July, 2005

LIDS Publication # 2656

Research supported in part by:

Air Force - DARPA - MURI award 009628001

A Multi-Agent Route Exploration Problem

Le Ny, J; Feron, E

A Multi-Agent Route Exploration Problem

Jerome Le Ny and Eric Feron

***Abstract*— We investigate a route exploration problem with N agents dropped randomly on the interval $[0,b]$ and discuss the impact of using multiple agents to perform this task. We consider both a discrete and a continous description of the path to explore. Independantly, we study an exploration problem with probabilistic agents having limited autonomy. In both problems , multi-agent scenarios are discussed with an emphasis on the number of agents necessary to obtain good performance.**

I. Introduction

Consider a line segment of length b, with coordinate x describing a position on the segment. The endpoints are $x = 0$ and $x = b$. The coordinates can take their values in the discrete set $[\![0,b]\!] := \{0,1,\ldots,b\}$, in which case we obtain a line graph with $(b+1)$ equally spaced sites, or they can take their value in the continous interval $[0,b]$.

We have N agents with initial positions x_i, $i = 1\ldots N$. When these initial positions are realizations of the associated random variables X_i, $i = 1\ldots N$, we denote the corresponding order statistics $(X_{1:N}, X_{2:N}, \ldots, X_{N:N})$, that is, the variables X_i arranged in increasing order: $X_{1:N} \leq X_{2:N} < \ldots < X_{N:N}$. We assume a continuous distribution function for the random variables X_i, and therefore we have $P(X_i = X_j) = 0$ for $i \neq j$ (see for example [1] p.29).

The agents can move along the continous line with the same speed v. When the line is discrete, we also discretize the time: in that case, at each period an agent can either move to a site that is next to its current position, or remain at its current position.

Finally we also consider non-compliant agents, that react probabilistically to given controls. More precisely, a non-compliant agent demonstrates the following behavior:

- in the continuous case: each agent moves with speed $v+\sigma W_t$, with σ a constant and W_t 1-dimensional white Gaussian noise with unit power spectral density.
- in the discrete case: when we tell the agent to move one step in a given direction, it might indeed move in that direction with probability p, but might go in the opposite direction with probability $q < p$ and also stay where it is with probability $1-p-q$. To a one step displacement corresponds a random variable X: its mean is analogous to the "speed" of the agent and therefore we write $v = p-q$. Its standard deviation is $\sigma = p+q+2pq-p^2-q^2$.

v and σ play a similar role in the continous and the discrete case, therefore we use the same notation in both cases. There will not be any confusion since we consider these models separately.

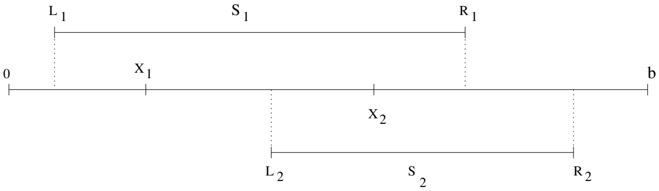

Fig. 1. Problem description and notation in the case of two agents.

Part II and III examine how to optimally explore the line with randomly dropped deterministic agents for a specific cost function, and how many agents we should use. Independantly, part IV and V discuss simple exploration strategies for non-compliant agents as well as trade-offs appearing in multi-agent exploration scenarios.

II. Deterministic Optimal Exploration Policy

In this part we consider the continous model, where the agents respond deterministically to the controls. Extension to the discrete case is straightforward. We assume a cost proportional to the distance that each agent travels. A possible motivation includes the risk of losing agents along the path in a hostile environment, increasing as the agents cover a longer distance. Another example could be that we want to minimize the amount of energy used by each agent. In the optimization problem, we seek to minimize the sum of the distances covered by all the agents. We have N agents initially at given distinct positions $0 \leq x_1 < x_2 < \ldots\ldots < x_N \leq b$. To agent $i = 1,\ldots,N$, we assign a part of the line to explore, called S_i, and let $L_i = \min S_i$ and $R_i = \max S_i$. Fig. 1 describes the notation in the case of two agents. Each agent explores its assigned region optimally by travelling a distance $d_i = [(R_i - L_i) + \min(x_i - L_i, R_i - x_i)]$, that is, it travels to the nearest endpoint first and then to the opposite endpoint.

The problem of minimum cost exploration For N agents becomes designing each set S_i so that when each of them is explored optimally by the corresponding agent, the sum of the minimum distances is minimized:

$$\begin{array}{rc}\text{minimize} & \sum_{i=1}^{N}[(R_i - L_i) + \min(x_i - L_i, R_i - x_i)] \\ \text{subject to} & R_i \geq x_i \geq L_i, \quad i = 1,\ldots N \\ & R_i \geq L_{i+1}, \quad i = 1,\ldots N-1 \\ & \min\{L_1,\ldots,L_N\} = 0 \\ & \max\{R_1,\ldots,R_N\} = b \end{array} \qquad (1)$$

This work was supported by Air Force-DARPA-MURI award 009628-001.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA (emails: jleny@mit.edu; feron@mit.edu).

