Spatial Distribution Statistics for Two-Agent Optimal Navigation with Cone-Shaped Local Observation

Jan De Mot and Eric Feron
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
{jdemot,feron}@mit.edu

Abstract— **In this paper, we study spatially synchronous two-agent navigation on a structured partially unknown graph. The general edge cost statistics are given, and the agents gather and share exact information on the cost of local edges. The agents purpose is to traverse the graph as efficiently as possible. In previous work, we formulate the problem as a Dynamic Program, and exploit the structure of an equivalent Linear Program to compute the optimal value function. Here, we use the optimal policy to formulate a Markov chain with an infinite number of states whose properties we analyze. We present a method that computes the steady state probability distribution of the agent separation, exploiting the repetitive structure of the Markov chain as the agent separation goes to infinity. The results confirms and quantify the intuition that the less rewards, the more beneficial for the agents to spread out.**

I. Introduction

In recent years, multi-agent navigation problems have become of major interest to many researchers. The main idea is that a group of multiple, possibly cheap and diverse agents execute tasks faster, more reliably, and more efficiently than a single agent. Applications range from coordinated terrain exploration, search and rescue operations, to coordinated thermal search and deceptive reconnaissance missions.

In general, one of the principal challenges multi-agent systems pose is the computational complexity of designing optimal agent controls. Therefore, many researchers focus on approximate sub-optimal strategy design, which is computationally simpler. In [8], a group of agents exploits the energy efficiency of flying in formation to coordinate the traversal of a set of agents as energy efficiently as possible. The two-agent case is solved exactly; however, for larger problems, approximate solutions were needed. Other examples include casting of the multi-agent problem in a the pursuit-evader framework in [2], where an agent clusters pursues a single evader. Approximate solutions are presented. Further, decentralization is often the adequate modeling tool suggested by nature, for example the swarms in [4], [7] where local single-agent navigation rules adequately model agent flocking. However, in many practical problems, a centralized decision maker is available and there is a need to rigorously study the potential benefits cooperation provides.

In this paper, we consider the problem of optimal two-agent graph traversal, where to each edge is associated a cost. *A priori*, only the edge cost statistics are given, but on a particular set of edges around the current agent position (the *local observation zone*), the exact edge cost is observed and shared. We exploit a trade-off between two competing tendencies. First, there is the tendency for the agents (A and B) to spread, increasing the size of the union of the local observation zones ($\mathcal{O} = \mathcal{O}^A \cup \mathcal{O}^B$). Indeed, agents close to each other have overlapping observation zones and not as many edge costs are observed. More information yields enhanced efficiency, hence the spreading tendency. On the other hand, the agents tend to converge so that agent A's reachable set of edges ($\mathcal{R}^A$) intersects with $\mathcal{O}^B$ and vice versa. Only then can each agent take advantage of potentially cheap edges the other agent observes and exploit the benefit multiple agents have over single agents.

In past work [5], [6], we present an algorithm to compute optimal two-agent policies. In particular, the algorithm computes the optimal value function in two steps. First, we solve a small linear program which yields the optimal value function for small agent separations, the zone where efficient cooperation takes place. Then, we simulate an autonomous linear time invariant system, which provides the optimal value function for large agent separations. The method is briefly reviewed in this paper.

The contribution of this paper is the study of the steady state characteristics of the underlying Markov Decision Process (MDP) under an optimal policy. The MDP has infinitely many states, but its structure can be exploited to obtain the exact steady state probability distribution of the agent separation. The method presented here is somewhat similar to traditional methods to derive properties of infinite state space Markov chains in queueing systems [3].

The paper is organized as follows. In section II we introduce the notation and formulate the problem. In section III, we present a mathematical model and briefly summarize the algorithm, developed in previous work, that computes the optimal value function. In section IV, we formulate the MDP and present a method to compute the steady state probability distribution of the agent separation, illustrated and discussed in section V. Finally, section VI concludes.

II. Notation and Problem Formulation

In this section, we introduce the notation and formulate the two-agent navigation problem as a graph traversal

0-7803-9098-9/05/$25.00 ©2005 AACC

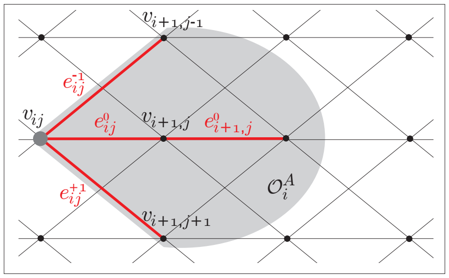

Fig. 1. Graph structure and notation. The gray area represents the local observation zone of an agent located at vertex v_{ij}; formally, $\mathcal{O}_i$ is the set of colored edges.

problem with partial edge cost information.

