EDITORIAL

Reviews and Responses for
Sampling-Based Aircraft Path Planning with Soft Actor-Critic

Authors: Dieudonne Groot, Joost Ellerbroek, and Jacco Hoekstra

Reviewers: Sameer Alam, Arsyi Aziz, and Filipo Perotto

Editor: Tatiana Polishchuk

1. Original paper

DOI for the original paper: https://doi.org/10.59490/joas.2025.7875

2. Review - round 1

2.1 Reviewer 1

This work adopts Soft Actor-Critic (SAC) RK algorithm for aircraft path planning within continuous airspace, taking into account the linear combination of fuel consumption and noise-related costs.

Bluesky air traffic simulator is used for simulation. A simplified representation of the Dutch airspace, i.e., a circular border with a radius of 150NM and centred at Schiphol airport, is used in the simulation environment for both training and evaluation. The noise emissions are calculated using population data obtained from Eurostat.

The Dijkstra algorithm is used as the baseline algorithm to generate optimal single-agent paths in discrete space. The proposed approach (SAC) outperforms Dijkstra by generating flight path with lower both noise cost, fuel cost and number of turns.

Then based on the learnt distribution of action, random sampling is applied to generate several paths. The experimental results show an improvement in performance by investigating and selecting the final path from that ensemble set.

My main concerns for this study are as follows:

1. It seems that this study is based on free routing concept, but still consider/mention waypoints. What are the waypoints considered in this study? Are they pre-defined or generated by each algorithm?
2. Can author further clarify the definition of turn in case of Dijkstra and SAC? Why are they different from each other? Can an unique definition is used for turn in both cases?
3. Is the state only contains the aircraft position sx, sy? If yes, the model was overfitted to the given airspace, i.e., population distribution. It significantly limits the generalization of the proposed approach by considering other inputs.
4. For single-agent path finding in continuous space, which is the state-of-the-art algorithm? By applying Dijkstra algorithm in this problem, some assumption on how the airspace can be discretized and how the graph can be constructed which can be the source of inefficiency of the

algorithm. It means, by a better way to construct the graph, not just with neighbouring cells, the performance of Dijkstra will be significantly improved.

5. No uncertainty and traffic/interactions with other flights are considered in this study. As the problem complexity will be significantly increased in the dynamic environment. This study can't justify the potential of using SAC or RL as a high-level path planner for HRL based multi-agent path planner.

6. Overall, the assumptions used in the study renders the contribution very limited.

2.2   Reviewer 2

The paper provides an intriguing exploration of hierarchical reinforcement learning applied to aircraft path planning, particularly focusing on implementing a high-level path planning using the soft actor-critic (SAC) algorithm. Unlike previous research that primarily focuses on tactical deconfliction models, this paper takes a different approach by concentrating on single path planning while incorporating constraints such as fuel consumption and noise pollution. Notably, the results demonstrate that the continuous policy trained under SAC outperforms the conventional discrete-space Dijkstra path planning method in scenarios where noise pollution takes precedence over fuel costs, such as in densely populated urban areas.

Although well written, I would like to suggest a few corrections:

1. **Equation Formatting**: The equations could be improved by incorporating proper punctuation. Additionally, it would be beneficial to integrate the equations more seamlessly into the narrative of the paragraphs wherever possible, to enhance readability and coherence.

2. **Clarity in Variables**: In equation (3), the variables n and m should be elaborated, as their meanings are unclear. Moreover, n could easily be confused with n_t, which denotes noise at time step t . If n and n_t represent distinct concepts, I recommend using different notations to avoid ambiguity.

3. **Soft Actor-Critic Description**: The description of SAC as having three networks (lines 166-167) is partially accurate but slightly misleading. SAC employs two Q-functions and a policy network, rather than the critic, value, and policy networks described in the paper. This distinction should be clarified to accurately represent the algorithm.

Additionally, I suggest that future research consider exploring path planning in three-dimensional space. Incorporating altitude into the model could capture its influence on noise production and fuel consumption, potentially leading to more robust and realistic planning frameworks.