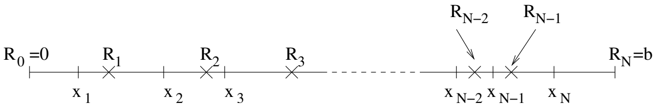

Fig. 2. Problem reformulation in the case of identical agents. The decision variables are the R_i's.

Our design variables are the L_i's and R_i's; the constraints make sure that the line is completely covered. The following lemma formalizes the intuitive result that in general exploration sets should not overlap.

Lemma 1: There exist an optimal solution for (1) satisfying:

$$L_1 = 0,\ R_N = b,\ R_i = L_{i+1},\ i = 1,\dots N-1.$$

Proof: Although this lemma is intuitively clear, due to the fact that we are considering identical agents, it is more tedious to prove formally. Consider an optimal solution for (1) (an optimal solution exists because of the interpretation of the problem, or alternatively because we are minimizing a concave function over a bounded polyhedron). Let j be an index such that $L_j = 0$. If $j \neq 1$, we consider a modification to the solution such that $L_j = R_1$ and $L_1 = 0$. Since the agents are identical, the fact that the leftmost part of the line is covered by agent j or agent 1 does not change the solution. Consequently we have an optimal solution such that $L_1 = 0$. Similarly we can choose the optimal solution such that $R_N = b$.

Now consider two agents i and $i+1$, $i \in [\![1, N-1]\!]$. If $R_i > L_{i+1}$ we consider a modification of the solution (denoted with a $'$) such that $R'_i = L'_{i+1} = \frac{R_i - L_{i+1}}{2}$. This modification implies no change on the other variables in the original solution. The interval $[L_{i+1}, R_i]$ was previously covered at cost $2(R_i - L_{i+1})$ in the case where the agents were not switching direction at the endpoints; it is now covered at cost $(R_i - L_{i+1})$. In the other cases where one or two agents switch direction to travel to their other endpoint, we verify that the cost is still divided by two. Therefore the initial solution could not be optimal, and the lemma is proved. ■

We do not have in general unicity of the optimal solution (as it will become clear in the following, there is an optimal solution using only one or two agents in any case). However the lemma is useful in restricting our analysis to some natural configurations. The problem then reduces to the following (see Fig. 2): refering to the point $R_i = L_{i+1}$ as the point R_i, $i = 1,\dots N-1$, $R_0 = 0$, $R_N = b$, we want to find the positions of the points R_i in order to minimize $\sum_{i=1}^{N}[(R_i - R_{i-1}) + \min(x_i - R_{i-1}, R_i - x_i)]$, which is rewritten:

$$\text{minimize } b + \sum_{i=1}^{N} \min(x_i - R_{i-1}, R_i - x_i) \qquad (2)$$

$$\text{subject to } R_0 = 0,\ R_N = b,$$
$$x_i \le R_i \le x_{i+1},\ i = 1,\dots N-1.$$

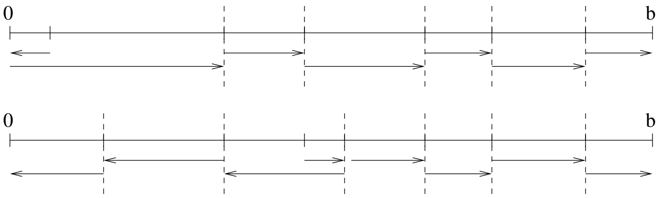

Fig. 3. Optimal Exploration Strategy to minimize the sum of individual distances. The two cases correspond to the leftmost interval being the shortest interval or not. Only one agent switches direction, and only the smallest initial interval between agents is covered twice.

Proposition 2: Let $x_0 = 0$, $x_{N+1} = b$, and let $j = \arg\min_{i=0,\dots,N}\{(x_{i+1} - x_i)\}$. Then we have:

- if $j = 0$, $R_i = x_{i+1}$, $i = 1,\dots,N$ is optimal for (2).
- if $j > 0$, $R_i = x_i$, $i = 0,\dots,(j-1)$, $R_i = x_{i+1}$, $i = j,\dots,N$ is optimal for (2).

The cost of an optimal solution of (2) is

$$Z_d(N) = b + \min_{i=0,\dots,N}\{(x_{i+1} - x_i)\}. \qquad (3)$$

Fig. 3 illustrates this optimal solution, which actually looks relatively clear. In words, we find the smallest interval between two consecutive agents. Next we choose one of these two agents, which will have to explore the intervals on both sides of its initial position. All other agents will have only one interval to explore. Again we do not have unicity of the solution, the choice for the directions of exploration is arbitrary for instance. It is also clear that we can perform the task with the same cost using only one agent if the shortest interval is at the extremities, or two agents in the other cases. However our solution is still interesting because it leads to an optimal solution in the discrete case as well: using only one or two agents on the discrete line forces them to travel over sites that are occupied by the other agents remaining idle. Whereas in the continuous model this does not contribute to an additional displacement cost because the agents positions are represented by points of measure 0, the additional cost of this policy in the discrete case appears clearly.