We grid the navigation terrain into sectors and associate a vertex $v \in V$ to each sector. Edges $e \in E$ connect pairs of vertices which reduces the navigation problem into a graph traversal problem on the graph $G(V, E)$. The structured transition graph we consider consists of an infinite number of vertical lines of vertices, having an infinite number of vertices each. A horizontal vertex array is referred to as a *lane*, while a *stage* refers to a vertical line of vertices. For the graph structure and related notation, see Fig 1. Note that at each vertex, three edges with associated superscript $k \in \{-1, 0, 1\}$ lead to the next stage.

To each edge e_{ij}^k is associated a cost c_{ij}^k. In this paper, we assume that the edge costs are independent identically distributed (i.i.d.) random variables with values picked from a finite set $\mathcal{L} = \{0, 1\}$. We denote p as the probability of encountering a zero edge cost. *A priori*, only the edge cost statistics are available and edge costs are time invariant.

The graph traversal problem is modeled as a discrete time decision problem where at time zero agents A and B are positioned at the vertices on lanes l_0^A and l_0^B ($\in \mathbb{N}$ and with $l_0^A \leq l_0^B$) at stage zero, indicated by the subscript. At each time step, each agent chooses to traverse one of the three available links to the next stage. Therefore, we let the time index coincide with the stage index and at time i the agents reach stage i at lanes l_i^A and l_i^B (where the agent are (re)labeled such that $l_i^A \leq l_i^B$, $i \geq 0$). Since at each time step agents A and B are positioned at the same stage, we refer to this as *spatially synchronous* motion.

At each stage i, we associate a local observation zone to both agents, namely $\mathcal{O}_i^A$ and $\mathcal{O}_i^B$, a set of edges of which the cost is observed and known. In this paper, we define the local observation zone for agent A positioned at vertex l_i^A, as (see Fig. 1)

$$\mathcal{O}_i^A = \{e_{ij}^{-1}, e_{ij}^{0}, e_{ij}^{1}, e_{i+1,j}^{0}\}.$$

Local observation zone $\mathcal{O}_i^B$ is defined similarly. Upon arrival at a vertex, each agent observes the costs of until then unobserved edges and communicates these to the other agent. At stage i the set of agents incurs the costs associated with the edges traversed to reach stage $i+1$. We formulate the following main problem.

Problem 1: **[Main]** Let a set of two agents be located at positions l_0^A and l_0^B of graph $G(V, E)$. The edge costs are time invariant and i.i.d. random variables over $\mathcal{L}$ with probability p for a zero edge cost. The agents navigate spatially synchronously through the graph in the direction of increasing stage indices, infinitely long. Furthermore, the agents share the costs of the edges in their respective local observation zones perfectly and instantaneously upon reaching a vertex.

Then, find the navigation strategy for both agents so that the expected discounted sum of costs incurred at each stage is minimized. Here, the expected value is taken over all initially unobserved edge costs; costs incurred at stage i are discounted by factor α^i, where $0 \leq \alpha < 1$. ■

III. Mathematical Problem Formulation and Solution Algorithm

In Section III-A, we present a *Dynamic Programming* (DP) formulation of the graph traversal problem. In Section III-B, we give a brief overview of the method to compute the optimal value function.

A. DP Formulation

We cast the two-agent navigation problem as a discounted cost, infinite horizon DP problem as follows [1]. Since graph G exhibits spatial invariance properties in the horizontal and vertical directions, and the agents are constrained to advance synchronously, we choose the system state $x \in \mathcal{S}$, to be independent of the stage and the absolute positions of the agent pair, dropping indices i and j. Let $\mathcal{C}^A$ and $\mathcal{C}^B$ denote the vectors with as entries the costs of the edges in $\mathcal{O}^A$ and $\mathcal{O}^B$, respectively, at the current position of A and B. Then, we define the system state $x = (s, \mathcal{C}^A, \mathcal{C}^B)$, where $s = l^B - l^A \in \mathbb{N}^+$ is the agent *separation*. Let $\mathcal{S}(s) \subset \mathcal{S}$ denote the set of states associated with separation s. Let $\mathbf{u} = (u^A, u^B) \in U = \{-1, 0, 1\}^2$ denote the decision vector where $u^{(\cdot)} \in \{-1, 0, 1\}$, representing the three possible decisions available to each agent. Given x_i and $\mathbf{u}_i$ at time step i, the agent cluster moves into the new state

$$x_{i+1} = f(x_i, \mathbf{u}_i), \tag{1}$$

where $f : \mathcal{S} \times U \rightarrow \mathcal{S}$ is the state transition function. The cost incurred in the transition from state x_i to x_{i+1} is the sum of the edge costs of the edges the agents traverse. Let policy $\mu : \mathcal{S} \rightarrow U$ be a particular two-agent policy. Then, the expected discounted cost $J_\mu(x_0)$ the agents incur in advancing for an infinite number of time steps, given initial state x_0 and under policy μ is

$$J_\mu(x_0) = \lim_{N \to \infty} E\left[\sum_{i=0}^{N} \alpha^i g(x_i, \mu(x_i))\right],$$

subject to the system equation (1), and where $0 \leq \alpha < 1$ is the discount factor. With the principle of optimality, the optimal value function $J^*(x)$ solves Bellman's equation,

$$J^*(x) = \min_{\mathbf{u} \in U} E\left[g(x,\mathbf{u}) + \alpha J^*(f(x,\mathbf{u}))\right]. \tag{2}$$

A policy that minimizes the right hand side of Eq. (2) is referred to as an optimal policy μ^*.

