Graphical model inference in optimal control of stochastic multi-agent systems. Journal of Artificial Intelligence Research 32(1), 95–122 (2008)
6. van den Broek, B., Wiegerinck, W., Kappen, B.: Optimal Control in Large Stochastic Multi-agent Systems. In: Tuyls, K., Nowe, A., Guessoum, Z., Kudenko, D. (eds.) Adaptive Agents and MAS III. LNCS (LNAI), vol. 4865, pp. 15–26. Springer, Heidelberg (2008)
7. Bullo, F., Cortes, J., Martinez, S.: Distributed control of robotic networks: A mathematical approach to motion coordination algorithms. Princeton University Press, Princeton (2009)
8. Dimarogonas, D.: On the rendezvous problem for multiple nonholonomic agents. IEEE Transactions on Automatic Control 52(5), 916–922 (2007)
9. Elkaim, G., Kelbley, R.: A Lightweight Formation Control Methodology for a Swarm of Non-Holonomic Vehicles. In: IEEE Aerospace Conference. IEEE, Big Sky (2006)
10. Fleming, W., Soner, H.: Logarithmic Transformations and Risk Sensitivity. In: Controlled Markov Processes and Viscosity Solutions, ch.6. Springer, Berlin (1993)
11. Freidlin, M., Wentzell, A.: Random Perturbations of Dynamical Systems. Springer (1984)
12. Gamerman, D.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall (1997)
13. Gelb, A.: Applied Optimal Estimation. The M.I.T. Press, Cambridge (1974)
14. Goldstein, H.: Classical Mechanics, 2nd edn. Addison-Wesley (1980)
15. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, 3rd edn. North-Holland (2007)
16. Kappen, H.: Linear Theory for Control of Nonlinear Stochastic Systems. Physical Review Letters 95(20), 1–4 (2005)
17. Kappen, H.: Path integrals and symmetry breaking for optimal control theory. Journal of Statistical Mechanics: Theory and Experiment 2005, P11,011 (2005)
18. Kappen, H., Wiegerinck, W., van den Broek, B.: A path integral approach to agent planning. In: Autonomous Agents and Multi-Agent Systems, Citeseer (2007)
19. Kappen, H.J.: Path integrals and symmetry breaking for optimal control theory. Journal of Statistical Mechanics, Theory and Experiment 2005, 21 (2005)
20. Kappen, H.J., Gómez, V., Opper, M.: Optimal control as a graphical model inference problem. Machine Learning 87(2), 159–182 (2012)
21. Khas'minskii, R.: On the principle of averaging the Itô's stochastic differential equations. Kybernetika 4(3), 260–279 (1968)
22. Kumar, V., Rus, D., Sukhatme, G.S.: Networked Robotics. In: Sciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics. ch. 41, pp. 943–958 (2008)

23. Kushner, H.J., Dupuis, P.: Numerical Methods for Stochastic Control Problems in Continuous Time, 2nd edn. Springer (2001)
24. Milutinović, D.: Utilizing Stochastic Processes for Computing Distributions of Large-Size Robot Population Optimal Centralized Control. In: Martinoli, A., Mondada, F., Correll, N., Mermoud, G., Egerstedt, M., Hsieh, M.A., Parker, L.E., Støy, K. (eds.) Distributed Autonomous Robotic Systems. STAR, vol. 83, pp. 359–372. Springer, Heidelberg (2013)
25. Milutinović, D., Garg, D.P.: Stochastic model-based control of multi-robot systems. Tech. Rep. 0704, Duke University, Durham, NC (2009)
26. Milutinović, D., Garg, D.P.: A sampling approach to modeling and control of a large-size robot population. In: Proceedings of the 2010 ASME Dynamic Systems and Control Conference. ASME, Boston (2010)
27. Oksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 6th edn. Springer, Berlin (2003)
28. Palmer, A., Milutinović, D.: A Hamiltonian Approach Using Partial Differential Equations for Open-Loop Stochastic Optimal Control. In: Proceedings of the 2011 American Control Conference, San Francisco, CA (2011)
29. Parker, L.E.: Multiple Mobile Robot Systems. In: Sciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics, ch.40, pp. 921–941. Springer (2008)
30. Powell, W.B.: Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley Interscience, Hoboken (2007)
31. Ren, W., Beard, R.: Distributed consensus in multi-vehicle cooperative control: Theory and applications. Springer, New York (2007)
32. Ren, W., Beard, R., Atkins, E.: A survey of consensus problems in multi-agent coordination. In: Proceedings of the 2005, American Control Conference, pp. 1859–1864 (2005)
33. Ryan, A., Zennaro, M., Howell, A., Sengupta, R., Hedrick, J.: An overview of emerging results in cooperative UAV control. In: 2004 43rd IEEE Conference on Decision and Control, vol. 1, pp. 602–607 (2004)
34. Särkkä, S.: Continuous-time and continuous-discrete-time unscented Rauch-Tung-Striebel smoothers. Signal Processing 90(1), 225–235 (2010)
35. Singh, A., Batalin, M., Stealey, M., Chen, V., Hansen, M., Harmon, T., Sukhatme, G., Kaiser, W.: Mobile robot sensing for environmental applications. In: Field and Service Robotics, pp. 125–135. Springer (2008)
36. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)
37. Theodorou, E., Buchli, J., Schaal, S.: A Generalized Path Integral Control Approach to Reinforcement Learning. The Journal of Machine Learning Research 11, 3137–3181 (2010)
38. Theodorou, E., Buchli, J., Schaal, S.: Reinforcement learning of motor skills in high dimensions: A path integral approach. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), vol. 4, pp. 2397–2403. IEEE (2010)
39. Todorov, E.: General duality between optimal control and estimation. In: 47th IEEE Conference on Decision and Control, vol. 5, pp. 4286–4292. IEEE, Cancun (2008)
40. Wang, M.C., Uhlenbeck, G.: On the theory of Brownian Motion II. Reviews of Modern Physics 17(2-3), 323–342 (1945)
41. Wiegerinck, W., van den Broek, B., Kappen, B.: Stochastic optimal control in continuous space-time multi-agent systems. In: Proceedings UAI. Citeseer (2006)
42. Wiegerinck, W., van den Broek, B., Kappen, B.: Optimal on-line scheduling in stochastic multiagent systems in continuous space-time. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, p. 1 (2007)

