Accepted for Proceedings of the 11th International Symposium on
Distributed Autonomous Robotic Systems (DARS), 2012

# A Stochastic Optimal Enhancement of Feedback Control for Unicycle Formations

Ross P. Anderson and Dejan Milutinović

**Abstract** We consider an optimal feedback control approach for multiple nonholonomic vehicles to achieve a distance-based formation with their neighbors using only local observations. Beginning with a non-optimal feedback control for collision-free flocking, each agent determines an additive correction term to its non-optimal control from an elliptic Hamilton-Jacobi-Bellman equation so that its actions are optimal and robust to uncertainty. In order to avoid offline spatial discretization of the stationary, high-dimensional cost-to-go function, we exploit the stochasticity of the distributed nature of the problem to develop an equivalent estimation problem in a continuous state space using a path integral representation. Consequently, each agent independently computes its optimal feedback control using a discrete-time Unscented Kalman smoother. Our approach is illustrated by a numerical example in which five agent achieve a pentagon with aligned and equal velocities.

## 1 Introduction

Nonholonomic vehicle formations, in which each agent is tasked with attaining and maintaining pre-specified distances from its neighbors, are beginning to demonstrate its significance and potential impact in a variety of applications in both the public and private sector [20]. The control approaches in relation to this problem have typically relied on stability analyses, or ad-hoc artificial potential functions [1, 4, 5, 6, 21, 22, 24].

Although previously-considered control approaches lead to satisfying results, they are non-optimal, and stability is usually only proven in the deterministic case. Because of this, we think it fruitful to examine the additional control input neces-

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

sary to drive the non-optimal system into a formation *optimally* and in a manner that is robust to uncertainty. In this work, we begin with a non-optimal feedback control policy [24], which provides an artificial potential function for distributed and collision-free flocking of deterministic nonholonomic vehicles. Then we introduce a stochastic optimal feedback control for each agent defining an additive control input in order to reach the formation in an optimal way. Consequently, our control approach is optimal, and, due to the adopted artificial potential function [24], it provides collision-free agent trajectories.

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 state space dimension of multi-agent systems makes this conventional approach to stochastic optimal control impossible. In this work, we exploit the distributed nature of the problem at hand in order to make its solution tractable. The distributed formation control problem is inherently stochastic – from the perspective of one agent, neighbors' control inputs are unknown, as well as the consequences of these inputs to agent trajectories due to agent model uncertainties. 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 a prior for the unknown control input of an agent can be robustly described as Brownian motion [12]. Based on our prior and the system kinematics, we can induce a probability distribution of the relative state $\mathbf{x}$ to all neighbors in an interval $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ at a particular future time [27]. 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. In particular, the relation between the solutions to optimal control PDEs and the probability distribution of stochastic differential equations [8, 18, 30], allows certain stochastic optimal control problems to be written as an estimation problem on the distribution of optimal trajectories in continuous state space, in a manner known as the path integral (PI) approach (see [13, 14, 25, 26] as well as [15] and references therein for a more recent review of results, and see [17, 19] for an analogous approach in the open-loop control case).

Multiagent systems have previously been studied using the PI approach [2, 3, 28, 29], but in these works, the agents cooperatively compute their control from a marginalization of the joint probability distribution of the group's trajectory. In this paper, we develop a method by which agents independently compute their controls without explicit communication. Moreover, previous works using the PI approach have formulated a receding horizon optimal control problem for which stability is difficult to guarantee [11]. We therefore consider an elliptic control problem over a planning horizon that ends only when the formation is reached. In this sense, each agent is estimating both their optimal control and the time that the formation will be achieved. Finally, our work differs from previous PI implementations in that the optimal feedback control is computed independently by each agent from nonlinear Kalman smoothing algorithms.

This paper is organized as follows. Section 2 introduces the formation control problem as viewed by a single agent in the group, followed by a derivation of a path integral representation in Section 3. Section 4 presents a Kalman smoother method for computing individual agent control. Section 5 illustrates our methods with a simulated five-agent formation, and we conclude with Section 6.

## 2 Control Problem Formulation

We consider a team of agents, each described by a Cartesian position $(x_m, y_m)$, a heading angle $\theta_m$, a speed $v_m$, and the kinematic model:

$$\begin{aligned}
dx_m(t) &= v_m \cos\theta_m dt \\
dy_m(t) &= v_m \sin\theta_m dt \\
d\theta_m(t) &= \omega_m dt + \sigma_{\theta,m} dw_{\theta,m} \\
dv_m(t) &= u_m dt + \sigma_{v,m} dw_{v,m},
\end{aligned} \tag{1}$$

where $\omega_m$ and $u_m$ are the feedback-controlled turning rate and acceleration, respectively, and where $dw_{\theta,m}$ and $dw_{v,m}$ are mutually independent Wiener process increments with corresponding intensities $\sigma_{\theta,m}$ and $\sigma_{v,m}$, respectively.

