INF8953DE

REINFORCEMENT LEARNING

PRESENTATION

HOW TO COMBINE TREE-SEARCH METHODS IN REINFORCEMENT LEARNING

A REPRODUCIBILITY STUDY

POLYTECHNIQUE MONTRÉAL

UNIVERSITÉ D'INGÉNIERIE

Adam Prévost (1947205)

Alexandre Morinvil (1897222)

Maude Nguyen-The (1843896)

Introduction

- Project track
  - Reproducibility study
  - Yonathan Efroni, Gal Dalal, Bruno Scherrer, and Shie Mannor. How to combine tree-search methods in reinforcement learning, 2019. https://arxiv.org/abs/1809.01843
- Goal of the project
  - Reproduce & validate the paper’s result
  - Confirm the mathematical results empirically
  - Compare the results

Presentation plan

1. Introduction
2. **<u>Presentation plan</u>**
3. Paper summary
4. Experiments
5. Conclusion
6. Q & A

Presentation plan

1. Introduction
2. Presentation plan
3. **Paper summary**
4. Experiments
5. Conclusion
6. Q & A

Paper summary

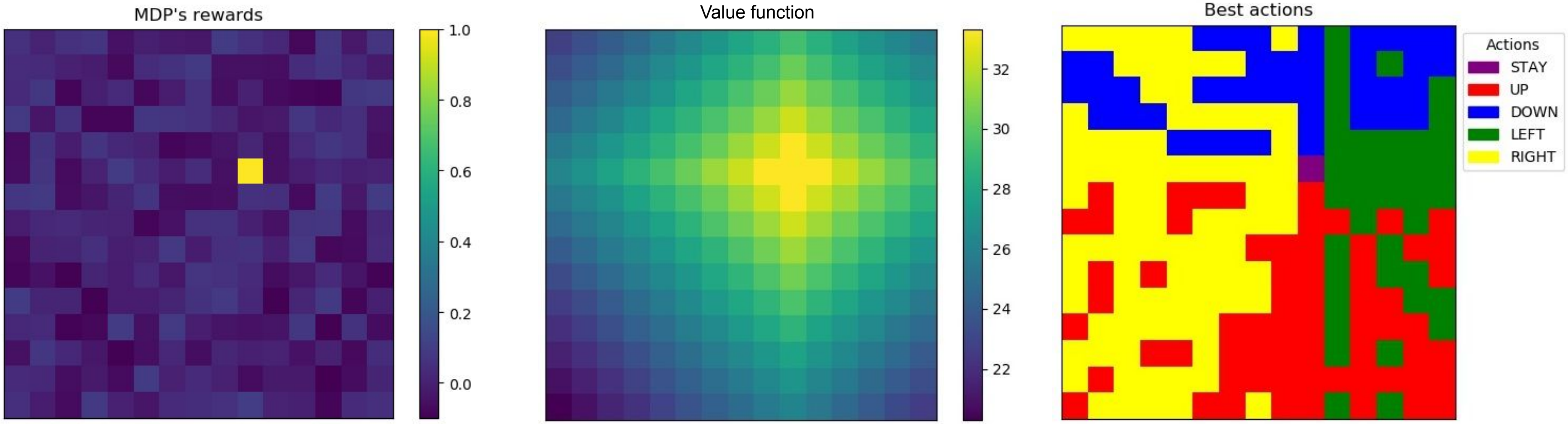
- Problem
    - Lookahead tree search methods don't always contract
    - Introduces h-greedy consistency to prove contraction
- Proposed algorithms
    - hm-PI
    - hλ-PI (no empirical results)

Presentation plan

1. Introduction
2. Presentation plan
3. Paper summary
4. **Experiments**
5. Conclusion
6. Q & A

Experiments: Environement

- 10 by 10 Gridworld
- One state with reward 1 & the rest with rewards drawn from [-0.1, 0.1]
- Actions : STAY, UP, DOWN, LEFT, RIGHT