43. Yin, G.G., Zhu, C.: Hybrid Switching Diffusions. Springer, New York (2010)
44. Yong, J.: Relations among ODEs, PDEs, FSDEs, BDSEs, and FBSDEs. In: Proceedings of the 36th IEEE Conference on Decision and Control, pp. 2779–2784. IEEE, San Diego (1997)

Appendix 1

Here we derive the switching model (15)-(18) from the original model (9)-(14). First, note that the relative angle and relative speed satisfy

$$\gamma_m(t) = (\theta(0) - \theta_m(0)) + \int_0^t \omega(\mathbf{x}(s))ds - \int_0^t \sigma_\theta dw_{m,\theta} \tag{65}$$

$$\kappa_m(t) = (v(0) - v_m(0)) + \int_0^t u(\mathbf{x}(s))ds - \int_0^t \sigma_v dw_{m,v}, \tag{66}$$

from which we may obtain

$$\mathbb{E}(\gamma_m) = \theta(0) - \theta_m(0) + \int_0^t \omega dt \tag{67}$$

$$\mathbb{E}(\gamma_m) - \gamma_m = \int \sigma_\theta dw_{m,\theta} \tag{68}$$

$$\mathbb{E}(\kappa_m) = v(0) - \kappa_m(0) + \int_0^t udt \tag{69}$$

$$\mathbb{E}(\kappa_m) - \kappa_m = \int \sigma_v dw_{m,v}. \tag{70}$$

Then the heading angles and speeds of the AiF and its neighbor m can both be encoded into γ_m and κ_m by the relations:

$$\theta(t) = \mathbb{E}\left\{ \gamma_m(t) + \theta_m(0) + \int_0^t \sigma_\theta dw_{m,\theta} \right\} = \mathbb{E}\left(\gamma_m(t)\right) + \theta_m(0) \tag{71}$$

$$\theta_m(t) = -\left(\gamma_m - \mathbb{E}\left(\gamma_m(t)\right)\right) + \theta_m(0) \tag{72}$$

$$v(t) = \mathbb{E}\left\{ \kappa_m(t) + v_m(0) + \int_0^t \sigma_\theta dw_{m,v} \right\} = \mathbb{E}\left(\kappa_m(t)\right) + v_m(0) \tag{73}$$

$$v_m(t) = -\left(\kappa_m - \mathbb{E}\left(\kappa_m(t)\right)\right) + v_m(0). \tag{74}$$

