Searching with Imperfect Information

F. Heukers
University of Twente

July 2, 2017

Abstract

When a non-optimal seeker is introduced into the cow path problem, the competitive ratio of the optimal deterministic algorithm increases. In this paper, we discuss the impact of the non-optimal seeker on the cow path problem using this optimal deterministic algorithm. Also, a method is proposed to improve this algorithm for the non-optimal seeker. The distance factor in the algorithm is made dependent on the probability of the seeker finding the object when he is upon it. Some examples are given were the competitive ratio of this adjusted algorithm is determined for the expected case and for 95% of the cases. Also, the impact of a non-optimal seeker on an extend version of the cow path problem, known as a star search, is discussed briefly.

1. Introduction

When searching for an object, we usually have some notion of where it could be. But what if the search area has no end? Then the object could be everywhere in that area and we are looking for an object in an unbounded domain: we have a search problem.

A well-known search problem is the cow path problem or linear search problem. This linear search problem was first introduced by Richard Bellman (**?**), and was independently researched by Wallace Franck (**?**) and Anatole Beck (**?**). Both their results were similar, but Franck's publication preceded Beck's. In essence Beck's paper provided a more comprehensible statement of the theorem already proved by Franck. Beck's next paper on this subject (**?**) gives some new results and gives some insight on a new area of study resulting from the theorem in his previous work.

The cow path problem can be described as a two-person zero-sum game between a hider H and a seeker S in an unbounded domain. The hider chooses a location T on the x-axis to hide an object. The seeker starts at the origin O and tries to find the object as soon as possible. It is assumed that the seeker has a constant speed of 1 and that changing direction can be done instantly. Beck and Newman (**?**) were the first to look at the cow path problem from this game theory perspective, resulting in them being able to prove that there exists a strategy for the seeker such that his loss never exceeds 9 times the distance from O to T. Later on, this loss became known as the *competitive ratio* of the algorithm.

The competitive ratio is a method of analyzing an online algorithm. We are dealing with an online algorithm, because we do not know all the input beforehand. If we did, we would know the location of the object and there would be no search problem. The competitive ratio of an online algorithm for this problem is the supremum of the ratio between the time the seeker needs to find the object and the time the seeker had needed if he had known the location.

Gal and Chazan showed that, for search games, the minimax trajectory is always a geometric sequence in the discrete case (**?**). A minimax trajectory is a search trajectory which minimizes the possible loss for a worst case scenario.

Gal was also the first to solve a generalized case of the cow path problem known as the *star search* (**?**). In this case the object is hidden somewhere in one of m rays. The cow path problem is in fact

a special case of the star search with $m = 2$. Gal showed that the the optimal strategy is to visit each ray periodically with monotonically increasing step sizes.
Beck and Beck considered the uniform and triangular distribution, where they showed that the optimal strategy is to go to one end and then the other if the hiding distribution is uniform, but the optimal strategy has infinity many turning points if the distribution is triangular (**?**). So, for some specific distribution, the problem is solved. But a solution to Bellman's original problem has still not been found.
But what is the appeal of this problem? Well, it is an easy problem to understand, but it is a lot harder than what you would have guessed at first sight. That can be concluded from the fact that even after more than fifty years the problem is still not solved.
In the nineties the cow path problem was formulated by computer scientists, while being unaware that this topic had already been researched by mathematicians. Baeza-Yates et al. also proved that the optimal deterministic strategy for the cow path problem has a competitive ratio of 9 (**?**). They named this optimal deterministic strategy *linear spiral search*. In linear spiral search the seeker alternates between going left and right and in each iteration the distance travelled from the origin is doubled.
In this paper a special case of the cow path problem will be discussed when a non-optimal seeker is in play. This special case was also mentioned as part b in Bellman's original problem (**?**), but since the first part of the problem is still not solved for a general probability distribution, not a lot of attention has been given to this question. This non-optimal seeker can be mathematically described with a probability p that he does not see the object when he is upon it. The expectation is that the seeker will have to search longer before finding the object. In theory the seeker can search arbitrarily long before finding the object. In this case it might be effective to turn earlier, so the same space is searched more often.
This paper is centred around the following question: What is the influence of a non-optimal seeker on the cow path problem? To answer this we first apply the linear spiral search algorithm on the cow path problem with a non-optimal seeker. Then, we will try to improve this algorithm for the non-optimal seeker by searching with a smaller step size. Also, we will look at the performance of the algorithm for the expected case and for 95% of the cases. Lastly, we take a quick look at a star search problem with a non-optimal seeker.