To achieve a distributed control, the problem is formulated from the perspective of just one agent, which we call the agent-in-focus, or AiF for short. We notate the state $(x, y, \theta, v)$ of the AiF *sans* subscript. The AiF observes $M$ neighbors, irrespective of the total number of agents in the population. This agent computes an optimal feedback control to reach a formation with respect to observed neighbors. The formation is achieved when the inter-agent distances $r_m = \sqrt{(x-x_m)^2 + (y-y_m)^2}$ reach a set of predefined nominal distances $\delta_m$ within a tolerance $\epsilon_r$, the agents' heading angles are aligned within a tolerance $\epsilon_\theta$, and the speeds are equal within a tolerance $\epsilon_v$. For the AiF, this occurs when its perspective of the system state $\mathbf{x}$ belongs to a set $\mathscr{F}$:

$$\begin{aligned}
\mathscr{F} = \big\{ \mathbf{x} : & \, |r_m - \delta_m| \le \epsilon_r, \\
& |\theta - \theta_m| \le \epsilon_\theta, \\
& |v - v_m| \le \epsilon_v, \;\; m = 1, \ldots, M \big\}.
\end{aligned} \tag{2}$$

The AiF further assumes no information about the observations made by its neighbors. In other words, the AiF assumes that its observed neighbors are only observing the AiF. This type of scenario is illustrated in Fig. 1.

We introduce collision avoidance by adding to the original kinematic model the artificial potential function [24] for collision-free velocity vector alignment in populations of vehicles described by a noiseless version of (1). The evolution equations for $\theta(t)$ and $v(t)$ become

$$d\theta(t) = \omega_{\mathrm{D}} dt + \omega dt + \sigma_\theta dw_\theta, \qquad dv(t) = u_{\mathrm{D}} dt + u dt + \sigma_v dw_v,$$

Accepted for Proceedings of the 11th International Symposium on Distributed Autonomous Robotic Systems (DARS), 2012

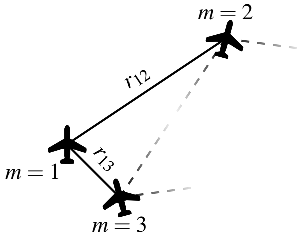

**Fig. 1** In this scenario, agent 1 (the AiF) observes neighbors 2 and 3 (solid lines) and attempts to achieve the inter-agent spacings $r_{12} = \delta_{12}$ and $r_{13} = \delta_{13}$, as well as alignment of heading angles and speeds. Agent 1 is unaware of any other observation connectivity (dashed lines).

so that the controls $\omega$ and $u$ are interpreted as the optimal *correction* terms to the known turning rate feedback control $\omega_{\text{D}}$ and known acceleration feedback control $u_{\text{D}}$, respectively. For brevity, the reader is directed to [24] for the relevant equations for $\omega_{\text{D}}$ and $u_{\text{D}}$. Suffice it to say that in the deterministic case ($\sigma_{\theta,m} = \sigma_{v,m} = 0$), the non-optimal feedback control $\omega_{\text{D}}$ and $u_{\text{D}}$ ensures collision avoidance and provides an alignment force of the vehicle velocity vectors. Moreover, it guarantees that the group will tend toward a minimum of the an artificially-constructed potential energy $V(r_m)$. Our specific choice of $V(r_m) = \delta_m^2 ||r_m||^{-2} + 2\log||r_m||$ (see [24]) suggests that the force applied to agents by the potential reaches minimum value when all inter-agent distances $r_m \to \delta_m$. Note that from the perspective of the AiF, the non-optimal controls executed by a neighboring agent $m$ are based on that neighbor's sole observation, i.e., the observation of the AiF, and independent of others in the population, as before.

Finally, we look to the evolution equations for the heading angle $\theta_m(t)$ and speed $v_m(t)$ of a neighbor $m$ to the AiF. In our formulation, the optimal controls $\theta_m(t)$ and $v_m(t)$ will be computed by the AiF, while assuming that agent $m$ is only observing the AiF. The distributed nature of this problem is preserved, since during simulation, the control executed by agent $m$ will differ from what is expected from it. Therefore, the control the AiF computes for agent $m$ is modeled as the mean value of a Gaussian random variable:

$$d\theta_m(t) = \omega_{\text{D},m}dt + N\left(\omega_m dt, \sigma_{\theta,m}^2 dt\right), \qquad dv_m(t) = u_{\text{D},m}dt + N\left(u_m dt, \sigma_{v,m}^2 dt\right),$$

where $\sigma_{\theta,m}$ and $\sigma_{v,m}$ take into account both kinematic uncertainty *and control uncertainty*, so that $\sigma_{\theta,m} > \sigma_\theta$ and $\sigma_{v,m} > \sigma_v$. In summary, we have a kinematic model for the AiF of the form

$$
\begin{aligned}
dx(t) &= v \cos\theta dt && (3)\\
dy(t) &= v \sin\theta dt && (4)\\
d\theta(t) &= \omega_{\text{D}} dt + \omega dt + \sigma_\theta dw_\theta && (5)\\
dv(t) &= u_{\text{D}} dt + u dt + \sigma_v dw_v && (6)\\
dx_m(t) &= v_m \cos\theta_m dt && (7)\\
dy_m(t) &= v_m \sin\theta_m dt && (8)\\
d\theta_m(t) &= \omega_{\text{D},m} dt + \omega_m dt + \sigma_{\theta,m} dw_{\theta,m} && (9)\\
dv_m(t) &= u_{\text{D},m} dt + u_m dt + \sigma_{v,m} dw_{v,m}, \qquad m = 1,\ldots,M. && (10)
\end{aligned}
$$

This model can be written in a general form:

$$d\mathbf{x}(t) = f(\mathbf{x})dt + B\mathbf{u}dt + \Gamma d\mathbf{w}, \tag{11}$$

where the state vector $\mathbf{x}$ includes the system state from the perspective of the AiF, $f(\mathbf{x})$ captures the kinematics including the deterministic, collision-avoiding controls, and $\mathbf{u} = [\omega, u, \omega_1, u_1, \ldots, \omega_M, u_M]^T$ is a vector of optimal feedback controls to be computed by the AiF. The Wiener process $d\mathbf{w}$ captures the uncertainty due to model kinematics for each agent, as well as the uncertainty due to the control executed by neighboring agents $m = 1,\ldots,M$.