B. Solution Algorithm: Review

Let $\mu_\infty : \mathcal{S}(s) \rightarrow U$ denote the policy where at separation s each agent follows the single agent optimal policy and where ties are broken by choosing the pair of decisions that minimizes the resulting agent separation. Then, we look for an optimal two agent policy μ^* that is such that for any particular p and α ($0 \leq p \leq 1, 0 \leq \alpha < 1$), there is an $\bar{s} \geq 0$ such that for all $s \geq \bar{s}$, $\mu^*(x) = \mu_\infty(x)$, for $x \in \mathcal{S}(s)$.

This allows for the formulation of the following LTI system:

$$\mathbf{X}^{s+1,*} = \mathbf{A}\mathbf{X}^{s,*} + \mathbf{B}, \tag{3}$$

where the entries of $\mathbf{X}^{s,*}$ are functions of the optimal value function at separations smaller than or equal to separation s. The simulation of system (3) allows to compute the optimal value function $J^*(x)$ at the separations $s = \bar{s}, \bar{s}+1, \ldots$.

Since the optimal value function is upper bounded by twice the (bounded) single agent optimal value function, the initial conditions of LTI-system (3) are required to only excite the stable system modes. This allows for the formulation of a well-defined LP, referred to as $\mathcal{LP}_{in}$ of which the solution is the optimal value function for separations $s < \bar{s}$. For a detailed treatment of this algorithm, we refer the reader to [6]. The paper can be found on `http://web.mit.edu/jdemot/Public/allerton`.

IV. Analysis of the Steady State Optimal Agent Behavior

In this section, we study the steady state agent behavior under an optimal policy. In particular, we present a method to compute the probability distribution of the agent separation after initial transition effects have died out.

With the optimal policy μ^*, the state evolution can be described by means of an infinite state Markov chain. Specifically, let $\delta^s_{qr} \in \mathcal{D}$ ($s \geq 0$, $q, r \in \mathcal{L}$) denote the state whereby the agent separation is s, and whereby the observation zone pair is any pair for which $c^0_A = q$ and $c^0_B = r$, representing the costs of first edge straight ahead of A and B, respectively. Other local observation zone edge costs are independent of $\mathbf{u}$, and hence states are lumped together as described (see [6], for details). Set $\mathcal{D}$ is the set of Markov chain states. For the remainder of this section, we denote by *state* the Markov chain state δ^s_{qr}. Fig. 2 shows the Markov chain structure conceptually, only depicting outgoing arcs for the three states associated with separation $s \geq 2$. This element is repeated for all separations $s \geq 2$ with natural extensions at separations $s = 0$ and $s = 1$.

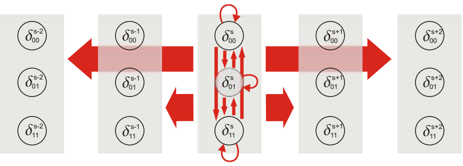

Fig. 2. Element of the two-agent Markov chain structure.

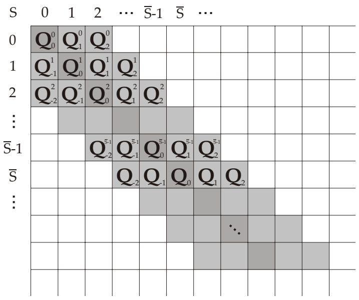
S
0
1
2
⋯
S̄-1
S̄
⋯
$\mathbf{Q}^0_0$ $\mathbf{Q}^0_1$ $\mathbf{Q}^0_2$
$\mathbf{Q}^1_{-1}$ $\mathbf{Q}^1_0$ $\mathbf{Q}^1_1$ $\mathbf{Q}^1_2$
$\mathbf{Q}^2_{-2}$ $\mathbf{Q}^2_{-1}$ $\mathbf{Q}^2_0$ $\mathbf{Q}^2_1$ $\mathbf{Q}^2_2$
$\mathbf{Q}^{\bar{S}-1}_{-2}$ $\mathbf{Q}^{\bar{S}-1}_{-1}$ $\mathbf{Q}^{\bar{S}-1}_0$ $\mathbf{Q}^{\bar{S}-1}_1$ $\mathbf{Q}^{\bar{S}-1}_2$
$\mathbf{Q}_{-2}$ $\mathbf{Q}_{-1}$ $\mathbf{Q}_0$ $\mathbf{Q}_1$ $\mathbf{Q}_2$

Fig. 3. Conceptual representation of the block multi-diagonal structure of Markov chain transition matrix $\mathbf{Q}$.