2. Applying Linear Spiral Search

The first problem that arises when a non-optimal seeker is introduced, is the fact that in theory the seeker could be searching for an infinite amount of time. This is the worst-case scenario. The probability that this will happen is zero, so therefore the average-case scenario will be analysed. To do this, the expected time E at which the seeker finds the objects is needed. The probability p of the seeker recognizing the object has a geometric distribution, since we are interested in the number of passings needed to get one success, which is the seeker recognizing the object when he is upon it. So the expected time of finding the object is $E = \frac{1}{p}$. Because the algorithm is discrete, it is assumed that $E \in \mathbb{N}$ [1].
For linear spiral search we use of a distance function $f_j = 2^{j-1}, \forall j \geq 1$. This function denotes the distance that is travelled during iteration j with respect to the origin in the direction opposite to the travelled direction during the previous iteration. So, the travelled distance with respect to the origin is doubled during each iteration.

[1] When $E \notin \mathbb{N}$, $\lceil E \rceil$ can be used instead of E.

Now, something can be said about the competitive ratio of the algorithm with a non-optimal seeker.

Theorem 2.1. *The competitive ratio of linear spiral search with a non-optimal seeker is* $2^{2+E} - 1$ *if* E *is even and* $2^{2+E} + 1$ *if* E *is odd.*

Proof. In the worst case, the distance d from the origin to the objective is $f_i + \epsilon$, where $f_i = 2^{i-1}$ and $\epsilon > 0$ is some small number. Then, the expected travel distance is:

$$2 \cdot \sum_{j=1}^{i+E} f_j + (-1)^{E+1} d.$$

Thus, there are two cases to be distinguished: E is even and E is odd. If E is even, the expected travel distance is:

$$\begin{aligned} 2 \cdot \sum_{j=1}^{i+E} f_j - d &= 2 \cdot \sum_{j=0}^{i+E-1} 2^j - d \\ &= 2 \cdot (2^{i+E} - 1) - 2^{i-1} - \epsilon \\ &= (2^{E+2} - 1) \cdot 2^{i-1} - 2 - \epsilon \\ &\leq (2^{E+2} - 1)(f_i + \epsilon) \\ &= (2^{E+2} - 1) \cdot d. \end{aligned}$$

If E is odd, the expected travel distance is:

$$\begin{aligned} 2 \cdot \sum_{j=1}^{i+E} f_j + d &= 2 \cdot \sum_{j=0}^{i+E-1} 2^j + d \\ &= 2 \cdot (2^{i+E} - 1) + 2^{i-1} + \epsilon \\ &= (2^{E+2} + 1) \cdot 2^{i-1} - 2 + \epsilon \\ &\leq (2^{E+2} + 1)(f_i + \epsilon) \\ &= (2^{E+2} + 1) \cdot d. \end{aligned}$$

So, the competitive ratio for the cow path problem with a non-optimal seeker is $2^{E+2} - 1$ if E is even, and $2^{E+2} + 1$ if E is odd. □

3. Improving the Algorithm

Now, we see whether this competitive ratio can be improved by changing the distance function $f_j = 2^{j-1}$. In this distance function we can see the number 2 as the distance factor, which multiplies the previously travelled distance. When the non-optimal seeker is introduced, we expect that the seeker needs to pass the object multiple times before finding it. It then makes sense to let the seeker search in the same places more often. So, a good strategy to improve the linear spiral search could be to make this distance factor smaller.
So, for a more general case, let $f_j = a^{j-1}$, where $1 < a \leq 2$. Here, a needs to be larger than 1, otherwise the algorithm would only cover the interval $[-1, 1]$, which would cause some difficulties when searching for the object. With this general distance function we can make some claims about the competitive ratio of the algorithm.