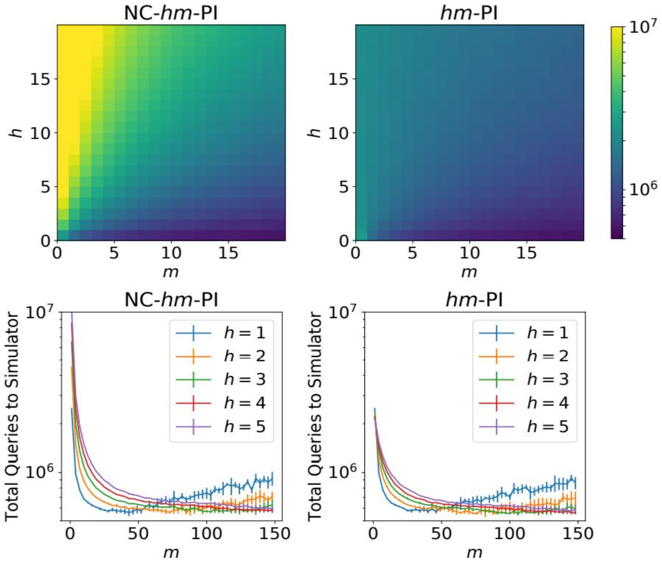


*For example purposes, the figures above are 15 by 15

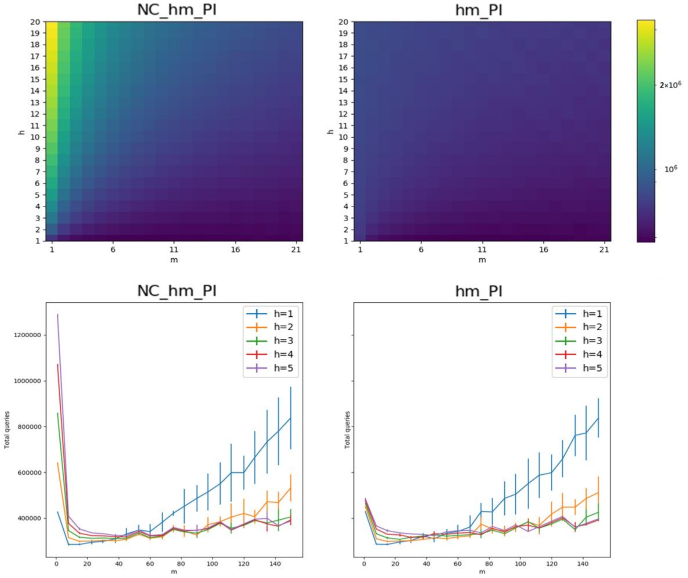
Experiments: Reproducing hm-PI experiment

- NC-hm-PI & hm-PI using the same *h* and *m* hyper parameters
- Total queries to measure the time performance

Paper's results



Experimental results

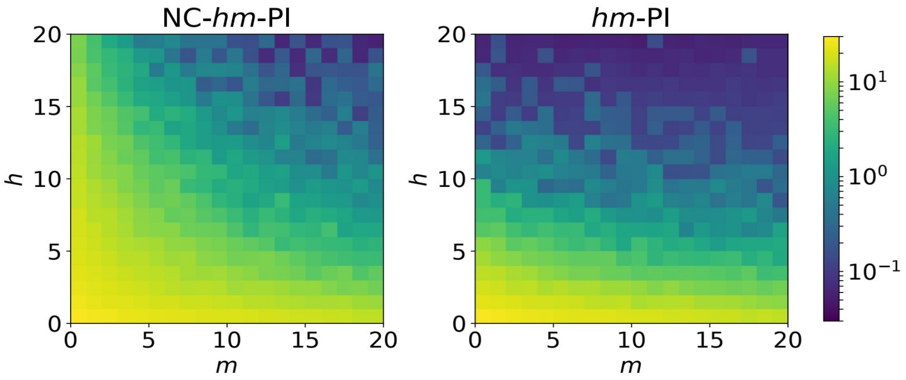


Alexandre Morinvil

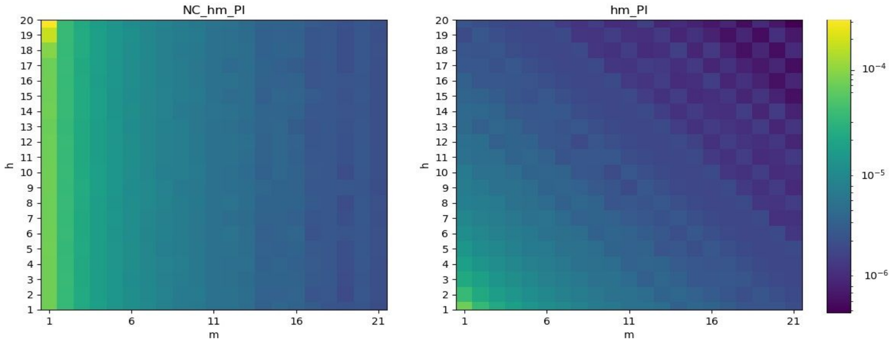
Experiments: Reproducing hm-PI experiments