Our goal is to compute the feedback controls $\mathbf{u}(\mathbf{x})$ that minimize the total accumulated cost until the formation is reached (a problem sometimes called control until a target set is reached). We define the following cost functional for the AiF:

$$J(\mathbf{x}) = \min_{\mathbf{u}} \mathbb{E}\left\{ \int_0^\tau \frac{1}{2}\Big(k(\mathbf{x}) + \mathbf{u}^T R \mathbf{u}\Big) ds \right\}, \tag{12}$$

where $\tau = \inf\{t > 0 : \mathbf{x}(t) \in \mathscr{F}\}$ is a (finite) first exit time, i.e., the first time that the state reaches a the formation $\mathscr{F}$ (2). We note that, unlike previous works that either use a receding horizon approach or fix a final time, the final time $\tau$ is not known in advance. The instantaneous state cost $k(\mathbf{x})$,

$$k(\mathbf{x}) = (h(\mathbf{x}) - \boldsymbol{\mu})^T Q (h(\mathbf{x}) - \boldsymbol{\mu}), \tag{13}$$

is a quadratic that reaches minimum value when the distances to each of the AiF's $M$ neighbors equal the nominal distances $\delta_m$, the heading angles are equal, and the speeds are equal:

$$
\begin{aligned}
h(\mathbf{x}) &= [r_1, \ldots, r_M, \theta - \theta_1, \ldots, \theta - \theta_M, v - v_1, \ldots, v - v_M]^T && (14)\\
\boldsymbol{\mu} &= [\delta_1, \ldots, \delta_M, 0, \ldots, 0]^T . && (15)
\end{aligned}
$$

where $Q$ is a diagonal positive definite matrix. Note that this instantaneous state cost reaches minimum value when the potential energy associated with noiseless control also reaches minimum value. However, the noiseless controls also prevent collisions using an infinite potential energy when inter-agent distances approach $r_m = 0$. Since the correction control $\mathbf{u}$ is penalized with a quadratic control penalty provided by $R$, it will not overcome this barrier, and collision avoidance is still ensured.

## 3 Path Integral Representation

In this section we show how the optimal control problem can be represented as a path integral over possible system trajectories. The derivation is similar to that used in previous works, but the new type of cost functional used in this paper warrants a new derivation. The (stochastic) Hamilton-Jacobi-Bellman equation for the model (11) and cost functional (12) is

$$0=\min_{\mathbf{u}}\left\{(f+B\mathbf{u})^T\partial_{\mathbf{x}}J+\frac{1}{2}\mathrm{Tr}\left(\Sigma\partial_{\mathbf{x}}^2J\right)+\frac{1}{2}k(\mathbf{x})+\frac{1}{2}\mathbf{u}^TR\mathbf{u}\right\}, \tag{16}$$

where $\Sigma = \Gamma\Gamma^T$. We have chosen boundary conditions for this PDE as

$$J(\mathbf{x}(\tau))=0, \qquad \mathbf{x}\in\mathscr{F}. \tag{17}$$

The HJB equation must typically be solved numerically in a discretized state space until a steady state is reached (see [16], for example). However, the structure of the problem at hand allows us to avoid this through a suitable transformation.

The time-invariant optimal control $\mathbf{u}(\mathbf{x})$ that minimizes (16) is

$$\mathbf{u}(\mathbf{x})=-R^{-1}B^T\partial_{\mathbf{x}}J, \tag{18}$$

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

$$0=f^T\partial_{\mathbf{x}}J-\frac{1}{2}\left(\partial_{\mathbf{x}}J\right)^TBR^{-1}B^T\partial_{\mathbf{x}}J+\frac{1}{2}\mathrm{Tr}\left(\Sigma\partial_{\mathbf{x}}^2J\right)+\frac{1}{2}k(\mathbf{x}(t)). \tag{19}$$

Next, we apply a logarithmic transformation [7] $J(\mathbf{x}) = -\lambda\log\Psi(\mathbf{x})$ for constant $\lambda > 0$ to obtain a new PDE

$$\begin{aligned}0=\frac{k(\mathbf{x})}{2\lambda}&-\frac{f^T}{\Psi}\partial_{\mathbf{x}}\Psi-\frac{1}{2}\frac{1}{\Psi}\mathrm{Tr}\left(\Sigma\partial_{\mathbf{x}}^2\Psi\right)\\&-\frac{1}{2}\frac{\lambda^2}{\Psi^2}\left(\partial_{\mathbf{x}}\Psi\right)^TBR^{-1}B^T\partial_{\mathbf{x}}\Psi+\frac{1}{2}\frac{\lambda}{\Psi^2}\left(\partial_{\mathbf{x}}\Psi\right)^T\Sigma\partial_{\mathbf{x}}\Psi.\end{aligned} \tag{20}$$