Theorem 3.1. *The competitive ratio of linear spiral search with a non-optimal seeker and a general distance function $f_{j+1} = a^j$ is $\frac{2a^{E+1}}{a-1} - 1$ if E is even and $\frac{2a^{E+1}}{a-1} + 1$ if E is odd.*

Proof. In the worst case, the distance from the origin to the objective d is $f_i + \epsilon$, where $f_i = a^{i-1}$ and $\epsilon > 0$ is some small number. Then, the expected travel distance is:

$$2 \cdot \sum_{j=1}^{i+E} f_j + (-1)^{E+1} d.$$

Thus, there are two cases to be distinguished: E is even and E is odd. If E is even, the expected travel distance is:

$$\begin{aligned} 2 \cdot \sum_{j=1}^{i+E} f_j - d &= 2 \cdot \sum_{j=0}^{i+E-1} a^j - d \\ &= 2 \cdot (\frac{a^{i+E} - 1}{a-1}) - a^{i-1} - \epsilon \\ &= (\frac{2a^{E+1}}{a-1} - 1) a^{i-1} - \frac{2}{a-1} - \epsilon \\ &\leq (\frac{2a^{E+1}}{a-1} - 1)(f_i + \epsilon) \\ &= (\frac{2a^{E+1}}{a-1} - 1) \cdot d. \end{aligned}$$

If E is odd, the expected travel distance is:

$$\begin{aligned} 2 \cdot \sum_{j=1}^{i+E} f_j + d &= 2 \cdot \sum_{j=0}^{i+E-1} a^j + d \\ &= 2 \cdot (\frac{a^{i+E} - 1}{a-1}) + a^{i-1} + \epsilon \\ &= (\frac{2a^{E+1}}{a-1} + 1) a^{i-1} - \frac{2}{a-1} + \epsilon \\ &\leq (\frac{2a^{E+1}}{a-1} + 1)(f_i + \epsilon) \\ &= (\frac{2a^{E+1}}{a-1} + 1) \cdot d. \end{aligned}$$

So, the competitive ratio for the cow path problem with a non-optimal seeker is $\frac{2a^{E+1}}{a-1} - 1$ if E is even, and $\frac{2a^{E+1}}{a-1} + 1$ if E is odd. □

Now we have the competitive ratios for the general distance function $f_j = a^{j-1}$. To improve this algorithm we need to choose a such that the competitive ratios are minimal. So we need to minimize $\frac{2a^{E+1}}{a-1}$ with respect to a.

Theorem 3.2. *For linear spiral search with a non-optimal seeker and general distance function $f_{j+1} = a^j$ the competitive ratio is optimal when $a = 1 + p$.*

Proof. The function $g(a) := \frac{2a^{E+1}}{a-1}$ needs to be minimized to obtain an optimal competitive ratio for this algorithm. Firstly, notice only a local minimum in the domain $1 < a \leq 2$ is necessary

to improve the competitive ratio. Secondly, the function g is a polynomial function, so g is a differentiable real function. Therefore, the critical point can be found by determining where the derivative of g is equal to zero:

$$g'(a) = 2\frac{a^E(Ea - E - 1)}{(a-1)^2} = 0$$

This results in a critical point at $a = 1 + \frac{1}{E} = 1 + p$, which lies inside the domain $1 < a \leq 2$, since $E \geq 1$.
Looking at the derivative of g right before the critical point, we see that it is negative there. Just after the critical point the derivative is positive. Therefore, the function g has a local minimum at $a = 1 + p$. So, linear spiral search with a non-optimal seeker and general distance function $f_{j+1=a^j}$ has an optimal competitive ratio when $a = 1 + p$. □