Proof: We prove the following result by induction on N: let N agents with initial positions $x_1 < \dots < x_N$ in the interval $[\alpha, \beta]$, $0 \le \alpha < \beta \le b$; then the optimal cost for the exploration of this interval is $\beta - \alpha + \min_{i=0,\dots,N}\{(x_{i+1} - x_i)\}$, with the convention $x_0 = \alpha$ and $x_{N+1} = \beta$.

The result is trivial for one agent. Now suppose the result true for all $k \le N-1$, and we want to prove it for N agents, $N \ge 2$. Since we have $N+1$ intervals to explore between the starting points, and only N agents, at least one agent has to switch direction and travel to R_{i-1} and R_i. Consider the agents $1,\dots,N$ in increasing order, and call p the first agent to switch direction. Agents $1,\dots,p$ are exploring $[\alpha, R_p]$, agents $p+1,\dots,N$ are exploring $[R_p, \beta]$, and the second group should explore its part optimally. Thus the induction hypothesis applies for the second group and the total exploration cost is therefore

$$\begin{aligned} &R_p - \alpha + \min(x_p - x_{p-1}, R_p - x_p) \\ &\quad + \beta - R_p + \min(x_{p+1} - R_p, x_{p+2} - x_{p+1}, \dots, \beta - x_N). \end{aligned}$$

Now if $x_p - x_{p-1} < R_p - x_p$, we obtain a cost of

$$\begin{aligned} &\beta - \alpha + (x_p - x_{p-1}) + \min(x_{p+1} - R_p, \dots, \beta - x_N) \\ &\geq \beta - \alpha + (x_p - x_{p-1}) \\ &\geq \beta - \alpha + \min_{i=0,\dots,N} \{(x_{i+1} - x_i)\}. \end{aligned}$$

If $x_p - x_{p-1} \geq R_p - x_p$, we obtain a cost of $\beta - \alpha + R_p - x_p + \min(x_{p+1} - R_p, \dots, \beta - x_N)$. Considering two cases for the last term, where the minimimum achieved is either $(x_{p+1} - R_p)$ or not, we see readily that in any case we obtain a cost lower bounded by an expression of the form $\beta - \alpha + x_{i+1} - x_i$ for some i, and therefore a lower bound on the cost is again $\beta - \alpha + \min_{i=0,\dots,N}\{(x_{i+1} - x_i)\}$.

Now it is also easy to see that the solution given in the proposition achieves this lower bound, which proves the recursion for N agents. ■

III. Agents Dropped Randomly on the Line

With the cost function considered in the previous part, clearly there is no benefit in using multiple agents if we have precise control on the initial position of these agents. We obtain the best possible solution by simply placing one agent at 0 and letting it travel to the other end of the route. However, we are interested in the case where agents are dropped randomly on the path. More precisely, we assume in this part that the initial positions are realizations of N iid random variables $X_1, \dots, X_N$ having uniform distribution on the interval [0,b]. Using more agents becomes beneficial in expectation, as we can reduce the minimum initial interval between them, which is the only variable part in the optimal cost (3).

As described in part I, we define the order statistics $X_{1:N}, \dots, X_{N:N}$, where the notation is useful to keep track of the number of agents N. Now denote $D_1 = X_{1:N}$, $D_i = X_{i:N} - X_{i-1:N}$ for $i = 2, \dots, N$, and $D_{N+1} = b - X_{N:N}$. The variables D_i are referred to as *spacings*. We are interested in the distribution and the expected value of $\min_{i=1,\dots,N+1}\{D_i\}$. The distribution of the spacings is a classical result:

Lemma 3: Let $X_1, \dots, X_N$ be iid random variables, uniformly distributed on $[0, b]$. Let $c_1, \dots, c_{N+1} \geq 0$, such that $\sum_{i=1}^{N+1} c_i \leq b$. Then we have

$$P(D_1 > c_1, \dots, D_{N+1} > c_{N+1}) = (1 - \frac{c_1}{b} - \dots - \frac{c_{N+1}}{b})^N.$$

Proof: For a proof of this result when $b = 1$, we refer for example to [2]. The result of the lemma follows by scaling, i.e. dividing all the quantities by b to obtain the base case. ■

Taking $c_1 = \dots = c_{N+1} = x$, we get $P(\min_{i=1,\dots,N+1}\{D_i\} > x) = P(D_1 > x, \dots, D_{N+1} > x) = (1 - (N+1)(x/b))^N$ if $0 \leq x \leq \frac{b}{N+1}$. So the expected value of the minimum interval length is

$$\begin{aligned} E[\min_i\{D_i\}] &= \int_0^{\frac{b}{N+1}} P(\min_i\{D_i\} > x)dx \\ E[\min_i\{D_i\}] &= \int_0^{\frac{b}{N+1}} (1 - (N+1)\frac{x}{b})^N dx \\ E[\min_i\{D_i\}] &= \frac{b}{(N+1)^2} \end{aligned} \tag{4}$$

and the expected optimal cost function is

$$Z_r(N) = b\left(1 + \frac{1}{(N+1)^2}\right) \tag{5}$$

It is clear now that we have a “saturation” effect when we use more agents, since the cost function goes asymptotically towards b. If we add a penalty for using more agents, for example we add a linear term $c_a N$ to $Z_r(N)$, we can solve for the optimal number of agents for the task

$$N^* = \left(\frac{2b}{c_a}\right)^{\frac{1}{3}} - 1 \tag{6}$$

where c_a represents the cost per agent. This solution has to be adapted slightly in order for N^* to be an integer.