2.3   Reviewer 3

In this article, the authors propose the use of RL, in particular the SAC algorithm, for path planning in an air traffic management context. A particular emphasis is given to the fact that a good trajectory must solve a multi-criteria and constrained optimization. The proposed experimental setting involves a trade-off between fuel costs and noisy-related costs. The authors initially compare SAC with a slightly modified version of classical path planning over graphs (Dijsktra), showing that SAC can outperform path planning, and attributing that advantage to the fact that SAC solves the problem directly in the continuous space. Methodologically, the authors propose an interesting usage of the learned police of actions, which outputs a gaussian distribution for the chosen actions on the continuous domain. It is used to sample different trajectories, and the experimental results indicate that it can improve the solution utility compared to the deterministic use of the mean.

The paper is clear, and the authors take time to explain every element necessary to understand the method, the experiments, the aeronautical context, and their contribution. From section 4, some phrases could be re-written, and some typos and grammar errors fixed (some suggestions at the end of my commentary).

For the final version, I think that the Hierarchical RL and multi-agent conjectures should not be in the abstract, since they are not developed in this paper, even if, as the authors said, this paper is part of a bigger project. In any case, the paper must be self-contained. That kind of conjectures could be better placed in the conclusion.

Technically, I was surprised that the airplane heading angle is not part of the input (i.e. not part of the observation), but just the position x and y. The cost of choosing an action (dx, dy) highly depends on that information. When reading the Section 3.2.1 Input representation, I also imagined the possibility of using a radial representation. The authors talk about that in the conclusion as future work. In the Section 3.2.2 Action Representation, the authors define that the action corresponds to a Cartesian displacement limited at a maximum distance of 50Km. It is only in the Conclusion that it is said that there is also a minimal distance of 1Km, very important point that should be declared before.

In the comparison, the problem is that the squared discretization used for Dijkstra can be the main handicap for that method. A smarter distribution of possible waypoints and neighbors certainly will increase the quality of that solution, of course, with a cost for increasing neighborhood. But, for example, I would like to see the result for the same discrete points but allowing to pass directly from one cell to another 2 ou 3 steps far away, i.e. from (i,j) to (i+1 or i-1 or i+2 or i-2, j+1 or j-1 or j+2, or j-2) etc. Or, instead of placing just a waypoint on the center of the cell, placing an additional waypoint in a lower population density of that cell.

Finally, a comparison with the existent airways, using the currently existent waypoints of the non-free-route-airspace, could be also interesting, even if they will result in more saturated paths.

Minor suggestions:

l.72 "during during"

alg 1 "model(S_i)" –> "model(s_i)" ?

l.278 purpose for this study –> purpose of this study; all methods compared –> all compared methods

l.279 Future studies on this subject however should investigate this and evaluate this impact –> Future studies should however investigate and evaluate its impact.

l.294 included for this metric –> included in this metric

l.296 the final indicator –> the last indicator

l.306 This because (?)

l.312 given in figure –> given in the Figure

l.319 the paths ... has a higher –> have a higher

tables –> maybe transform it in bar charts ?

l.403 it is important that the network architecture used is carefully considered –> the choice of network architecture must be carefully considered

l.414 a results of –> a consequence of

l.459 the costs... is normally distributed –> are

3. Response - round 1

3.1   Responses to Reviewer 1

"*It seems that this study is based on free routing concept, but still consider/mention waypoints. What are the waypoints considered in this study? Are they pre-defined or generated by each algorithm?*"

> **Response**
>
> In this study, a waypoint is defined as an intermediate freely chosen target location, which is defined by lat,lon coordinates. To ensure that this is more clear to the reader, a description of what a waypoint is for the context of this study has been added to the section that first mentions 'waypoints'.

"*Can author further clarify the definition of turn in case of Dijkstra and SAC? Why are they different from each other? Can an unique definition is used for turn in both cases?*"

> **Response**
>
> The definition is now reformulated to be any required change in heading, this does not change the original definition, but hopefully makes it more clear to the reader.

"*Is the state only contains the aircraft position sx, sy? If yes, the model was overfitted to the given airspace, i.e., population distribution. It significantly limits the generalization of the proposed approach by considering other inputs.*"

> **Response**
>
> That is correct, because this study was a global planning task, the entire environment is known upfront, similarly to gridworld examples where only x and y are given as states. The main benefit from this representation is that the output of the model also directly serves as the input, allowing for fast recursive generation of paths from any state in the environment. This reasoning is now also added to the paper, to clarify for future readers.
>
> For different environments the model should be retrained, but this is not too different from methods such as RRT(*) which also do not use any information from the environment. It is true that more complex state representations would potentially allow for better performance / generalization, however for the goal of this study, this state representation was explicitly chosen.
>
> We do understand the concerns of the reviewer, and agree that this could be viewed as overfitting, however, because it is impossible to sample the entire continuous domain of coordinates, an argument could also be made that the model learned to interpolate between the observed samples, allowing for an efficient representation of the population map within the neural network.