In the model (3)-(10), it can be seen that the optimal controls $\mathbf{u}(\mathbf{x})$ act as a correction term to the deterministic controls and the stochastic noise. Penalizing this control (12) suggests that the optimal control is that which is, in some sense, “close” to the passive, deterministic process (see [26] for a more precise definition in terms of Kullback-Leibler divergence). Moreover, this implies that the possibility of a large stochastic disturbance (either due to neighbors’ unknown controls or model kinematics) requires the possibility of a greater control input. Because of this, we assume that the noise in the controlled components is inversely proportional to the control penalty, or

$$\Sigma = \Gamma\Gamma^T = \lambda B R^{-1} B^T. \tag{21}$$

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

$$R = \lambda \; diag\left(\sigma_\theta^{-2}, \sigma_v^{-2}, \sigma_{\theta,1}^{-2}, \sigma_{v,1}^{-2}, \ldots, \sigma_{\theta,M}^{-2}, \sigma_{v,M}^{-2}\right). \tag{22}$$

From (22), the quadratic terms on the second line of (20) cancel, and the remaining PDE for $\Psi$ is linear:

$$0 = f^T \partial_{\mathbf{x}} \Psi(\mathbf{x}) + \frac{1}{2}\mathrm{Tr}\left(\Sigma \partial_{\mathbf{x}}^2 \Psi\right) - \frac{k(\mathbf{x})}{2\lambda}\Psi(\mathbf{x}) \tag{23}$$

$$\Psi(\mathbf{x}) = 1, \qquad \mathbf{x} \in \mathscr{F} \tag{24}$$

As before, this could be solved numerically until a steady state is reached. However, the Feynman-Kac equations [18, 30] connect certain linear differential operators to adjoint operators that describe the evolution of a *forward* diffusion process beginning from the current state $\widetilde{\mathbf{x}}(0) = \widetilde{\mathbf{x}}_0 = \mathbf{x}$. From the Feynman-Kac equations, the solution to (23) is [8]:

$$\Psi(\mathbf{x}) = \mathbb{E}_{\widetilde{\mathbf{x}},\tau|\widetilde{\mathbf{x}}_0}\left\{\Psi(\widetilde{\mathbf{x}}(\tau))\exp\left(-\frac{1}{2\lambda}\int_0^\tau k(\widetilde{\mathbf{x}}(s))ds\right)\right\}. \tag{25}$$

where $\widetilde{\mathbf{x}}(t)$ satisfies the path integral-associated, uncontrolled dynamics (cf. (11)),

$$d\widetilde{\mathbf{x}}(t) = f(\widetilde{\mathbf{x}}(t))dt + \Gamma d\mathbf{w}, \tag{26}$$

with initial condition $\widetilde{\mathbf{x}}(0) = \mathbf{x}$. The expectation in (25) is taken with respect to the joint distribution of $(\widetilde{\mathbf{x}}, \tau)$ of sample paths $\widetilde{\mathbf{x}} = \widetilde{\mathbf{x}}(t)$ that begin at $\widetilde{\mathbf{x}}_0 = \mathbf{x}$ and evolve as (26) until hitting the formation $\widetilde{\mathbf{x}}(\tau) \in \mathscr{F}$ at time $\tau$. Unlike previous PI works, where the terminal time is fixed and known in advance, this stopping time is a property of the set of stochastic trajectories $\widetilde{\mathbf{x}}(t)$.

The distribution $(\widetilde{\mathbf{x}}, \tau)$ is difficult to obtain. Monte Carlo techniques may be used to sample trajectories $\widetilde{\mathbf{x}}$, but hitting the formation is a rare event unless there is an artificial mechanism to "guide" the trajectory into the formation. In this work, we determine the trajectory $\widetilde{\mathbf{x}}|\tau, \mathbf{x}_0$ conditioned on its hitting time. From the law of total expectation,

$$\Psi(\mathbf{x}) = \mathbb{E}_{\tau|\mathbf{x}_0}\left\{\mathbb{E}_{\widetilde{\mathbf{x}}|\tau,\widetilde{\mathbf{x}}_0}\left[\Psi(\widetilde{\mathbf{x}}(\tau))\exp\left(-\frac{1}{2\lambda}\int\limits_0^{\tau} k(\widetilde{\mathbf{x}}(s))ds\right)\right]\right\} \tag{27}$$

$$= \mathbb{E}_{\tau|\mathbf{x}_0}\left\{\Psi(\mathbf{x}_0|\tau)]\right\}. \tag{28}$$