Let $p^k(\delta)$ denote the probability for being at state δ after k state transitions. Let

$$\pi_k = \left[\begin{array}{cccc} (\mathbf{p}^k_0)^T & (\mathbf{p}^k_1)^T & (\mathbf{p}^k_3)^T & \ldots \end{array}\right]^T,$$

where $\mathbf{p}^k_s = \left[\begin{array}{ccc} p^k(\delta^s_{00}) & p^k(\delta^s_{01}) & p^k(\delta^s_{11}) \end{array}\right]^T$. In words, π_k contains the state probability distribution after k state transitions. We define $q(\delta'|\delta)$ as the probability the next state is δ', given the current state δ. Given the edge cost statistics and μ^*, we can compute $q(\delta'|\delta)$ for all $\delta, \delta' \in \mathcal{D}$. Let matrix $\mathbf{Q}$ denote the state transition matrix. Since from separation s, only the separations $s+\sigma$ (for $\sigma = -2, \ldots, 2$) can be reached, the matrix $\mathbf{Q}$ has a block multi-diagonal structure (see Fig. 3). In $\mathbf{Q}$, the entry on the row and column corresponding to δ' and δ, respectively, contains $q(\delta'|\delta)$. Sub-matrix $\mathbf{Q}^s_\sigma \in \mathbb{R}^{3\times 3}$ ($\sigma = \max\{-2, -s\}, \ldots, 2$) describes the transitions from the states associated with separation s to the states associated with separation $s+\sigma$.

We have that policy μ_∞ is optimal for all $s \geq \bar{s}$. Therefore, the matrix $\mathbf{Q}$ has a recurring structure for all $s \geq \bar{s}$. In particular, let $\mathbf{Q}_s$ contain the rows of $\mathbf{Q}$ corresponding to separation s, for $s \geq \bar{s}$. Then,

$$\begin{aligned}\mathbf{Q}_s = \big[\, \mathbf{0} \ \cdots \ \mathbf{0} \ \ \mathbf{Q}_{-2} \ \ \mathbf{Q}_{-1} \ \ \mathbf{Q}_0 \\ \mathbf{Q}_1 \ \ \mathbf{Q}_2 \ \ \mathbf{0} \ \ \cdots \,\big]^T,\end{aligned}$$

where the zero-matrices are of the appropriate dimensions. Fig. 3 shows the recurring structure of $\mathbf{Q}$ for $s \geq \bar{s}$.

We wish to compute

$$\pi^* = \lim_{k\rightarrow\infty} \pi^k,$$

the steady state probability distribution. The vector π^* is the eigenvector corresponding to the unique unit eigenvalue of $\mathbf{Q}$. We exploit the structure of $\mathbf{Q}$ to obtain π^*, using a principle similar to the one underlying the algorithm to compute the optimal value function presented earlier. Specifically, we formulate an autonomous LTI system whose state trajectories are the steady state probabilities for $s \geq \bar{s}$, where the separation s plays the role of "time" in a classical LTI system. We then formulate a set of linear equalities whose solution are the steady state probabilities for $s < \bar{s}$. The set includes linear equalities that ensure the computation of initial conditions so that only the stable LTI modes are excited. Lastly, the condition that the probabilities are required to sum up to one ensures the set of equalities yields a unique solution. The latter condition is equivalent to the maximization in $\mathcal{LP}_{in}$.

We have that

$$\mathbf{Q}\pi^* = \pi^*. \tag{4}$$

Therefore, for $s \geq \bar{s}$, we have that

$$\mathbf{p}_s^* = \sum_{\sigma=-2}^{2} \mathbf{Q}_\sigma \mathbf{p}_{s+\sigma}^*, \tag{5}$$

where $\mathbf{p}_s^*$ is the steady state version of $\mathbf{p}_s^k$. We transform this equation into a forward recursive equation. In particular, we have that $\text{rank}(\mathbf{Q}_2) = 1$. Therefore,

$$\mathbf{Q}_2 = \mathbf{q}_2(\mathbf{r}_1)^T, \tag{6}$$

where $\mathbf{q}_2, \mathbf{r}_1 \in \mathbb{R}^{3\times 1}$ are non-zero vectors. Since $\text{rank}(\mathbf{Q}_1) = 2$ and

$$\text{rank}\left(\left[\begin{array}{cc} \mathbf{Q}_1 & \mathbf{Q}_2 \end{array}\right]\right) = 2,$$

we have that

$$\mathbf{Q}_1 = \mathbf{q}_{11}(\mathbf{r}_1)^T + \mathbf{q}_{12}(\mathbf{r}_2)^T, \tag{7}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2 \in \mathbb{R}^{3\times 1}$ are linearly independent and $\mathbf{q}_{11}, \mathbf{q}_{12} \in \mathbb{R}^{3\times 1}$ are non-zero vectors. Further, $\text{rank}(\mathbf{I} - \mathbf{Q}_0) = 3$. Hence, we have that