- NC-hm-PI & hm-PI using the same *h* and *m* hyper parameters
- Distance from optimal value function

Paper's results



Experimental results



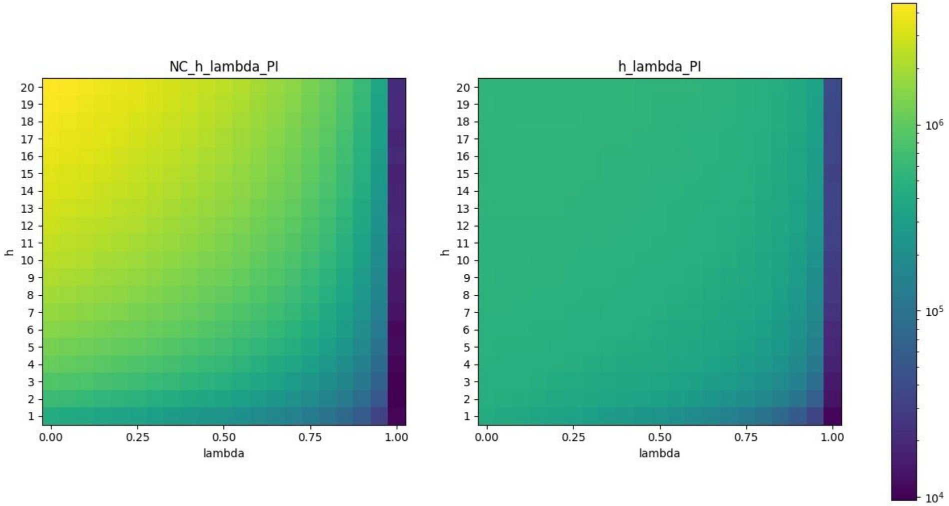
Alexandre Morinvil

Experiments: Exploring results with hλ-PI

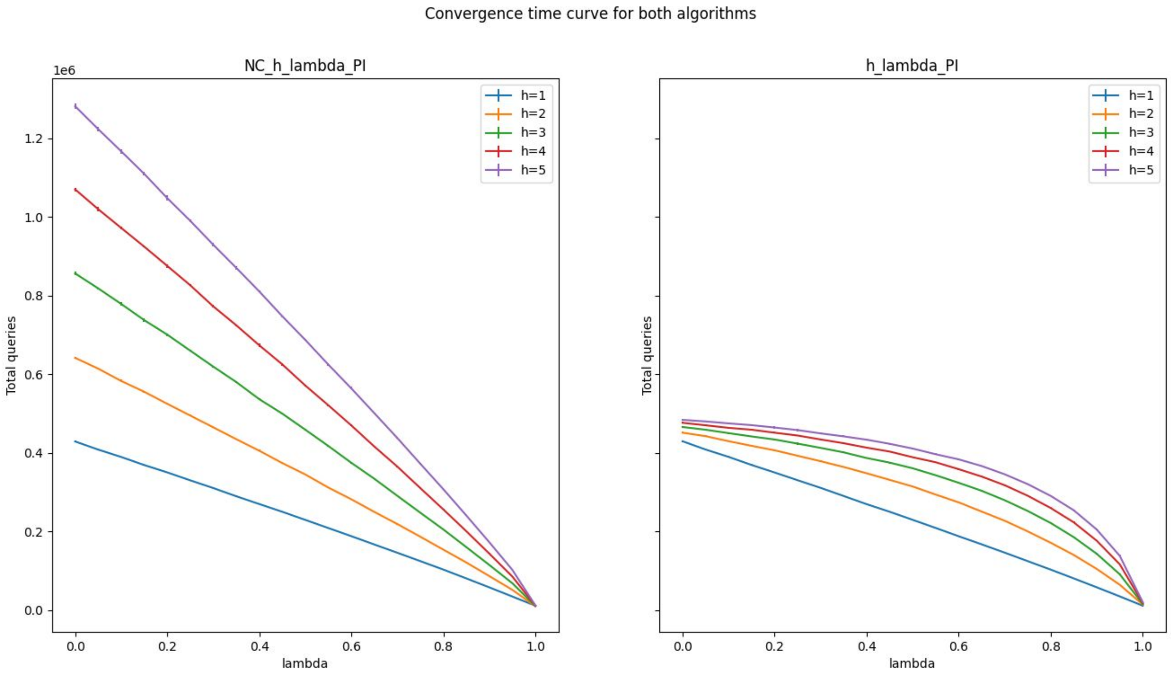
- NC-hλ-PI & hλ-PI using the same *h* and *λ* hyper parameters
- Total queries to measure the time performance