In practice, we find that the inner distribution $\Psi(\mathbf{x}_0|\tau)$ exhibits small tails for most $\tau$ and has high probability for just a small range of $\tau$. Moreover, the range of $\tau$ with higher likelihood $\Psi(\mathbf{x}_0|\tau)$ is that which appears to equally balance state and control costs. Therefore, we consider a discrete set $(\tau_1,\ldots,\tau_{N_\tau})$ of $N_\tau$ possible values for $\tau$ with non-informative prior probabilities. This implies that the distribution of the hitting times is implicitly encoded in the length (and the ensuing cost) of the path $\widetilde{\mathbf{x}}|\tau_i,\mathbf{x}_0$. Since $\Psi(\mathbf{x}(\tau_i)) = 1$ from (24), the solution (27) can be expanded as:

$$\Psi(\widetilde{\mathbf{x}}) = \frac{1}{N_\tau}\sum_{i=1}^{N_\tau}\mathbb{E}_{\widetilde{\mathbf{x}}|\widetilde{\mathbf{x}}_0,\tau_i}\left\{\Psi(\widetilde{\mathbf{x}}(\tau_i))\exp\left(-\frac{1}{2\lambda}\int\limits_0^{\tau_i} k(\widetilde{\mathbf{x}}(s))ds\right)\right\} \tag{29}$$

$$= \frac{1}{N_\tau}\sum_{i=1}^{N_\tau}\mathbb{E}_{\widetilde{\mathbf{x}}|\widetilde{\mathbf{x}}_0,\tau_i}\left\{\exp\left(-\frac{1}{2\lambda}\int\limits_0^{\tau_i} k(\widetilde{\mathbf{x}}(s))ds\right)\right\}. \tag{30}$$

By discretizing the interval $[0,\tau_i]$ into $N_i$ intervals of equal length $\Delta t$, $t_0 < t_1 < \ldots < t_{N_i} = \tau_i$, we can consider a sample of the discretized trajectory $\widetilde{\mathbf{x}}^N|\mathbf{x}_0,\tau_i = (\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_{N_i})$. Under this discretization in time, the solution (27) can be written as

$$\Psi(\widetilde{\mathbf{x}}) = \frac{1}{N_\tau}\lim_{\Delta t\to 0}\sum_{i=1}^{N_\tau}\int d\widetilde{\mathbf{x}}^N p(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)\exp\left[-\frac{\Delta t}{2\lambda}\sum_{k=1}^{N_i}k(\widetilde{\mathbf{x}}_k)\right], \tag{31}$$

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

$$p(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i) = \prod_{k=0}^{N_i-1} p(\widetilde{\mathbf{x}}_{k+1}|\widetilde{\mathbf{x}}_k,\tau_i). \tag{32}$$

Since the uncontrolled process (26) is driven by Gaussian noise with zero mean and covariance $\Sigma = \Gamma\Gamma^T$, the transition probabilities may be written as

$$\begin{aligned} p(\widetilde{\mathbf{x}}_{k+1}|\widetilde{\mathbf{x}}_k,\tau_i) \propto \exp\Big(&-\frac{1}{2}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k - f(\widetilde{\mathbf{x}}_k)\Delta t\right)^T \\ &\times\left(\Delta t\lambda BR^{-1}B^T\right)^{-1}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k - f(\widetilde{\mathbf{x}}_k)\Delta t\right)\Big) \end{aligned} \tag{33}$$

for $k < N_i - 1$, and $p(\widetilde{\mathbf{x}}_{k+1}|\widetilde{\mathbf{x}}_k,\tau_i) = \mathbb{1}_{h(\widetilde{\mathbf{x}}_{k+1})=\boldsymbol{\mu}}(\widetilde{\mathbf{x}}_{k+1})$ for $k = N_i - 1$.

The path integral representation of $\Psi(\widetilde{\mathbf{x}})$ is obtained from equations (31-33), and can be written as an exponential of an "action" [10] $S(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)$ along the time-discretized sample trajectories $(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N)$:

$$\Psi(\widetilde{\mathbf{x}}) \propto \lim_{\Delta t\to 0}\sum_{i=1}^{N\tau}\int d\widetilde{\mathbf{x}}^N \exp\left(-S(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)\right) \tag{34}$$

$$\begin{aligned} S(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0,\tau_i) &= \sum_{k=1}^{N_i}\frac{\Delta t}{2\lambda}k(\widetilde{\mathbf{x}}_k) + \sum_{k=0}^{N_i-1}\frac{1}{2}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k-\Delta t f(\widetilde{\mathbf{x}}_k)\right)^T \\ &\quad\times\left(\lambda\Delta t B R^{-1}B^T\right)^{-1}\left(\widetilde{\mathbf{x}}_{k+1}-\widetilde{\mathbf{x}}_k-\Delta t f(\widetilde{\mathbf{x}}_k)\right). \end{aligned} \tag{35}$$

The optimal control (18) is given by

$$\begin{aligned} \mathbf{u}(\widetilde{\mathbf{x}}) &= \lim_{\Delta t\to 0}\lambda R^{-1}B^T\partial_{\widetilde{\mathbf{x}}}\log\Psi \\ &= \lim_{\Delta t\to 0}\sum_{i=1}^{N_\tau}\int d\widetilde{\mathbf{x}}^N P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)\mathbf{u}_L(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i) \\ &= \lim_{\Delta t\to 0}\sum_{i=1}^{N_\tau}\mathbb{E}_{P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)}\left\{\mathbf{u}_L(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)\right\} \end{aligned} \tag{36}$$

where $\lim_{\Delta t\to 0}P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i) = P(\widetilde{\mathbf{x}}|\widetilde{\mathbf{x}}_0,\tau_i)$ is the probability of an *optimal* trajectory conditioned to hit the formation at time $\tau_i$:

$$P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i) \propto e^{-S(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)}, \tag{37}$$

and the local controls $\mathbf{u}_L(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)$ are

$$\mathbf{u}_L(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i) = \frac{\widetilde{\mathbf{x}}_1-\widetilde{\mathbf{x}}_0}{\Delta t} - f(\widetilde{\mathbf{x}}_0). \tag{38}$$