It is interesting to note that N^* grows relatively slowly with b. For example, the intuition that in order to explore a line of length $2b$ we need twice the number of agents used to explore a line of length b leads to a large overestimate.

IV. Agents with Probabilistic Behavior: Continuous Model

A. Feedback Strategy

We now turn to a situation where we have agents with a probabilistic behavior as described in part I. First, we consider the exploration problem with a single agent in the continuous model. Suppose the agent initially at x_0 moves along the continuous version of the line $[0, b]$ with speed $u(t)v + \sigma W_t$, where v is a positive constant, $u(t) \in [-1, +1]$ is the control, σ the standard deviation is a positive constant and W_t is white Gaussian noise with unit power spectral density. The position of the agent follows the following stochastic differential equation:

$$dX_t = u(t)v\,dt + \sigma dB_t \tag{7}$$

where B_t is 1-dimensional Brownian motion. Since v is constant, the individual cost function can equivalently be the time spent moving or the distance travelled.

Several strategies can be employed to explore the line with one such agent. We first review the optimal control result, i.e. a strategy minimizing the expected exploration time, that can be implemented on an agent with positioning or communication capacities.

Proposition 4: Note $X(t)$ the position of the agent on the line. The optimal feedback law $u(X(t))$ to bring the agent in minimum time to 0 or b is:

$$\begin{cases} u = -1 & \text{if } X(t) \in [0, \frac{b}{2}] \\ u = +1 & \text{if } X(t) \in (\frac{b}{2}, b] \end{cases}$$

The minimum expected time to hit the boundary at 0 or b is:

$$\begin{cases} \frac{x_0}{v} - \frac{\sigma^2}{2v^2} e^{\frac{-vb}{\sigma^2}} \left[e^{\frac{2vx_0}{\sigma^2}} - 1 \right] \text{ if } x_0 \leq \frac{b}{2} \\ \frac{b-x_0}{v} - \frac{\sigma^2}{2v^2} e^{\frac{-vb}{\sigma^2}} \left[e^{\frac{2v(b-x_0)}{\sigma^2}} - 1 \right] \text{ if } x_0 \geq \frac{b}{2}. \end{cases} \tag{8}$$

Proof: Let $f(x,t) = \min_u E_x\{\tau - t\}$, where τ denotes the first time the agent hits the boundary at 0 or b, and x is the initial position of the agent (E_x denotes the expected value given $X(0) = x$). The dynamic programming equation for f is [3]:

$$-\frac{\partial f}{\partial t}(x,t) = min_{u \in [-1,+1]} \left\{ 1 + uv\frac{\partial f}{\partial x}(x,t) + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}(x,t) \right\}$$

It is readily seen that the minimization occurs for $u = -sign\left(\frac{\partial f}{\partial x}(x,t)\right)$, and the twice continuously differentiable solution of the corresponding equation with boundary conditions $f(0,t) = f(b,t) = 0$ is given in the proposition. ■

Proposition 4 tells us how to hit optimally the first boundary. It is intuitively clear that once the first boundary has been hit, the control should remain constant telling the agent to travel as fast as possible to the other boundary. The process describing the agent position then reduces to a Brownian motion with drift between a reflecting barrier (the boundary already hit) and an absorbing barrier (the second boundary to reach). Here we add a lemma based on the calculations in [4] for this process, that is useful to obtain closed-form results in the following.

Lemma 5: Consider the agent subject to constant control, moving towards a target $h = 0$ or b. If $E(\tau|x_0)$ is the expected time for the agent to reach the target, we have:

$$E(\tau|x_0) = \frac{|h - x_0|}{v} - \frac{\sigma^2}{2v^2} \left[e^{-\frac{2v(b-|x_0-h|)}{\sigma^2}} - e^{-\frac{2vb}{\sigma^2}} \right] \tag{9}$$

Hence, using (9) and proposition 4 gives immediately the total optimal expected time T_{opt} necessary to explore the line with one agent using position feedback. For example in the case $x_0 \leq b/2$ we get:

$$T_{opt} = \frac{b + x_0}{v} - \frac{\sigma^2}{2v^2} \left[1 - e^{\frac{-2vb}{\sigma^2}} + e^{-\frac{v(b-2x_0)}{\sigma^2}} - e^{-\frac{vb}{\sigma^2}} \right] \tag{10}$$

Note that the agent is faster than in the deterministic case.

B. Open-Loop Strategy for One Non-Compliant Agent.

The implementation of the optimal policy requires that the agent knows its exact position on the line at each instant, at least with respect to the point $b/2$. In the rest of this section, we discuss "open-loop" strategies, assuming the agents used do not have the capacities to receive feedback instructions during the exploration. A straightforward approach is to see at the beginning of the mission which is the endpoint closest to the initial position. Next, tell the agent to move first towards that endpoint until it reaches it, and then towards the other endpoint until it reaches it.

