A ML-inspired Approach to Symmetric Rendezvous Problem on the Line

Stéphane Liem Nguyen
{stephane.nguyen@etu.unige.ch}

Switzerland, University of Geneva

A new method for the symmetric rendezvous problem on the line (SRPL), inspired by reinforcement learning, provides better results than random strategies.

A large variety of situations, such as emergencies, require humans or objects to meet in the least expected amount of time. This can happen in environments ranging from real life to mathematical structures. Entities involved in these tasks need strategies to inform them how they should move as close as possible to optimal in their environment.

This mentioned optimality or sub-optimality of the strategies depends on many factors which the entities involved can or cannot control. For this reason, instead of having a lot of entities living in some complex environment, our work is mainly focused on the problem of finding a common strategy for two entities living on an infinite line that should meet in the least expected amount of time.

Deciding the least expected meeting time in the SRPL is still an open problem in the field of rendezvous search since 1995 [1], while Alpern solved the asymmetric problem in the same year. The SRPL is a problem where two blind agents, living on a line, want to meet as fast as possible in expectation while having the same strategy[1] telling them how to move at any specific instant. The agents are at an initial distance of 2 apart but don't know if the other agent is on its left or its right. Their velocities are also bounded by 1.

To the best of our knowledge, most previous works used methods that could become extremely expensive computationally [2–4]. For instance, Han et al. [4] mentioned in their conclusion that their computational resources limited them. In our work, we try a new approach inspired by reinforcement learning, but that still does not resolve the computational resources problem. However, our results show that the strategies we obtained are better than random strategies, which is optimistic news.

[1] This is why "symmetric" appears in the name of the problem.

Intuitively, our approach consists in reinforcing paths that lead to meeting. To achieve this, we sample each time a pair of trajectories where both agents meet at their end, and then we increase the probabilities of repeating that pair. Note that because agents do not know whether the other agent is on its left or right, we also need to increase the probabilities of the flipped trajectories. We change the probabilities by adding some value to a component of each probability vector that the trajectories affected and projecting the modified vector back onto the space of probability vectors using the projected gradient method [5].

During each time unit, our agents only either move at a maximum speed of 1 to their left or their right or do not move.

Strategies allowed are only those that let the agents move freely (initialized uniformly) for some time and then force them to go back to their initial positions if they can't move fast enough to reach their initial positions before or at a time that is a multiple of N. We then repeat the process until both agents meet (see Figure 1). When the agents meet, the episode ends and we increase the probabilities. For this, we tried two methods, the first one only uses the last repetition and the second one uses all the repetitions (the information from the whole episode).

For the first method, the value added to a component of an affected probability vector is inversely proportional to the number of times both agents picked the particular action from their positions relative to their initial positions at this time unit modulo N.

For the second method, we took inspiration from REINFORCE algorithm [7] but clipped the gradient of the logarithm of the strategy directly so that the updates to the probabilities are not too important.

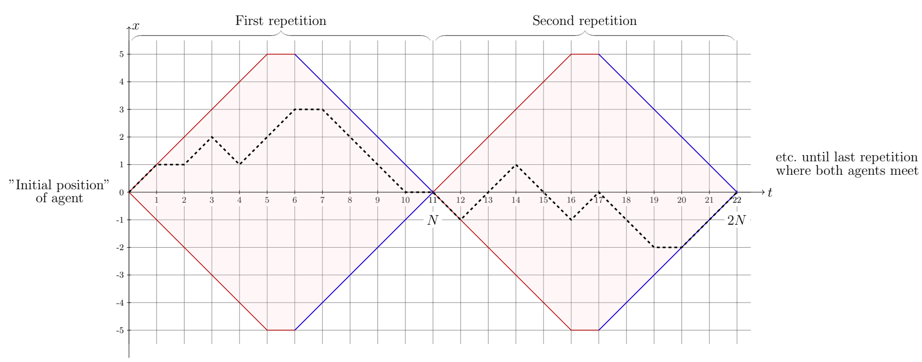

Fig. 1. Example of a trajectory of one of the agents relative to its initial position for $N = 11$. This figure only shows the two first repetitions. Agents always only move on the line. Black: trajectory/position relative to its initial position with respect to time, red: possible relative positions of agents at a given time, blue: forcing agent to go back to its initial position.

Table 1 summarizes the results.

Table 1. Expected meeting times of: our strategies vs Han et al.'s strategies [4]

N	First method	REINFORCE-like	Han et al. [4]
2	7.0006	8.3785	7.0000
3	10.0004	11.1097	5.0000
4	6.0004	8.5138	5.9441
5	8.0005	–	4.8827
6	5.7638	–	4.4634

Our first method gave better expected meeting times than the REINFORCE-like algorithm but worse than in Han et al. [4] because we work in a different search space where we bounded the positions of agents relative to their initial positions, meaning that a lot of strategies are not reachable except if we let agents free to move for a longer, potentially infinite, period of time N. However, strategies obtained have similar shapes as those in the work of Han et al. [4].

The strategies we obtained in the SRPL were not as good as the strategies obtained in the work of Han et al. [4] and did not solve their computational resources issue. However, our work demonstrates that strategies better than random are learnable even if agents are completely blind.

Our next step would be to improve our strategies to either reach similar or better results than in Han et al. [4]. One promising research direction would be to use parameterized function approximations with a fixed number of parameters to define the strategy instead of our current tabular-like method.

Note

This work is a summary of my bachelor project [6] supervised by Prof. Pierre Leone. This text contains 820 words without taking into account this section, the title, the references, the figure, the table and the captions.

References

1. Alpern, S. The Rendezvous Search Problem. *SIAM Journal on Control and Optimization 33*, 3 (May 1995), 673–683. Publisher: Society for Industrial and Applied Mathematics.
2. Anderson, E. J., and Essegaier, S. Rendezvous Search on the Line with Indistinguishable Players. *SIAM Journal on Control and Optimization 33*, 6 (Nov. 1995), 1637–1642. Publisher: Society for Industrial and Applied Mathematics.
3. Baston, V. Note: Two rendezvous search problems on the line. *Naval Research Logistics (NRL) 46*, 3 (1999), 335–340.

4. Han, Q., Du, D., Vera, J., and Zuluaga, L. F. Improved Bounds for the Symmetric Rendezvous Value on the Line. *Operations Research 56*, 3 (June 2008), 772–782.
5. Luenberger, D. G., and Ye, Y. *Linear and Nonlinear Programming*, 4th ed. Springer, July 2015.
6. Nguyen, S. L., and Leone, P. Mise en oeuvre de méthodes d'apprentissage pour le problème de rendez-vous symétrique sur la ligne. [Unpublished BSc project], 2021.
7. Sutton, R. S., and Barto, A. G. *Reinforcement learning: An Introduction*, second ed. Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts, 2018.