Although the resulting control law is stationary, the state space is too large for it to be computed offline. Because of this, after computing $\mathbf{u}(\mathbf{x}) = \mathbf{u}(\widetilde{\mathbf{x}}_0)$, each agent executes only the first increment of that control, at which point the optimal control is recomputed. Then (36) is

$$\begin{aligned} \mathbf{u}(\mathbf{x}) &= \frac{\mathbb{E}_{P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}-\mathbf{x}}{\Delta t} - f(\mathbf{x}) \\ &= \frac{\mathbb{E}_\tau\mathbb{E}_{P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau)}\{\widetilde{\mathbf{x}}_1\}-\mathbf{x}}{\Delta t} - f(\mathbf{x}). \end{aligned} \tag{39}$$

In other words, the control (39) applied by an agent in state $\mathbf{x}$ is constructed from a realization of the unknown or random dynamics of the system that maximizes the probability of the trajectory that starts from $\mathbf{x}$ and evolves until hitting the formation. This probability is weighted by the cost accumulated along the path. One may compute the optimal control (39) once the joint probability $P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0,\tau_i)$ 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 to compute the control in (39). Although Monte Carlo techniques can be used to generate samples of the maximally-likely trajectory $P(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_N|\widetilde{\mathbf{x}}_0)$, we find them to be slow in practice due to the high dimension of this problem ($\widetilde{\mathbf{x}}^N \in \mathbb{R}^{4NM}$). Moreover, when sampling a trajectory $\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}0,\tau_i$, the trajectory must be conditioned to hit the formation at $\tau_i$. Finally, is is not necessary to sample the entire distribution $P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0)$ since only the estimate $\widehat{\mathbf{x}}_1 \equiv \mathbb{E}_{P(\widetilde{\mathbf{x}}^N|\widetilde{\mathbf{x}}_0)}\{\widetilde{\mathbf{x}}_1\}$ is needed.

Therefore, in this work, we treat the temporal discretization of the optimal trajectory $\widetilde{\mathbf{x}}^N$ as the hidden state of a stochastic process, where appropriately-chosen measurements of this hidden state are related to the system goal $\boldsymbol{\mu}$ (15). The optimal control can then be computed from the optimal estimate $\widehat{\mathbf{x}}_1$ given the process and measurements over a fixed interval $t_1,\ldots,\tau_i$. We define the following nonlinear smoothing problem.

*Nonlinear Smoothing Problem*:

Given measurements $\mathbf{y}_k = \mathbf{y}(t_k)$ for $t_k = t_1,\ldots,t_N = \tau_i$, 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 hidden state-space model:

$$\widetilde{\mathbf{x}}_{k+1} = \widetilde{\mathbf{x}}_k + \Delta t f(\widetilde{\mathbf{x}}_k) + \varepsilon_k \tag{40}$$

$$\mathbf{y}_k = h(\widetilde{\mathbf{x}}_k) + \eta_k, \tag{41}$$

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

$$\mathbb{E}\left(\varepsilon_k \varepsilon_k^T\right) = \lambda \Delta t B R^{-1} B^T \tag{42}$$

$$\mathbb{E}\left(\eta_k \eta_k^T\right) = \begin{cases} \frac{\lambda}{\Delta t} Q^{-1} & k = 1,\ldots,N-1 \\ 0 & k = N \end{cases}. \tag{43}$$

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 $\mathbf{y}_k = \boldsymbol{\mu}$. □

To show the relation between the nonlinear smoothing problem and the stochastic optimal control problem, we write the probability of an estimated hidden state $\widehat{\mathbf{x}}_k$ in the filtering algorithm predication/update steps [9], which is proportional to the measurement likelihood and the predicted state:

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

where

$$
\begin{aligned}
p(\mathbf{y}_k|\widehat{\mathbf{x}}_k) \equiv p(\boldsymbol{\mu}_k|\widehat{\mathbf{x}}_k) &= N\left(h(\widehat{\mathbf{x}}_k), \eta_k \eta_k^T\right) \\
&\propto \exp\left\{ -\frac{\Delta t}{2\lambda} (h(\widehat{\mathbf{x}}_k) - \boldsymbol{\mu})^T Q (h(\widehat{\mathbf{x}}_k) - \boldsymbol{\mu}) \right\} \qquad (44) \\
p(\widehat{\mathbf{x}}_k|\widehat{\mathbf{x}}_{k-1}) &= N\left(\widehat{\mathbf{x}}_{k-1} + \Delta t f(\widehat{\mathbf{x}}_{k-1}), \Delta t \Sigma\right) \\
&\propto \exp\left\{ -\frac{1}{2} \left(\widehat{\mathbf{x}}_k - \widehat{\mathbf{x}}_{k-1} - \Delta t f(\widehat{\mathbf{x}}_{k-1})\right)^T \right. \\
&\qquad \left. \times \left(\lambda \Delta t B R^{-1} B^T\right)^{-1} \left(\widehat{\mathbf{x}}_k - \widehat{\mathbf{x}}_{k-1} - \Delta t f(\widehat{\mathbf{x}}_{k-1})\right) \right\}. \qquad (45)
\end{aligned}
$$

Comparing the right hand sides of (44-45) with (35), it can be seen that the increments of the nonlinear filtering problem are equivalent to those in the stochastic optimal control problem.