$$\mathbf{I} - \mathbf{Q}_0 = \mathbf{q}_{01}(\mathbf{r}_1)^T + \mathbf{q}_{02}(\mathbf{r}_2)^T + \mathbf{q}_{03}(\mathbf{r}_3)^T,$$

where $\mathbf{r}_3 \in \mathbb{R}^{3\times 1}$ is linearly independent of the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, and where $\mathbf{q}_{01}, \mathbf{q}_{02}, \mathbf{q}_{03} \in \mathbb{R}^{3\times 1}$ are non-zero vectors. Let $\rho_s^{i,*} = (\mathbf{r}_i)^T \mathbf{p}_s^*$, for $i = 1, 2, 3$. Furthermore, let

$$\rho_s^* = \mathbf{R}_\rho \mathbf{p}_s^*, \tag{8}$$

where

$$\rho_s^* = \left[\begin{array}{ccc} \rho_s^{1,*} & \rho_s^{2,*} & \rho_s^{3,*} \end{array}\right]^T.$$

Then, we can write Eq. (5) as

$$\begin{aligned}\left[\mathbf{q}_{01} \;\; \mathbf{q}_{02} \;\; \mathbf{q}_{03}\right]\rho_s^* \;\; = \;\; & \mathbf{Q}_{-2}\mathbf{R}_\rho\rho_{s-2}^* + \mathbf{Q}_{-1}\mathbf{R}_\rho\rho_{s-1}^* + \\ & \left[\mathbf{q}_{11} \;\; \mathbf{q}_{12}\right]\left[\begin{array}{c} \rho_{s+1}^{1,*} \\ \rho_{s+1}^{2,*} \end{array}\right] + \mathbf{q}_2\rho_{s+2}^{1,*},\end{aligned} \tag{9}$$

It can be verified that $\mathbf{q}_2$, $\mathbf{q}_{12}$ and $\mathbf{q}_{03}$ are linearly independent. Therefore, we can write recursion (9) as

$$\mathbf{Q}_2'\rho_{s+2}'^* = \sum_{\sigma\in\{1,0,-1,-2\}} \mathbf{Q}_\sigma'\rho_{s+\sigma}'^*, \tag{10}$$

where

$$\mathbf{Q}_2' = \left[\begin{array}{ccc} -\mathbf{q}_2 & -\mathbf{q}_{12} & \mathbf{q}_{03} \end{array}\right]$$

is invertible, where

$$\rho_s'^* = \left[\begin{array}{ccc} \rho_s^{1,*} & \rho_{s-1}^{2,*} & \rho_{s-2}^{3,*} \end{array}\right]^T,$$

and where $\mathbf{Q}_\sigma'$ (for $\sigma \in \{1, 0, -1, -2\}$) can be determined from Eq. (9). Note that the last column of $\mathbf{Q}_{-1}'$ and the last two columns of $\mathbf{Q}_{-2}'$ have zero entries. Eq. (10) is a forward recursion which we convert in the autonomous LTI system

$$\left[\begin{array}{c} \rho_{s+2}'^* \\ \rho_{s+1}'^* \\ \rho_s'^* \\ \rho_{s-1}'^* \end{array}\right] = \left[\begin{array}{cccc} \mathbf{Q}_1'' & \mathbf{Q}_0'' & \mathbf{Q}_{-1}'' & \mathbf{Q}_{-2}'' \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \rho_{s+1}'^* \\ \rho_s'^* \\ \rho_{s-1}'^* \\ \rho_{s-2}'^* \end{array}\right], \tag{11}$$

where $\mathbf{Q}_\sigma'' = (\mathbf{Q}_2')^{-1}\mathbf{Q}_\sigma'$, for $\sigma = 1, 0, -1, -2$. We write system equation (11) in short as $\bar{\rho}_{s+2}^* = \mathbf{A}_\rho\bar{\rho}_{s+1}^*$. The subscript s of $\bar{\rho}_s^*$ indicates the largest separation s at which some linear combination of the steady state probability distribution is known, given $\bar{\rho}_{s'}^*$ for $s' \leq s$.

We now use the LTI system to compute

$$\Sigma_{\bar{s}}^* = \sum_{s=\bar{s}}^{\infty} \left(p^*(\delta_{00}^s) + p^*(\delta_{01}^s) + p^*(\delta_{11}^s)\right), \tag{12}$$

where $p^*(\delta)$ is the steady state version of $p^k(\delta)$. It can be verified that $\mathbf{A}_\rho$ is diagonalizable. In particular, we have that $\mathbf{V}_\rho^{-1}\mathbf{A}_\rho\mathbf{V}_\rho = \Lambda_\rho$, where ith column of $\mathbf{V}_\rho$ is the ith eigenvector of $\mathbf{A}_\rho$ and where

$$\Lambda_\rho = \text{diagonal}(\lambda_1, \ldots, \lambda_{12})$$

with λ_i the ith eigenvalue of $\mathbf{A}_\rho$ (for $i = 1, \ldots, 12$). The system has three stable modes and nine unstable modes. We divide $\mathbf{V}_\rho$, $\mathbf{V}_\rho^{-1}$ and Λ_ρ accordingly into submatrices. Specifically,