We assume that only one pair (γ_m, κ_m) evolves at a time. Introducing $\delta_{\xi(t/\varepsilon),m}$ as the Kronecker delta selecting the evolution of the pair m, we would have that

$$d\gamma_m(t) = (\omega dt - \sigma_\theta dw_{\theta,m})\delta_{\xi(t/\varepsilon),m} \tag{75}$$

$$d\kappa_m(t) = (udt - \sigma_v dw_m)\delta_{\xi(t/\varepsilon),m}. \tag{76}$$

However, we wish for the average evolution of the states to be the same as in the original problem formulation. Since each pair m is selected with frequency M^{-1}, we write the evolution of these relative states as

$$d\gamma_m(t) = \left(M\omega dt - \sqrt{M}\sigma_\theta dw_{\theta,m}\right)\delta_{\xi(t/\varepsilon),m} \tag{77}$$

$$d\kappa_m(t) = \left(Mudt - \sqrt{M}\sigma_v dw_m\right)\delta_{\xi(t/\varepsilon),m}. \tag{78}$$

Next, we substitute (72) and (74) into $\theta_m(t)$ and $v_m(t)$, respectively, in the kinematic model for $r_m(t)$ and $\varphi_m(t)$. Finally, we also substitute the averages for $\theta(t)$ and $v(t)$:

$$\theta(t) = \frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\gamma_j(t)\right) + \theta_m(0)\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\gamma_j(t)\right)\right) + \overline{\theta_m}(0) \tag{79}$$

$$v(t) = \frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\kappa_j(t)\right) + v_m(0)\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\mathbb{E}\left(\kappa_j(t)\right)\right) + \overline{v_m}(0). \tag{80}$$

Appendix 2

Here we develop a first approximation to (37), reproduced here:

$$\partial_t\Psi_i = \frac{1}{2\lambda}k(x)\Psi_i - f^T\partial_{\mathbf{x}}\Psi_i - \frac{1}{2}\mathrm{Tr}\left(\Gamma_i\Gamma_i^T\partial_{\mathbf{x}}^2\Psi_i\right) - \Psi_i\frac{Q(t)}{\varepsilon}\log\Psi(\mathbf{x},t)(i).$$

This derivation follows closely to that in Chapter 11 of [43].

We seek to find an approximation to $\Psi_i(\mathbf{x},t)$, and begin with an asymptotic expansion to $J_i(\mathbf{x},t)$ of the form

$$J_i(\mathbf{x},t) = A_0(\mathbf{x},t,i) + \varepsilon A_1(\mathbf{x},t,i) + B_0(\mathbf{x},\tau,i) + \varepsilon B_1(\mathbf{x},\tau,i), \qquad i = 1,\ldots,M$$

where $\tau = (t_f - t)/\varepsilon$ is a stretched-time variable, the $A_k(\cdot)$'s are outer expansion terms, and $B_k(\cdot)$'s are terminal layer correction terms. The expansion terms are matched at terminal condition (28) with

$$A_0(\mathbf{x},t_f,i) + B_0(\mathbf{x},0,i) = \phi(\mathbf{x}) = 0 \tag{81}$$

$$A_1(\mathbf{x},t_f,i) + B_1(\mathbf{x},0,i) = 0. \qquad i = 1,\ldots,M \tag{82}$$

From (33), define the transformed expansion terms as

$$a_k(\mathbf{x},t,i) = \exp(-A_k(\mathbf{x},t,i)/\lambda) \tag{83}$$

$$b_k(\mathbf{x},\tau,i) = \exp(-B_k(\mathbf{x},\tau,i)/\lambda), \qquad i = 1,\ldots,M,\ k = 0,1. \tag{84}$$

Substituting the outer expansion terms $a_k(\cdot)$ into (37) and collecting terms by powers of ε, we obtain

$$\varepsilon^0 \;:\; Q(t)a_0(\mathbf{x},t,\cdot)(i) = 0 \tag{85}$$

$$\begin{aligned}\varepsilon^1 \;:\; \partial_t a_0(\mathbf{x},t,i) &= \frac{1}{2\lambda}k(\mathbf{x})a_0(\mathbf{x},t,i) - f^T\partial_{\mathbf{x}}a_0(\mathbf{x},t,i) - \frac{1}{2}\mathrm{Tr}\left(\Gamma_i\Gamma_i^T\partial_{\mathbf{x}}^2 a_0(\mathbf{x},t,i)\right)\\ &\quad - a_0(\mathbf{x},t,i)Q(t)\log a_1(\mathbf{x},t,\cdot)(i).\end{aligned} \tag{86}$$