With a constant control, we know that an agent will eventually reach a given target with probability one: in the discrete case for example, this is given by the ergodicity of the underlying Markov chain describing the position. Again, (9) can be used to obtain the expected time T_{ol} necessary to finish the exploration. In the case $x_0 \leq b/2$ we get:

$$T_{ol} = \frac{b + x_0}{v} - \frac{\sigma^2}{2v^2} \left[1 + e^{-\frac{2v(b-x_0)}{\sigma^2}} - 2e^{-\frac{2vb}{\sigma^2}} \right]. \tag{11}$$

The difference with T_{opt} is then found to be bounded in every case as follows:

$$T_{ol} - T_{opt} \leq \frac{\sigma^2}{2v^2} \left(1 - e^{\frac{-vb}{\sigma^2}} \right)^2,$$

which tells us that we are not loosing much if we do not implement any feedback.

Because in practice the agent has limited autonomy, it is useful to know more about the distribution of the exploration time. Let us describe the position of the agent by the process X_t starting at $X_0 = x_0$, and consider the time necessary for the agent to be absorbed at 0 with a high enough probability, treating b as a reflecting barrier. A bound on this time is obtained by considering an auxilliary process Y_t describing the movement of an agent with the same dynamics but on an infinite line (i.e. without barriers at 0 and b). We have

$$P(X_t = 0 | X_0 = x_0) \geq P(Y_t \leq 0 | Y_0 = x_0). \tag{12}$$

The reason is simply that before hitting 0, both processes have the same behavior; but once X_t hits 0 we know for sure that it will remain there, whereas Y_t might become positive again after it hits 0 for the first time. From this idea we get the following proposition.

Proposition 6: Let $0 < \varepsilon < 1$ and α be given by $\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{z^2}{2}} dz = 1 - \varepsilon$. Suppose without loss of generality that the agent is initially at $x_0 \leq b/2$ and therefore told to move to 0 first. The agent has reached 0 with probability at least $1 - \varepsilon$ for $t \geq t_0$, where t_0 is given by

$$t_0 = \frac{x_0}{v} + \frac{\alpha\sigma}{v}\left(\frac{x_0}{v}\right)^{\frac{1}{2}} + \frac{\alpha^2\sigma^2}{v^2} \tag{13}$$

Proof: Write $Y_t = (x_0 - vt) + \sigma\sqrt{t}\chi$, where χ is a random variable with standard normal distribution. Then we solve for $P\left(\chi \leq \frac{vt-x_0}{\sigma\sqrt{t}}\right) \geq 1 - \varepsilon$. This is obtained for $\frac{vt-x_0}{\sigma\sqrt{t}} \geq \alpha$. Solving for equality in this inequality, we obtain for t_0:

$$t_0 = \left[\frac{\alpha\sigma + \sqrt{\alpha^2\sigma^2 + 4x_0v}}{2v} \right]^2$$

This is simplified to obtain a lower bound as follows:

$$\begin{aligned} \left[\frac{\alpha\sigma}{2v} + \left(\frac{x_0}{v} + \frac{\alpha^2\sigma^2}{4v^2} \right)^{\frac{1}{2}} \right]^2 &= \frac{\alpha^2\sigma^2}{2v^2} + \frac{x_0}{v} \\ &\quad + \frac{\alpha\sigma}{v}\left(\frac{x_0}{v} + \frac{\alpha^2\sigma^2}{4v^2} \right)^{\frac{1}{2}} \\ &\leq \frac{\alpha^2\sigma^2}{v^2} + \frac{x_0}{v} + \frac{\alpha\sigma}{v}\left(\frac{x_0}{v}\right)^{\frac{1}{2}} \end{aligned}$$

using $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x,y \geq 0$. ■

Remark 7: For $\alpha > 0$, we have $\int_\alpha^\infty e^{-\frac{z^2}{2}} dz \leq \frac{1}{\alpha} e^{-\frac{\alpha^2}{2}}$. In particular when $\alpha \geq 1$ and $\varepsilon < \frac{1}{\sqrt{2\pi}}$, we obtain a more conservative bound which is easier to apply , choosing α such that $e^{-\frac{\alpha^2}{2}} \leq \sqrt{2\pi}\varepsilon$ i.e. $\alpha \geq \sqrt{2 \ln \frac{1}{\sqrt{2\pi}\varepsilon}}$.

Remark 8: An exact expression for the distribution of the exploration time can be seen as a special case of the calculations in [5]. However, our bound above is easier to interpret and sufficient for our purpose.

From the proposition, we can immediately conclude that for $0 < \varepsilon < 1/2$ and the agent going to 0 and then to b, the line will be completely explored with probability at least $1 - 2\varepsilon$ for $t \geq t_1$, where:

$$t_1 = \frac{b + x_0}{v} + \frac{\alpha\sigma}{v}\left[\left(\frac{b}{v}\right)^{\frac{1}{2}} + \left(\frac{x_0}{v}\right)^{\frac{1}{2}}\right] + \frac{2\alpha^2\sigma^2}{v^2}$$

Note that this simple open-loop strategy is asymptotically optimal in the limit where $b \rightarrow \infty$, since $(b + x_0)/v$ is the time for a deterministic agent to explore the line. Also, at the limit when $\alpha = 0$, i.e. $\varepsilon = 1/2$, we obtain the same speed as in the deterministic case, provided we can be satisfied with a very low probability of success. In fact the bound is not tight due to the crude use of Boole's inequality to obtain $1 - 2\varepsilon$ for the probability of success in the two successive travel periods. If we look only at the travel from x_0 to 0 in the proposition, we see that in fact we can be *faster* than in the deterministic case by allowing ε to be greater than $1/2$, i.e. $\alpha < 0$. This appears natural as we can exploit the possibility that the speed might take values well above its mean.