4. More Than The Expected Case

In this section we will expand the results of the previous chapters. Up until now we have only considered the expected time of finding E, but what if we wanted to know how the algorithm performs in at least 95% of the cases? Then, we need to take a look at the cumulative distribution function of the geometric distribution used for this non-optimal seeker. We know that the cumulative distribution function of a geometric distributed random variable X with probability p of success is:

$$F(X) = P(X \leq t) = 1 - (1-p)^t.$$

So, if we want to know the worst case search time for 95% of the cases, we deduce for what t the cumulative distribution function is equal to 95%. We denote that value as $t_{95\%}$. This value gives us the amount of times the seeker walks past the object before finding it in the worst case, when we look at 95% of the cases. For a general probability p of finding the object, we have:

$$t_{95\%} = \lceil \frac{\log 0.05}{\log(1-p)} \rceil.$$

We use the ceiling here, so that $t_{95\%} \in \mathbb{N}$. This is necessary, since the seeker can only walk an integral amount of times past the object. This is why earlier we said that we wanted to know how the algorithm performs for *at least* 95% of the time. Because when we are using the ceiling, we are actually gaining an upper bound for a bit more than 95% of the cases.
Now we have $t_{95\%}$ we can substitute this in the competitive ratios given by Theorem 3.1. If $t_{95\%}$ is even, we have a competitive ratio of:

$$\frac{2a^{t_{95\%}+1}}{a-1} - 1.$$

And if $t_{95\%}$ is odd, we have a competitive ratio of:

$$\frac{2a^{t_{95\%}+1}}{a-1} + 1.$$

5. Some Examples

To get some concrete numbers for different probabilities, we are now going to apply the previously presented results to some example distributions. Recall that we have a geometric distribution with

probability p of the seeker finding the object when he is upon it. In Table 1, the expected time of finding E, the optimal distance factor a, the expected competitive ratio $CR_{expected}$, the worst case search time for 95% of the cases $t_{95\%}$, and the competitive ratio for 95% of the cases $CR_{95\%}$ can be found for some example probabilities p.

p	E	a	$CR_{expected}$	$t_{95\%}$	$CR_{95\%}$
1	1	2	9	-	-
$\frac{3}{4}$	2	$1\frac{3}{4}$	13.29	3	26.01
$\frac{2}{3}$	2	$1\frac{2}{3}$	12.89	3	24.15
$\frac{1}{2}$	2	$1\frac{1}{2}$	12.5	5	46.56
$\frac{1}{3}$	3	$1\frac{1}{3}$	19.96	8	78.91
$\frac{1}{4}$	4	$1\frac{1}{4}$	23.41	11	117.42

Table 1: *The expected competitive ratio and the competitive ratio for 95% of the cases for some example probabilities p.*

If we look at the probabilities with an integral expected time of finding E, we see, as expected, that the competitive ratios increase when the probability of finding the object decreases. For the probabilities $\frac{3}{4}$ and $\frac{2}{3}$ this relation does not hold. This can be explained by the fact that the expected time of finding E for these cases is not an integer. Therefore, the ceiling of the expected time of finding the object is used. In the table these are denoted as E, since the seeker can only search in discrete steps.
When looking at the difference in the competitive ratio for the expected case and the competitive ratio for 95% of the cases, we see that the difference is almost a factor 4. This does not hold for the probabilities $\frac{3}{4}$ and $\frac{2}{3}$, but this can be explained by the fact that for these probabilities the ceiling of E is used. This difference in competitive ratios implies that the distribution of p has a large standard deviation, which is correct, since it is a geometric distribution. For a geometric distribution the standard deviation is given as $\sqrt{\frac{1-p}{p^2}} = \frac{\sqrt{1-p}}{p}$, which, when compared to the mean $E = \frac{1}{p}$, is not that much smaller when p is small.

6. Expanding the Search Space

Here, we are going to take a look at a larger search space. Up until now the object was hidden on a line, but what if the object is hidden somewhere on one of m concurrent rays? Then the problem changes to a *star search*. To solve the normal star search *general linear spiral search* is used. The distance function is $f_j = \frac{m}{m-1}^{j-1}, \forall\, j \geq 1$, and the competitive ratio is $1 + 2\frac{m^m}{(m-1)^{m-1}}$ for large m (**?**). We are going to expand this competitive ratio for a non-optimal seeker.