Since the optimal control (39) is based on the probability of a full trajectory of fixed length and values $\boldsymbol{\mu}$ are available in advance, the expected value of the trajectory originating from state $\widehat{\mathbf{x}}_0$ conditioned to hit the formation at time $\tau_i$, 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 values $\boldsymbol{\mu}_k$ using a nonlinear Kalman smoother. Such an algorithm assumes that the increments given by (44) and (45) 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.

Once this estimated trajectory has been computed for each $\tau_i$, the expectation over $\tau_i$ may be computed using (35) and (37). This results in an average of the controls $u_L(\widetilde{\mathbf{x}}|\widetilde{\mathbf{x}}_0, \tau_i)$ to be applied, weighted by the the probability of the optimal trajectory for each $\tau_i$. In other words, each agent is estimating both the optimal system trajectory (from its perspective) given the time the formation will hit *and* estimating the hitting time of the formation. Hitting the target sooner would save on state costs, but may cause an increase in control costs, and vice versa.

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 $[t_0, \tau_i]$. The computational complexity of the smoother used in this work is analyzed in [23].

## 5 Results

Next, these methods are employed so that five agents achieve the formation of a regular pentagon, where each agent is individually estimating the hidden optimal trajectory based on the relative kinematics of all of its neighbors. The agents observe all others, but, as described in Section 2, only the inter-agent connections *known* an agent are used when computing the control. The instantaneous state cost (15) penalizes the mean squared distance from the unicycle to all of its $M = 4$ neighbors

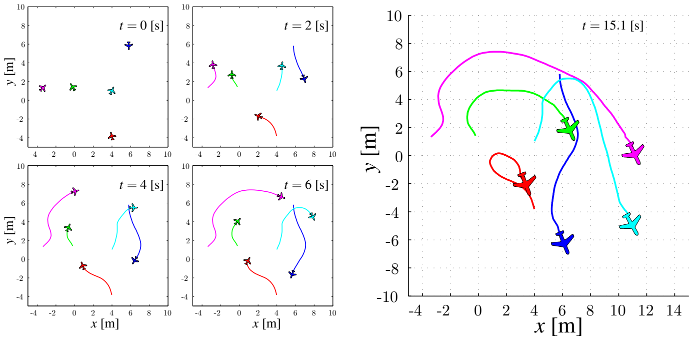

**Fig. 2** Five agents, starting from random initial positions and a common speed $v = 2.5$ [m/s], achieve a regular pentagon formation by an individually-optimal choice of acceleration and turning rate, without any active communication. The frame at 4 [s] shows an example of collision avoidance between the two upper-left agents.

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 $\delta_m$, $m = 1, \ldots, 4$.

System and methodological parameters were chosen as $\lambda = 100$, $\sigma_\theta = 0.1, \sigma_v = 0.05, \sigma_{\theta,m} = 1, \sigma_{v,m} = 1, N_\tau = 10$, $\tau \in (1, \ldots, 10)$, $Q = 100I$, $\epsilon_r = \epsilon_v = 0.1$, $\epsilon_\theta = 10^\circ$, and $\Delta t = 0.1$ s. The control was computed from the result of a Discrete-time Unscented Kalman Rauch-Tung-Striebel Smoother [23]. Fig. 2 shows the trajectories of all agents, while the the inter-agent distances and agents' angles and speeds can be seen in Fig. 3. The actual stopping time was $\tau = 15.1$ [s], and the agents do not collide. Without the addition of the optimal controls, the agents form a loose pentagon, but the collision-avoiding controls acting alone led to oscillatory trajectories, and the formation tolerances (2) were not reached in the first 100 [s] of simulation.

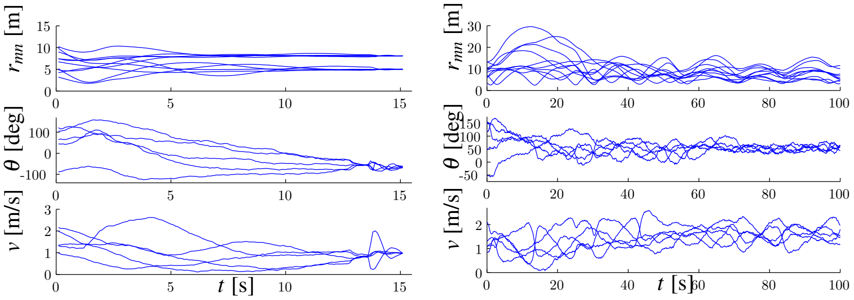

**Fig. 3** Inter-agent distances $r_{mn}$, agent heading angles $\theta$, and agent speeds $v$ as a function of time using the stochastic optimal control (left) and the deterministic feedback control (right).

## 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 huge state space, we exploit the stochasticity inherent in distributed multi-agent control problems in order to apply path integral method.