C. Agents with Limited Autonomy.

Suppose we have an infinite number of non-compliant agents that we can use to explore the interval $[0, b]$, all starting from b at the beginning of the mission. The mission terminates when an agent reaches 0. These agents work under the open-loop policy described in the previous paragraph, since it was argued that in general, adding position feedback does not increase dramatically the performance. We illustrate in this part applications of the previous results for two multi-agent exploration scenarios.

If every agent can only run for a time t_0, the line segment exploration problem can be seen as a Monte-Carlo algorithm[6]; that is, the algorithm might sometimes produce an incorrect answer but we are able to bound the probability of that incorrect answer using (13). The running time of this "algorithm" is guaranteed to be $t_0(\varepsilon)$ (where the notation is showing the dependance in ε explicitely) and the probability of the result being correct is at least $1 - \varepsilon$. To improve the probability of success of a Monte-Carlo algorithm, we simply run it repetetively, trading-off running time. This means for our task that we can send multiple agents successively, and let each of them run for t_0, until one of them finishes the task. The expected time it takes to finish the exploration is then upper bounded by $t_0(\varepsilon)/(1 - \varepsilon)$, since the sucess event follows a geometric distribution with parameter $1 - \varepsilon$.

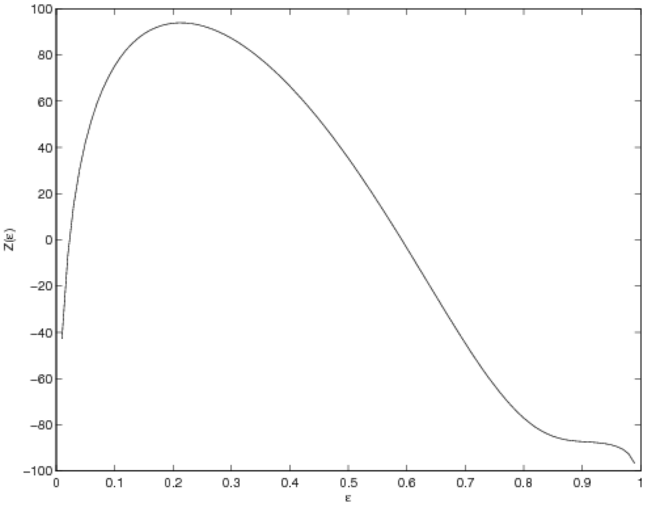

Fig. 4. Expected exploration reward. Parameters: $v = 1, \sigma = 2, b = 100, R = 1000, \gamma_1 = 10^{-2}, \gamma_2 = 1$. The optimal reward $Z_{opt} = 94$ is obtained for $\varepsilon = 0.21$.

Consider the following scenario. We assume that we collect an expected reward which is a function of the *expected time* to finish the exploration. Hence, let us consider an expected reward of the form $Re^{(-\gamma_1 t_0/(1-\varepsilon))}$ (more precisely, this should be a lower bound on the expected reward that we can achieve). There is also a linear cost $\gamma_2 t_0$ associated to the use of agents with a greater autonomy which are more expensive. The total expected reward is then

$$Z(\varepsilon) = Re^{\frac{-\gamma_1 t_0(\varepsilon)}{1-\varepsilon}} - \gamma_2 t_0(\varepsilon), \quad R, \gamma_1, \gamma_2 > 0 \qquad (14)$$

Since we can use multiple agents to finish the task, we will not necessarily need to require a high probability for one agent to finish correctly. There is a trade-off between the development of better agents with a greater autonomy and the reward that we can collect from them. Moreover, with multiple agents we should be able to use the cases where the random component of the speed allows a faster execution. Therefore, the optimal ε increases with the variance σ.

Various shapes can be obtained for the function $Z(\varepsilon)$ for different choices of parameters. Fig. 4 is an illustration for specific values.

Once we have computed the optimal ε, we might be interested in knowing how many agents will actually be necessary to perform the exploration in practice. As mentionned earlier, the expected number of agents used until one of them finishes is $1/(1 - \varepsilon)$. Standard Chernoff arguments apply to the corresponding geometric sequence to show that the number of agents used will be close to this expectation with high probability. Looking back at the example illustrated on Fig. 4, we obtain an optimal number of agents of about 1.26. Obviously a number of different scenarios can be studied in a similar way, but this tells us again that the cost of using multiple agents should be included in reasonable models, because the saturation effect already encountered in the previous part can be dominant.

V. Agents with Probabilistic Behavior: Discrete Model

In this part we extend some results of the previous section to the discrete model. These results would translate directly to the discrete case, except that using the central limit theorem to determine the bound on the expected exploration time would only give us a result in the limit $b \to \infty$. However, a concentration inequality allows us to obtain finite-time bounds.