Theorem 6.1. *The competitive ratio of a star search on m concurrent rays with a non-optimal seeker and distance function $f_j = (\frac{m}{m-1})^{j-1}$ is $\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1})-1} - 1$ if E is even and $\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1})-1} + 1$ if E is odd.*

Proof. In the worst case, the distance from the origin to the objective d is $f_i + \epsilon$, where $f_i = (\frac{m}{m-1})^{i-1}$

and $\epsilon > 0$ is some small number. Then, the expected travel distance is:

$$2 \cdot \sum_{j=1}^{i+E} f_j + (-1)^{E+1} d.$$

Thus, there are two cases to be distinguished: E is even and E is odd. If E is even, the expected travel distance is:

$$\begin{aligned}
2 \cdot \sum_{j=1}^{i+E} f_j - d &= 2 \cdot \sum_{j=0}^{i+E-1} (\frac{m}{m-1})^j - d \\
&= 2 \cdot (\frac{(\frac{m}{m-1})^{i+E} - 1}{(\frac{m}{m-1}) - 1}) - (\frac{m}{m-1})^{i-1} - \epsilon \\
&= (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} - 1)(\frac{m}{m-1})^{i-1} - \frac{2}{(\frac{m}{m-1}) - 1} - \epsilon \\
&\leq (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} - 1)(f_i + \epsilon) \\
&= (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} - 1) \cdot d.
\end{aligned}$$

If E is odd, the expected travel distance is:

$$\begin{aligned}
2 \cdot \sum_{j=1}^{i+E} f_j + d &= 2 \cdot \sum_{j=0}^{i+E-1} (\frac{m}{m-1})^j + d \\
&= 2 \cdot (\frac{(\frac{m}{m-1})^{i+E} - 1}{(\frac{m}{m-1}) - 1}) + (\frac{m}{m-1})^{i-1} + \epsilon \\
&= (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} + 1)(\frac{m}{m-1})^{i-1} - \frac{2}{(\frac{m}{m-1}) - 1} + \epsilon \\
&\leq (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} + 1)(f_i + \epsilon) \\
&= (\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1}) - 1} + 1) \cdot d.
\end{aligned}$$

So, the competitive ratio for the cow path problem with a non-optimal seeker is $\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1})-1} - 1$ if E is even, and $\frac{2(\frac{m}{m-1})^{E+1}}{(\frac{m}{m-1})-1} + 1$ if E is odd. □

7. Conclusions

The main topic of this paper is the influence of introducing a non-optimal seeker into the cow path problem. We saw that the competitive ratio of linear spiral search increased when this non-optimal seeker was introduced. This increase is dependent on the expected time E of finding the object. When E is even, the competitive ratio is $2^{E+2} - 1$, and when E is odd, the competitive ratio is $2^{E+2} + 1$.

These competitive ratios can be improved by adjusting the algorithm. The algorithm is adjusted

by making the distance factor dependent on the probability p of finding the object. This results in a competitive ratio of $\frac{2a^{E+1}}{a-1} + 1$ when E is odd, and $\frac{2a^{E+1}}{a-1} - 1$ when E is even.
These competitive ratios can be changed to be valid for at least 95% of the cases. When looking at the difference between the competitive ratio for the expected case and for 95% of the cases, it is evident that there is a major difference. The competitive ratio for 95% of the cases was almost 4 times the competitive ratio for the expected case. This implies that the distribution of p has a large standard deviation, which is the case for a geometric distribution.
In conclusion, when a non-optimal seeker is introduced in the cow path problem, the competitive ratio of linear spiral search increases. By then reducing the step size of the algorithm, the competitive ratio can be improved quite a lot.

Further Research. In this research only the linear spiral search algorithm was considered. There might be other algorithms that perform better with this instance of the cow path problem, such as the randomized algorithm *SmartCow* (**?**). This algorithm gives an improvement of almost a factor 2 for the normal cow path problem.
This paper focused mostly on the cow path problem or linear search problem. Only one section was dedicated to a star search. This is an area where it could be interesting to research it further. Also, other search problems, such as searching a lattice or searching in a plane could be interesting extensions of searching with imperfect information.