Writing $a_0(\mathbf{x},t) = [a_0(\mathbf{x},t,1),\ldots,a_0(\mathbf{x},t,M)]^T$, we have from (85) that

$$Q(t)a_0(\mathbf{x},t) = 0.$$

Then from (22)-(23), the rank of $Q(t)$ is $M-1$, implying that the null-space of $Q(t)$ is one dimensional and spanned by a vector of all ones, $\mathbb{1} = [1,\ldots,1]^T$. Then $a_0(\mathbf{x},t)$ must be independent of i, and so

$$a_0(\mathbf{x},t) = \Psi_0(\mathbf{x},t)\mathbb{1}. \tag{87}$$

Note that this condition on $Q(t)$ further implies the existence of a quasi-stationary distribution [43] $\nu(t) = [\nu_1(t),\ldots,\nu_M(t)]$ with the properties that $\sum_{i=1}^M \nu_i = 1$ and $\nu(t)Q(t) = 0$. Substituting (87) into (86), left multiplying by ν_i, and summing over i gives

$$\begin{aligned}\sum_{i=1}^M \nu_i\partial_t\Psi_0(\mathbf{x},t) &= \sum_{i=1}^M \nu_i\frac{1}{2\lambda}k(\mathbf{x})\Psi_0(\mathbf{x},t) - \sum_{i=1}^M \nu_i f^T\partial_{\mathbf{x}}\Psi_0(\mathbf{x},t)\\ &\quad - \sum_{i=1}^M \nu_i\frac{1}{2}\mathrm{Tr}\left(\Gamma_i\Gamma_i^T\partial_{\mathbf{x}}^2\Psi_0(\mathbf{x},t)\right) - \Psi_0(\mathbf{x},t)\sum_{i=1}^M \nu_i Q(t)\log a_1(\mathbf{x},t,\cdot)(i).\end{aligned} \tag{88}$$

The properties of $\nu(t)$ cause the last term to drop out, and the ν_i's in the first three sums add to one.

$$\partial_t\Psi_0(\mathbf{x},t) = \frac{1}{2\lambda}k(\mathbf{x})\Psi_0(\mathbf{x},t) - f^T\partial_{\mathbf{x}}\Psi_0(\mathbf{x},t) - \sum_{i=1}^M \nu_i\frac{1}{2}\mathrm{Tr}\left(\Gamma_i\Gamma_i^T\partial_{\mathbf{x}}^2\Psi_0(\mathbf{x},t)\right). \tag{89}$$

Next, since Γ_i selects the i^{th} pair of $(\gamma_m^{\pm},\kappa_m^{\pm})$ and multiplies them by $\sqrt{M}$, the remaining sum represents a consolidation of the diffusion terms associated with each of the pairs $(\gamma_m^{\pm},\kappa_m^{\pm})$. Then in light of the chosen symmetry of $Q(t)$ (22)-(23), this sum reduces to an average diffusion with covariance

$$\Sigma = \frac{1}{M}\sum_{i=1}^M \Gamma_i\Gamma_i^T,$$

which is (41), and $\Psi_0(\mathbf{x},t)$ satisfies

$$\partial_t \Psi_0(\mathbf{x},t) = \frac{1}{2\lambda} k(\mathbf{x}) \Psi_0(\mathbf{x},t) - f^T \partial_{\mathbf{x}} \Psi_0(\mathbf{x},t) - \frac{1}{2} \mathrm{Tr}\left(\Sigma \partial_{\mathbf{x}}^2 \Psi_0(\mathbf{x},t)\right). \tag{90}$$

which is (37), and with terminal condition $\Psi_0(\mathbf{x},t_f) = \exp(\phi(\mathbf{x})) = 1$, which is (42).

Next we consider the terminal correction terms $b_k(\mathbf{x},\tau,i)$. Rewriting the original PDE in the timescale of τ,

$$-\frac{1}{\varepsilon} \partial_\tau \Psi_i = \frac{1}{2\lambda} k(x) \Psi_i - f^T \partial_{\mathbf{x}} \Psi_i - \frac{1}{2} \mathrm{Tr}\left(\Gamma_i \Gamma_i^T \partial_{\mathbf{x}}^2 \Psi_i\right) - \Psi_i \frac{Q(t_f - \varepsilon\tau)}{\varepsilon} \log \Psi(\mathbf{x},t)(i), \quad i = 1,\ldots,M \tag{91}$$

and expanding $Q(\cdot)$ around t_f,

$$Q(t_f - \varepsilon\tau) \approx Q(t_f) - (\varepsilon\tau)\, Q'(t)\big|_{t=t_f}, \tag{92}$$

we can obtain, using the same method as before,

$$\partial_\tau b_0(\mathbf{x},\tau,i) = b_0(\mathbf{x},\tau,i) Q(t_f) \log b_0(\mathbf{x},\tau,\cdot)(i). \tag{93}$$

From (81) and (42), this implies that $b_0(\mathbf{x},\tau,i) = 1$ for all time and states i. We do not derive asymptotic error bounds here.