"*For single-agent path finding in continuous space, which is the state-of-the-art algorithm? By applying Dijkstra algorithm in this problem, some assumption on how the airspace can be discretized and how the graph can be constructed which can be the source of inefficiency of the algorithm. It means, by a better way to construct the graph, not just with neighbouring cells, the performance of Dijkstra will be significantly improved.*"

> **Response**
>
> We agree with this statement, and have therefore changed the way the graph is constructed for Dijkstra to include more edges from a given cell. Initially little effort was given to the Dijkstra implementation because the main focus of the study was to compare mean-based with sampling-based SAC. However, considering reviewer comments, we have improved the Dijkstra implementation. The new results improve Dijkstra, however, they do not alter the main conclusion of the paper.

"No uncertainty and traffic/interactions with other flights are considered in this study. As the problem complexity will be significantly increased in the dynamic environment. This study can't justify the potential of using SAC or RL as a high-level path planner for HRL based multi-agent path planner."

> **Response**
>
> We agree, initially the HRL part was included because this paper is part of a larger study, however, as other reviewers have rightfully pointed out, a paper should be stand-alone. We have therefore limited the mentioning of HRL / multi-agent systems and instead focus more on the main contributions in the revised manuscript.

"Overall, the assumptions used in the study renders the contribution very limited."

> **Response**
>
> We have implemented your and the other reviewers feedback to clarify the scope and the goal of the study, as well as rerunning some experiments to strengthen the Dijkstra baseline. By doing so and rewriting parts of the paper, the contribution of the paper, showing that improvements to SAC as a (path planning) method can be obtained by altering the evaluation strategy, has been made more explicit. Many thanks for the good advice.

3.2   Responses to Reviewer 2

"The equations could be improved by incorporating proper punctuation. Additionally, it would be beneficial to integrate the equations more seamlessly into the narrative of the paragraphs wherever possible, to enhance readability and coherence."

> **Response**
>
> Referencing to the equations in text has been improved, and some redundant equations, such as summing have been removed to enhance readability. We, however, are unclear what is meant with incorporating proper punctuation. If this is still a point of concern in the revised manuscript we are more than willing to improve this if given more instructions.

"In equation (3), the variables n and m should be elaborated, as their meanings are unclear. Moreover, n could easily be confused with n_t , which denotes noise at time step t . If n and n_t represent distinct concepts, I recommend using different notations to avoid ambiguity."

> **Response**
>
> We agree with the ambiguity of using n & m for the sums. The equation is now rewritten in the form of a double sum with 'i' and 'j' iterating from zero till 'i_{max}' and 'j_{max}' respectively. We hope this clarifies the notations.

"The description of SAC as having three networks (lines 166-167) is partially accurate but slightly misleading. SAC employs two Q-functions and a policy network, rather than the critic, value, and policy networks described in the paper. This distinction should be clarified to accurately represent the algorithm."

Response

While this is true for many implementations of SAC found online, the original paper mentions the usage of both a State-Action value function (Q) and a State value function (V). The authors of the original SAC paper do mention that in principle there is no need to implement a separate State value function as it is related to Q and PI, however, including this function V can stabilize training [Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor]. In a later paper the same authors indeed identify that the State value function can be dropped, and instead be approximated using Q and PI, resulting in more light weight training [Soft Actor-Critic Algorithms and Applications] For this study it was chosen to use the implementation that includes V as it would also allow visualization of the value map, without requiring actions to be sampled from PI, creating a heatmap of values over the entire state space, which could be used to enhance explainability. This benefit however was dropped from the paper due to not fitting in the scope.

3.3   Response to Reviewer 3

"*Technically, I was surprised that the airplane heading angle is not part of the input (i.e. not part of the observation), but just the position x and y. The cost of choosing an action (dx, dy) highly depends on that information.* "

Response

We agree that this might enhance the performance of the model, and that indeed the cost of an action does depend on the current heading of the aircraft. However, the simplified implementation was chosen to more closely resemble the state representation used in gridworld environments and to allow for recursive generation of paths by simply adding the 'action' to the state to get the new state. This would then allow for paths to be generated without requiring intermediate simulation of the state transitions. Additionally, because of the convergent nature of the solution space (e.g. paths all converge into multiple streams), later during training the heading is largely correlated for a given state and therefore implied. We have updated the section on the state description to better clarify the reasoning behind the state representation.