$$\mathbf{V}_\rho = \left[\begin{array}{cc} \mathbf{V}_{\rho 1} & \mathbf{V}_{\rho 2} \end{array}\right],$$

where the columns of $\mathbf{V}_{\rho 1}$ and $\mathbf{V}_{\rho 2}$ are the stable and unstable mode eigenvectors, respectively; similarly,

$$\Lambda_\rho = \text{diagonal}(\Lambda_{\rho 1}, \Lambda_{\rho 2})$$

and

$$\mathbf{V}_\rho^{-1} = \left[\begin{array}{cc} \bar{\mathbf{V}}_{\rho 1}^T & \bar{\mathbf{V}}_{\rho 2}^T \end{array}\right]^T.$$

Let

$$\bar{\rho}_s'^* = \mathbf{V}_\rho^{-1}\bar{\rho}_s^*. \tag{13}$$

Then, the diagonalized system equation becomes $\bar{\rho}_{s+2}'^* = \Lambda_\rho\bar{\rho}_{s+1}'^*$. We have that

$$\Sigma_{\bar{s}}^* = \left\|\sum_{k=0}^{\infty} \mathbf{p}_{\bar{s}+k}^*\right\|,$$

where $\|(\cdot)\|$ denotes the sum of the rows of $(\cdot)$. Let $\mathbf{T}_3 \in \mathbb{R}^{3\times 12}$ be such that $\rho_s^* = \mathbf{T}_3\bar{\rho}_s^*$. Then, with Eqs (8) and (13), we have that

$$\begin{aligned}\sum_{k=0}^{\infty}\mathbf{p}_{\bar{s}+k}^* &= \mathbf{R}_\rho^{-1}\sum_{k=0}^{\infty}\rho_{\bar{s}+k}^*,\\ &= \mathbf{R}_\rho^{-1}\sum_{k=2}^{\infty}\mathbf{T}_3\bar{\rho}_{\bar{s}+k}^*,\\ &= \mathbf{R}_\rho^{-1}\mathbf{T}_3\sum_{k=2}^{\infty}\mathbf{V}_{\rho 1}\bar{\rho}_{\bar{s}+k,1}^{\prime *},\\ &= \mathbf{R}_\rho^{-1}\mathbf{T}_3\mathbf{V}_{\rho 1}(\mathbf{I}-\Lambda_{\rho 1})^{-1}\bar{\mathbf{V}}_{\rho 1}\bar{\rho}_{\bar{s}+2}^*,\end{aligned}$$

where $\bar{\rho}_{\bar{s}+k,1}^{\prime *}$ denotes the part of $\bar{\rho}_{\bar{s}+k}^{\prime *}$ associated with the stable system modes, and where we use the fact that the unstable system modes are not excited, since otherwise $\mathbf{p}_s^*$ diverges which is impossible. For

$$\mathbf{W} = \|\mathbf{R}_\rho^{-1}\mathbf{T}_3\mathbf{V}_{\rho 1}(\mathbf{I}-\Lambda_{\rho 1})^{-1}\bar{\mathbf{V}}_{\rho 1}\|,$$

we have that

$$\Sigma_{\bar{s}}^* = \mathbf{W}\bar{\rho}_{\bar{s}+2}^*. \tag{14}$$

In words, Eq. (14) provides the sum of the steady state probabilities for $s \geq \bar{s}$ as a linear equation in $\bar{\rho}_{\bar{s}+2}^*$.

We now formulate a set of linear equalities with $\mathbf{p}_s^*$ for $s \leq \bar{s}+1$ as solution. In particular, let

$$\mathbf{p}^* = \begin{bmatrix} (\mathbf{p}_0^*)^T & \dots & (\mathbf{p}_{\bar{s}+1}^*)^T & \rho_{\bar{s}+2}^{1,*} & \rho_{\bar{s}+2}^{2,*} & \rho_{\bar{s}+3}^{1,*}\end{bmatrix}^T.$$

The first $2+3(\bar{s}-1)$ equalities in Eq. (4) are then

$$\begin{bmatrix} \mathbf{Q}_{0,\mathbf{I}}^0 & \mathbf{Q}_1^0 & \mathbf{Q}_2^0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0}\\ \mathbf{Q}_{-1}^1 & \mathbf{Q}_{0,\mathbf{I}}^1 & \mathbf{Q}_1^1 & \mathbf{Q}_2^1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0}\\ \mathbf{Q}_{-2}^2 & \mathbf{Q}_{-1}^2 & \mathbf{Q}_{0,\mathbf{I}}^2 & \mathbf{Q}_1^2 & \mathbf{Q}_2^2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0}\\ & & \vdots & & & & \ddots & & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Q}_{-2}^{\bar{s}-1} & \mathbf{Q}_{-1}^{\bar{s}-1} & \mathbf{Q}_{0,\mathbf{I}}^{\bar{s}-1} & \mathbf{Q}_1^{\bar{s}-1} & \mathbf{Q}_2^{\bar{s}-1} & \mathbf{0}\end{bmatrix}\mathbf{p}^* = \mathbf{0}, \tag{15}$$