$$T^{\pi}_{\lambda} v \overset{\text{def}}{=} (1-\lambda) \sum_{j=0}^{\infty} \lambda^j (T^{\pi})^{j+1} v$$
$$= v + (I - \gamma \lambda P^{\pi})^{-1} (T^{\pi} v - v).$$

Time of convergence for both algorithms (in number of calls)

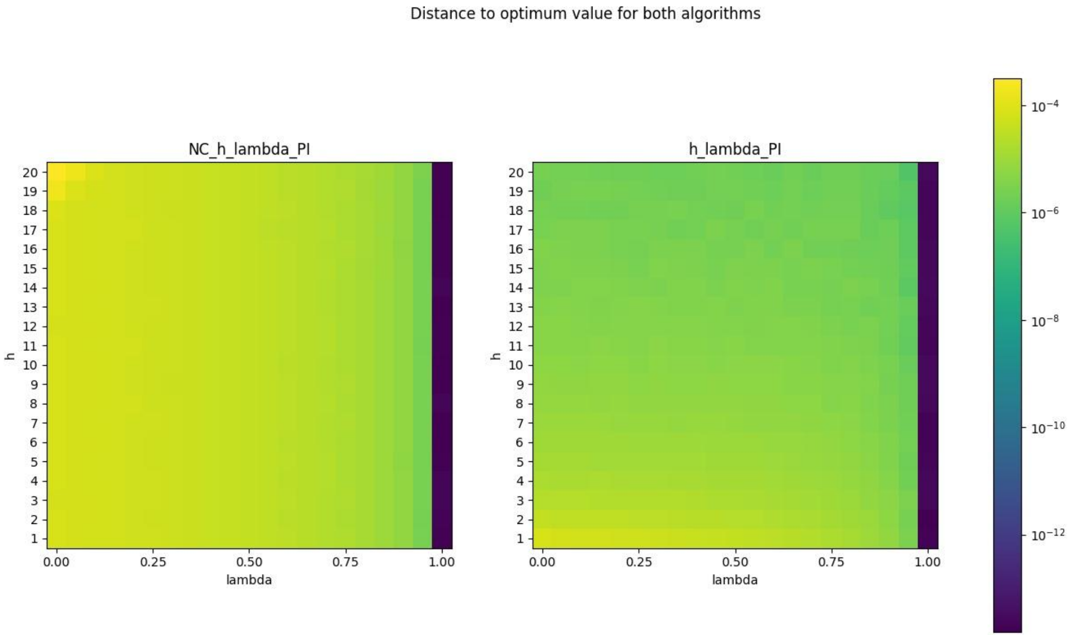


Experiments: Exploring results with hλ-PI



Maude Nguyen-The

Experiments: Exploring results with hλ-PI



Maude Nguyen-The

Presentation plan

1. Introduction
2. Presentation plan
3. Paper summary
4. Experiments
5. **<u>Conclusion</u>**
6. Q & A

Conclusion

- Results are similar
- We can affirm the findings reported in the original work
- Reproducibility issues & caveats
  - Missing parameters (ex: the size of the gridworld)
  - Unspecified procedures (ex: how to count queries *in practice*)
  - Unintuitive notation
- Link to our study and code
  https://github.com/AdamPrevost/INF8953.git

Presentation plan

1. Introduction
2. Presentation plan
3. Paper summary
4. Experiments
5. Conclusion
6. **<u>Q & A</u>**

Q & A