A. Optimal Closed-Loop Policy for a Single Agent

The behavior of a non-compliant agent in the discrete case was described in part I. Remember that now we want to explore a line graph with $b+1$ vertices. An agent moves on the line following a controlled random walk: at a given period, it goes in the required direction with probability p, stays where it is with probability $1-p-q$ and goes in the opposite direction with probability q. The characteristics of the boundaries are as follows: if the agent is at 0 and told to go left, it will remain where it is with probability $1-q$ and go right with probability q. If told to go right, it will go right with probability p and stay at 0 with probability $1-p$ (asume $p>q$). The boundary at b is described symmetrically. Suppose that a single agent is initially at site x_0 on the discrete line, and that we want to explore the line while minimizing the expected exploration time. To determine the optimal policy minimizing the expected cover time for the corresponding controlled Markov chain, we can use a standard dynamic programming approach. This is summarized in the following proposition, which parallels the continuous case. It can be proved using the value iteration method as described in [7], for a stochastic shortest path problem on a finite number of states. Under these conditions, Bellman's equation holds. Since the result is now intuitively clear and the proof is straightforward but lengthy, we omit it.

Proposition 9: The optimal policy to explore the discrete line in minimum expected time with a non-compliant agent is to always send the agent towards the nearest still unvisited endpoint.

Since we know the optimal policy, we can compute the corresponding optimal expected cost (at least numerically) as a solution of the linear system corresponding to Bellman's equation, where we know the result of the minimization for each state. Solving the linear system analytically is difficult compared to the continuous case calculation, and instead we simply provide a lower bound result analogous to lemma 5

Lemma 10: Let $E(\tau|x_0)$ be the optimal expected travel cost for a single non-compliant agent initially at site x_0 and moving under constant control towards 0. We have

$$E(\tau|x_0) \geq \frac{x_0}{1-2q} + \frac{q^{b+1}}{(1-q)^b(1-2q)^2}\left[1-\left(\frac{1-q}{q}\right)^{x_0}\right] \tag{15}$$

Proof: This lower bound is obtained as follows: the dynamics of the agent follow a random walk between an absorbing barrier (the endpoint to reach) and a reflecting barrier (the endpoint already visited). The expected time is obtained as a solution of the corresponding subsystem in Bellman's equation. For the case $q=1-p$, this system was solved in [8], [9]. In our case however, we can have $p+q<1$. But if we consider the random walk with parameters p_1, q_1 such that $q_1 = q$ and $p_1 = 1-q$ (i.e. when the agent would remain idle in the original process, in the modified process it moves in the right direction), we obviously reach the target in a shorter time. The lower bound given in the proposition can therefore be obtained from [8], for our case where $q<1/2$. ■

This result can be used as before to argue that adding position feedback does not add a lot to the performance of the agent. This is because even in the optimal case, the agent will have to travel from the first hit boundary to the second one, and on this phase there is no difference between open-loop and closed-loop strategy. Using (15), we know then that the optimal policy will have a cost of at least $b/(1-2q)$. During the first phase, we can expect from the continuous model result that the feedback performance is also relatively close to the open-loop performance. We do not make the argument more formal here.

B. Open-Loop Policy

As in the continuous case, we consider simple open-loop policies that are in practice a lot easier to implement and should perform relatively well with respect to the optimum. So for an agent with limited autonomy, we tell the agent to go towards the closest endpoint for a fixed maximum number of steps, and then to switch direction and go towards the other endpoint again for a fixed number of steps. If the agent has infinite autonomy, it goes in each direction until it reaches its target, which happens with probability one. The implementation of the policy only involves mission pre-planning and no online re-planning.

We can derive a result analogous to the bound on the exploration time (13) in the continuous model. Consider an agent travelling under constant control from its initial position x_0 towards 0. If X_n represents the position at time n of the agent moving between the two barriers (absorbing at 0, reflecting at b), and Y_n is the position of an agent starting from x_0 and moving on an infinite discrete line which is an extension of our interval, with the same transition probabilities as X_n (but without barriers), we have:

$$P(X_n = 0 | X_0 = x_0) \geq P(Y_n \leq 0 | Y_0 = x_0), \quad \forall n$$

Define $Z_1, Z_2, \ldots$ iid random variables with $P(Z_i = -1) = p$, $P(Z_i = 0) = 1-p-q$, $P(Z_i = 1) = q$. Then we have:

$$Y_0 = x_0$$
$$Y_n = Y_0 + \sum_{i=1}^{n} Z_i\ , \quad n \geq 1$$

Let μ and σ be the mean and the variance of Z_i.

$$\mu = -p+q\ , \quad \sigma = p+q+2pq-p^2-q^2$$

We assume $p > q$ and therefore $\mu < 0$. Notice that $v = |\mu|$. With these notations, we have:

Proposition 11: Let $0 < \varepsilon < 1$, and $\alpha = \sqrt{2ln\frac{1}{\varepsilon}}$. Assume without loss of generality that the non-compliant agent starts at $0 < x_0 \leq \lfloor \frac{b}{2} \rfloor$. Then the agent moving under constant control towards 0 has reached 0 with probability at least $(1-\varepsilon)$ for $n \geq n_0$, with

$$n_0 = \frac{x_0}{v} + \frac{\alpha\sigma}{v}\left(\frac{x_0}{v}\right)^{\frac{1}{2}} + \frac{\alpha^2\sigma^2}{v^2} + \frac{1}{3v} \qquad (16)$$

If $i = 0$, obviously we take $n_0 = 0$ since it means that we start at the absorbing barrier. Note the similarity to the expression obtained in the continuous case, in particular when we use the expression for α mentionned in remark 7. In the limit where x_0 is large, we have $n_0 = \frac{x_0}{v}(1+\mathrm{o}(1))$. If we interpret $v = |\mu|$ as the mean speed of the agent, this result says that asymptotically for x_0 and b large we do not have to wait a lot more in the stochastic case than in a deterministic situation where we have an agent moving at speed v.

Proof: Let $S_n = \sum_{i=1}^{n} Z_i$. We will use Bernstein's inequality for our distribution on Z_i (see for example [10] for a survey of concentration inequalities):

$$\forall \delta > 0,\ P(S_n - \mu n \geq \delta n) \leq \exp\left(-\frac{n\delta^2}{2\sigma^2 + 2\delta/3}\right) \qquad (17)$$

Since $\mu < 0$, we can choose n_0 integer such that $n_0\mu < -x_0$. Then let $\delta = -\frac{x_0}{n_0} - \mu$, we have $\delta > 0$.

Now let n be an integer, $n \geq n_0$. Then we have $[-x_0, +\infty) \subset [-x_0\frac{n}{n_0}, +\infty)$ therefore $P(S_n \geq -x_0) \leq P(S_n \geq -x_0\frac{n}{n_0})$. Moreover, using (17) and our definition of δ, we have

$$P(S_n \geq -x_0\frac{n}{n_0}) = P(S_n - n\mu \geq (-\frac{x_0}{n_0} - \mu)$$
$$P(S_n \geq -x_0\frac{n}{n_0}) \leq \exp\left(-\frac{n(\frac{x_0}{n_0} + \mu)^2}{2\sigma^2 - \frac{2}{3}(\frac{x_0}{n_0} + \mu)}\right)$$

Note that with our constraint on δ, $2\sigma^2 - \frac{2}{3}(\frac{x_0}{n_0} + \mu) > 0$. Since $n \geq n_0$ and $P(Y_n \geq 0 | Y_0 = x_0) = P(S_n \geq -x_0)$, we obtain finally

$$P(X_n \neq 0 | X_0 = x_0) \leq \exp\left(-\frac{(x_0 + n_0\mu)^2}{2n_0\sigma^2 - \frac{2}{3}(x_0 + n_0\mu)}\right) \qquad (18)$$

To obtain $P(X_n \neq 0 | X_0 = x_0) \leq \varepsilon$ for $\varepsilon > 0$, it is sufficient to have

$$\frac{(x_0 + n_0\mu)^2}{2n_0\sigma^2 - \frac{2}{3}(x_0 + n_0\mu)} \geq \ln\frac{1}{\varepsilon}$$

We obtain the value for n_0 given in the proposition by considering only the solution greater than $x_0/|\mu|$. The final expression is simplified as in the proof of proposition 6.

■

Since proposition 11 is almost identical to proposition 6, it follows that our discussion on multi-agent exploration in the continuous model is valid for the discrete model as well.

VI. Conclusions

Two simple multi-agent line exploration problems were considered in this paper. The optimal policy for exploring the line with N agents seeking to minimize the sum of their travelled distances was obtained. For agents dropped randomly on the line, it was shown that adding a cost proportional to the number of agents leads to an optimal number of agents to use for the task. In a second part, we considered an exploration problem using non-compliant agents with limited autonomy. Again it was argued that the number of agents used to perform a given task should be considered as an important question. In practice using more agents has an associated cost and might not always lead to a dramatic increase in the final performance.

References

[1] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Duxbury, 2004.

[2] H. David and H. Nagaraja, *Order Statistics*, 3rd ed. Wiley, 2003.

[3] W. Fleming and R. Rishel, *Deterministic and Stochastic Optimal Control*. Springer, 1975.

[4] M. Dominé, “Moments of the first passage time of a wiener process with drift between two elastic barriers,” *Journal of Applied Probability*, vol. 32, pp. 1007–1014, 1995.

[5] ——, “First passage time distribution of a wiener process with drift concerning two elastic barriers,” *Journal of Applied Probability*, vol. 33, pp. 164–175, 1996.

[6] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995.

[7] D. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2001, vol. 1.

[8] B. Weesakul, “The random walk between a reflecting and an absorbing barrier,” *The Annals of Mathematical Statistics*, vol. 32, no. 3, pp. 765–769, 1961.

[9] A. Blasi, “On a random walk between a reflecting and an absorbing barrier,” *The Annals of Probability*, vol. 4, no. 4, pp. 695–696, 1976.

[10] S. Boucheron, O. Bousquet, and G. Lugosi, *Concentration Inequalities*. Springer-Verlag, 2004, vol. 1099, pp. 208–240.