Each agent computes its optimal control using a nonlinear Kalman smoothing algorithm. The measurement noise and process noise of the dual smoothing problem are created using the structure of the cost function and stochastic kinematics. Aside from instantaneous observations of neighbors, the formation is created and maintained without any communication among agents.

In order to prevent collisions among agents, the optimal turning rate and acceleration controls affect the system alongside a non-optimal feedback control law based on an artificial potential. This suggests that the type of stochastic optimal control problem considered in this work may provide a way to analyze or optimize other deterministic feedback control laws used in swarm robotics.

**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: V. Blondel, S. Boyd, H. Kimura (eds.) Recent Advances in Learning and Control, pp. 15–34. Springer-Verlag (2008)
2. 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)
3. van den Broek, B., Wiegerinck, W., Kappen, B.: Optimal control in large stochastic multi-agent systems. Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning pp. 15–26 (2008)
4. Bullo, F., Cortes, J., Martinez, S.: Distributed control of robotic networks: A mathematical approach to motion coordination algorithms. Princeton University Press, Princeton, NJ (2009)
5. Dimarogonas, D.: On the rendezvous problem for multiple nonholonomic agents. IEEE Transactions on Automatic Control **52**(5), 916–922 (2007)
6. Elkaim, G., Kelbley, R.: A Lightweight Formation Control Methodology for a Swarm of Non-Holonomic Vehicles. In: IEEE Aerospace Conference. IEEE, Big Sky, MT (2006)
7. Fleming, W., Soner, H.: Logarithmic Transformations and Risk Sensitivity. In: Controlled Markov Processes and Viscosity Solutions, chap. 6. Springer, Berlin (1993)
8. Freidlin, M.: Functional Integration and Partial Differential Equations. Princeton University Press, Princeton, NJ (1985)
9. Gelb, A.: Applied Optimal Estimation. The M.I.T. Press, Cambridge, MA (1974)
10. Goldstein, H.: Classical Mechanics, 2nd edn. Addison-Wesley (1980)
11. Jadbabaie, A., Hauser, J.: On the stability of unconstrained receding horizon control with a general terminal cost. In: Proceedings of the 40th IEEE Conference on Decision and Control, vol. 5, pp. 4826–4831. IEEE, Orlando, FL (2001)
12. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, 3rd edn. North Holland (2007)

13. Kappen, H.: Linear Theory for Control of Nonlinear Stochastic Systems. Physical Review Letters **95**(20), 1–4 (2005)
14. Kappen, H.J.: Path integrals and symmetry breaking for optimal control theory. Journal of Statistical Mechanics, Theory and Experiment **2005**, 21 (2005)
15. Kappen, H.J., Gómez, V., Opper, M.: Optimal control as a graphical model inference problem. Machine Learning **87**(2), 159–182 (2012)
16. Kushner, H.J., Dupuis, P.: Numerical Methods for Stochastic Control Problems in Continuous Time, 2nd edn. Springer (2001)
17. Milutinović, D.: Utilizing Stochastic Processes for Computing Distributions of Large-Size Robot Population Optimal Centralized Control. In: Proceedings of the 10th International Symposium on Distributed Autonomous Robotic Systems. Lausanne, Switzerland (2010)
18. Oksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 6th edn. Springer-Verlag, Berlin (2003)
19. 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)
20. Parker, L.E.: Multiple Mobile Robot Systems. In: B. Sciliano, O. Khatib (eds.) Springer Handbook of Robotics, chap. 40, pp. 921–941. Springer (2008)
21. Ren, W., Beard, R.: Distributed consensus in multi-vehicle cooperative control: Theory and applications. Springer Verlag, New York, NY (2007)
22. Ryan, A., Zennaro, M., Howell, A., Sengupta, R., Hedrick, J.: An overview of emerging results in cooperative UAV control. 2004 43rd IEEE Conference on Decision and Control pp. 602–607 Vol.1 (2004)
23. Särkkä, S.: Continuous-time and continuous-discrete-time unscented Rauch-Tung-Striebel smoothers. Signal Processing **90**(1), 225–235 (2010)
24. Tanner, H., Jadbabaie, A., Pappas, G.: Coordination of multiple autonomous vehicles. In: IEEE Mediterranean Conference on Control and Automation. IEEE, Rhodes, Greece (2003)
25. Todorov, E.: General duality between optimal control and estimation. In: 47th IEEE Conference on Decision and Control, 5, pp. 4286–4292. IEEE, Cancun, Mexico (2008)
26. Todorov, E.: Efficient computation of optimal actions. Proceedings of the National Academy of Sciences of the United States of America **106**(28), 11,478–83 (2009)
27. Wang, M.C., Uhlenbeck, G.: On the theory of Brownian Motion II. Reviews of Modern Physics **17**(2-3), 323–342 (1945)
28. Wiegerinck, W., van den Broek, B., Kappen, B.: Stochastic optimal control in continuous space-time multi-agent systems. In: Proceedings UAI. Citeseer (2006)
29. Wiegerinck, W., van den Broek, B., Kappen, B.: Optimal on-line scheduling in stochastic multiagent systems in continuous space-time. Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems p. 1 (2007)
30. Yong, J.: Relations amng ODEs, PDEs, FSDEs, BDSEs, and FBSDEs. In: Proceedings of the 36th IEEE Conference on Decision and Control, December, pp. 2779–2784. IEEE, San Diego, CA (1997)