where $\mathbf{Q}_{0,\mathbf{I}}^s = \mathbf{Q}_0^s - \mathbf{I}$. We adapt the equalities of Eq. (4) associated with $\mathbf{p}_{\bar{s}}^*$ and $\mathbf{p}_{\bar{s}+1}^*$ using Eqs (6) and (7) and the definition of ρ_s^* [Eq. (8)], which yields

$$\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_{-2} & \mathbf{Q}_{-1} & \mathbf{Q}_0-\mathbf{I} & \mathbf{Q}_1 & \cdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{-2} & \mathbf{Q}_{-1} & \mathbf{Q}_0-\mathbf{I} & \end{bmatrix}$$
$$\begin{matrix} \mathbf{q}_2 & 0 & 0 \\ \mathbf{q}_{11} & \mathbf{q}_{12} & \mathbf{q}_2 \end{matrix}\Bigg]\mathbf{p}^* = \mathbf{0}. \tag{16}$$

Further, we add a set of equalities that ensures that the unstable system modes are not excited. In particular, with $\mathbf{T}_\rho$ such that $\bar{\rho}_{\bar{s}+2} = \mathbf{T}_\rho\mathbf{p}$, we have as necessary condition

$$\bar{\mathbf{V}}_{\rho 2}\mathbf{T}_\rho\mathbf{p}^* = \mathbf{0}. \tag{17}$$

In the set of equations (15-17), we have $n_e = 6+3\bar{s}$ equations and an equal number of unknowns. However, less than n_e equations are linearly independent, since otherwise only the zero vector is a solution, which is impossible. We believe that for most relevant p and α, there are exactly $n_e - 1$ linearly independent equations in the set, which is the subject of further investigation. The n_eth linearly independent equation originates from the fact that the probabilities sum up to one and, with Eq. (12), can be written as

$$\sum_{s=0}^{\bar{s}-1}\mathbf{p}_s^* + W\mathbf{T}_\rho\mathbf{p} = 1.$$

We now have a set of n_e linearly independent equations in n_e unknowns, with as unique non trivial solution $\mathbf{p}_s^*$ for $s = 0, 1, \dots, \bar{s}+1$. For separations $s \geq \bar{s}+2$, we simulate the autonomous LTI system in Eq. (11) with $\mathbf{T}_\rho\mathbf{p}^*$ as initial condition. From LTI system (11), it is clear that for $s \to \infty$, $\mathbf{p}_s^*$ decays exponentially to zero.

V. Discussion and Examples

In this section, we give examples of two-agent spatial distributions for a set of probabilities p, and discuss some properties.

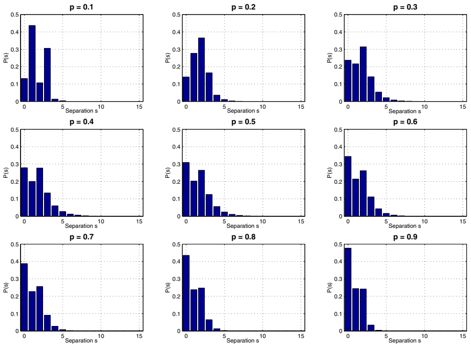

Fig. 4. Steady state agent separation probability distributions for different values of p, under an optimal policy.

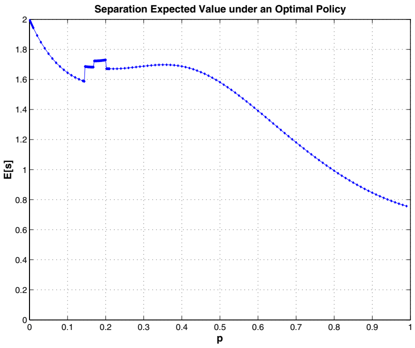

Fig. 5. The agent separation expected value under an optimal policy as function of p.

Fig. 4 shows the probability distribution of the two-agent separation in steady state and under an optimal policy for $p = 0.1, 0.2, \dots, 0.9$. The associated expected value of the steady state agent separation is shown in Fig. 5. It can be seen that for p close to one, where almost all edges have zero cost, the expected agent separation is the smallest and equals approximately 0.75. On the other hand, for p close to zero, we have the largest expected agent separation, equal to 2. This reinforces the idea that as fewer opportunities are observed, the agents spread out more to increase the size of the environment where exact edge costs are observed, increasing the probability of encountering a zero edge cost. Conversely, for large p, the agents remain close to take advantage of opportunities the other agent observes in the case of an unfavorable situation.