"*In the Section 3.2.2 Action Representation, the authors define that the action corresponds to a Cartesian displacement limited at a maximum distance of 50Km. It is only in the Conclusion that it is said that there is also a minimal distance of 1Km, very important point that should be declared before.*"

Response

This was a mistake of wording, we meant to use 1km not as a minimal distance, but as an example that in theory very small actions could be used, but in practice, large waypoint steps were preferred by the method. This wording has been changed to infinitesimal distance in the hope to clarify this misunderstanding.

" *I think that the Hierarchical RL and multi-agent conjectures should not be in the abstract, since they are not developed in this paper, even if, as the authors said, this paper is part of a bigger project. In any case, the paper must be self-contained. That kind of conjectures could be better placed in the conclusion.*"

Response

We agree, and thank the reviewer for this comment, the paper has been rewritten to better fit the scope and to be more self-contained, and we belief it has let to a better paper overall. Hopefully these changes better reflect the contributions of the paper.

"*A smarter distribution of possible waypoints and neighbors certainly will increase the quality of that solution, of course, with a cost for increasing neighborhood. But, for example, I would like to see the result for the same discrete points but allowing to pass directly from one cell to another 2 ou 3 steps far away, i.e. from (i,j) to (i+1 or i-1 or i+2 or i-2, j+1 or j-1 or j+2, or j-2) etc. Or, instead of placing just a waypoint on the center of the cell, placing an additional waypoint in a lower population density of that cell.*"

> **Response**
>
> We agree with this statement, and have therefore changed the way the graph is constructed for Dijkstra to include more edges from a given cell as is suggested by the reviewer. Initially little effort was given to the Dijkstra implementation because the main focus of the study was to compare mean-based with sampling-based SAC. However, considering your comments, we have improved the Dijkstra implementation based on the provided suggestions. The new results improve Dijkstra, however, they do not alter the main conclusion of the paper.

"*Finally, a comparison with the existent airways, using the currently existent waypoints of the non-free-route-airspace, could be also interesting, even if they will result in more saturated paths.*"

> **Response**
>
> This is indeed a possible field for the results to be compared with, however comparing the results with the current airways was out of the chosen scope of this research.

"Minor suggestions:

l.72 "during during"

alg 1 "model(S_i)" –> "model(s_i)" ?

l.278 purpose for this study –> purpose of this study; all methods compared –> all compared methods

l.279 Future studies on this subject however should investigate this and evaluate this impact –> Future studies should however investigate and evaluate its impact.

l.294 included for this metric –> included in this metric

l.296 the final indicator –> the last indicator

l.306 This because (?)

l.312 given in figure –> given in the Figure

l.319 the paths … has a higher –> have a higher

tables –> maybe transform it in bar charts ?

l.403 it is important that the network architecture used is carefully considered –> the choice of network architecture must be carefully considered

l.414 a results of –> a consequence of

l.459 the costs… is normally distributed –> are"

> **Response**
>
> Thanks for the detailed suggestions! We have adressed all comments mentioned under minor suggestions except for 'tables –> maybe transform it in bar charts?' based on the following reasoning:
>
> Bar charts were discussed initially, but because of the large disparity in costs between paths originating from water bodies and from land it caused a lot of outliers and wide spreads which did not contribute to the legibility. This is why tables were used in the end.
>
> For sampling-based SAC, bar charts were used because the ensembles all started from the same initial bearings per bar, reducing the width of the spread and actually highlighting the fraction of paths improving on the mean-based solutions.

4. Review - round 2

4.1   Reviewer A

My recommendations were addressed, and the revisions have strengthened the paper. Great job!

4.2   Reviewer B

Thanks for this new revised version of your paper.

In my opinion, the authors answered to the main first review points, and the paper can be accepted for publication.

Some typos:

In the abstract: "These three methods; ..." –> "These three methods: ..."

In the introduction: "airline preference" –> "airline preferences"

Paragraph in lines 166-170 : It seems an answer to the reviewer. Please, reformulate for a better integration in the text.