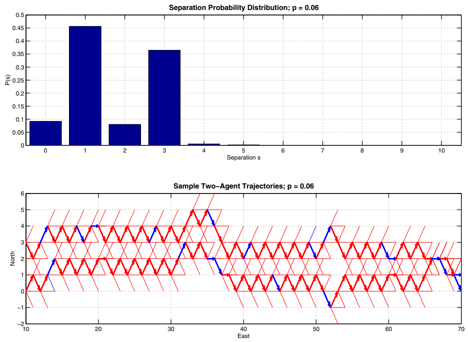

Fig. 6. Top: Steady state agent separation probability distribution for a p close to zero ($p = 0.06$), under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.06$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively.

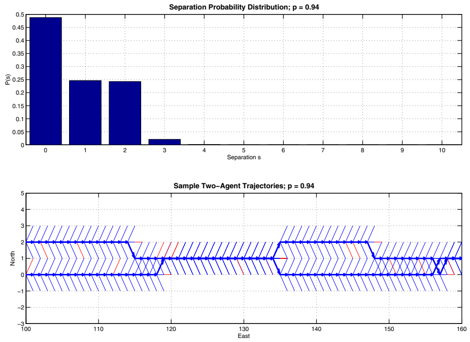

Fig. 7. Top: Steady state agent separation probability distribution for a p close to one ($p = 0.94$) under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.94$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively.

Two extreme but instructive cases, for $p = 0.06$ and $p = 0.94$, are shown in Fig. 6 and in Fig. 7, respectively; in the top half, the steady state separation spatial distribution is shown, while in the bottom half a set of sample trajectories is depicted. For $p = 0.06$, *i.e.* the case where ones are abundant, we see (Fig. 6) that the separations $s = 1$ and $s = 3$ are most probable. In the corresponding sample trajectories, one can see that agents tend to go to separation one, increasing the probability of agent A observing a cheap edge that is reachable for agent B and vice versa. However, with high probability, only ones are observed, driving the agents apart to $s = 3$, where a set of eight previously unobserved edges is observed, thus maximizing the probability of encountering a zero. Again, with high probability, only ones are in sight, and the agents converge again to $s = 1$, where eight previously unobserved edges enter the observation zone, finishing the "expected" cycle.

For $p = 0.94$, the case where zeros are abundant, we see (Fig. 7) that the separation $s = 0$ is most likely, followed by $s = 1$ and $s = 2$, both equally likely. The sample trajectories indicate the mechanics of cooperation here. In particular, let the agents start at $s = 1$, the most advantageous separation. Most likely, only zeros are observed, and agents continue straight ahead. With probability $p_1 = 2p(1-p) = 0.11$ (for p = 0.94), an edge of cost one enters an observation zone two stages ahead (see for example stage 118 in Fig. 7). Consequently, the agents converge to $s = 0$, which is maintained till again a one appears, with probability $p_2 = 1 - p = 0.06$ (for $p = 0.94$). The agents split to $s = 2$, maximizing the number newly observed edges. Again, one edge cost of one enters an observation zone with probability p_1, which causes the agents to converge to $s = 1$, where the "expected" cycle is repeated. The difference in magnitude of p_1 and p_2 clarifies the difference of the probabilities with which agent are at separation $s = 0$ and at the separations $s = 1$ and $s = 2$.

VI. Conclusion

In this paper, we study two-agent navigation on a partially unknown graph. The agents observe and share the cost of the edges in a local observation zone. On unobserved edge costs, only *a priori* statistics are available. In this paper, we build on previous work to compute the steady state probability distribution of the agent separation under an optimal policy by exploiting the structure of the underlying Markov chain. The results confirm and quantify the intuition that the more 'hostile' the environment, the more beneficial for the agents to spread out and observe more, increasing the probability of encountering an opportunity.

VII. Acknowledgments

This research is supported in part by Cooperative Control of Distributed Autonomous Vehicles in Adversarial Environments, 0205-G-CB222, University of California and by STTR16, Reinforced Learning and Genetic Learning Classifier Systems for Sensor Management, Nascent Technology Corporation.

References

[1] D. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 1995.

[2] J. Hespanha, H. Kim, and S. Sastry. Multiple-agent probabilistic pursuit-evasion games. In *IEEE CDC*, December 1999.

[3] F. Hillier and G. Lieberman. *Introduction to Operations Research*. Holden-Day, Inc., San Francisco, 1974.

[4] A. Jadbabaie, J. Lin, and A. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, June 2003.

[5] J. De Mot and E. Feron. Spatial distribution of two-agent clusters for efficient navigation. In *IEEE Conf. on Decision and Control*, December 2003.

[6] J. De Mot and E. Feron. Optimal two-agent navigation with local environment information. In *42nd Annual Allerton Conference on Communication, Control and Computing*, 2004.

[7] R. Olfati-Saber and R. Murray. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on Automatic Control (submitted)*, June 2004.

[8] G. Ribichini and E. Frazzoli. Energy-efficient coordination of multiple-aircraft systems. In *IEEE Conference on Decision and Control*, 2003.