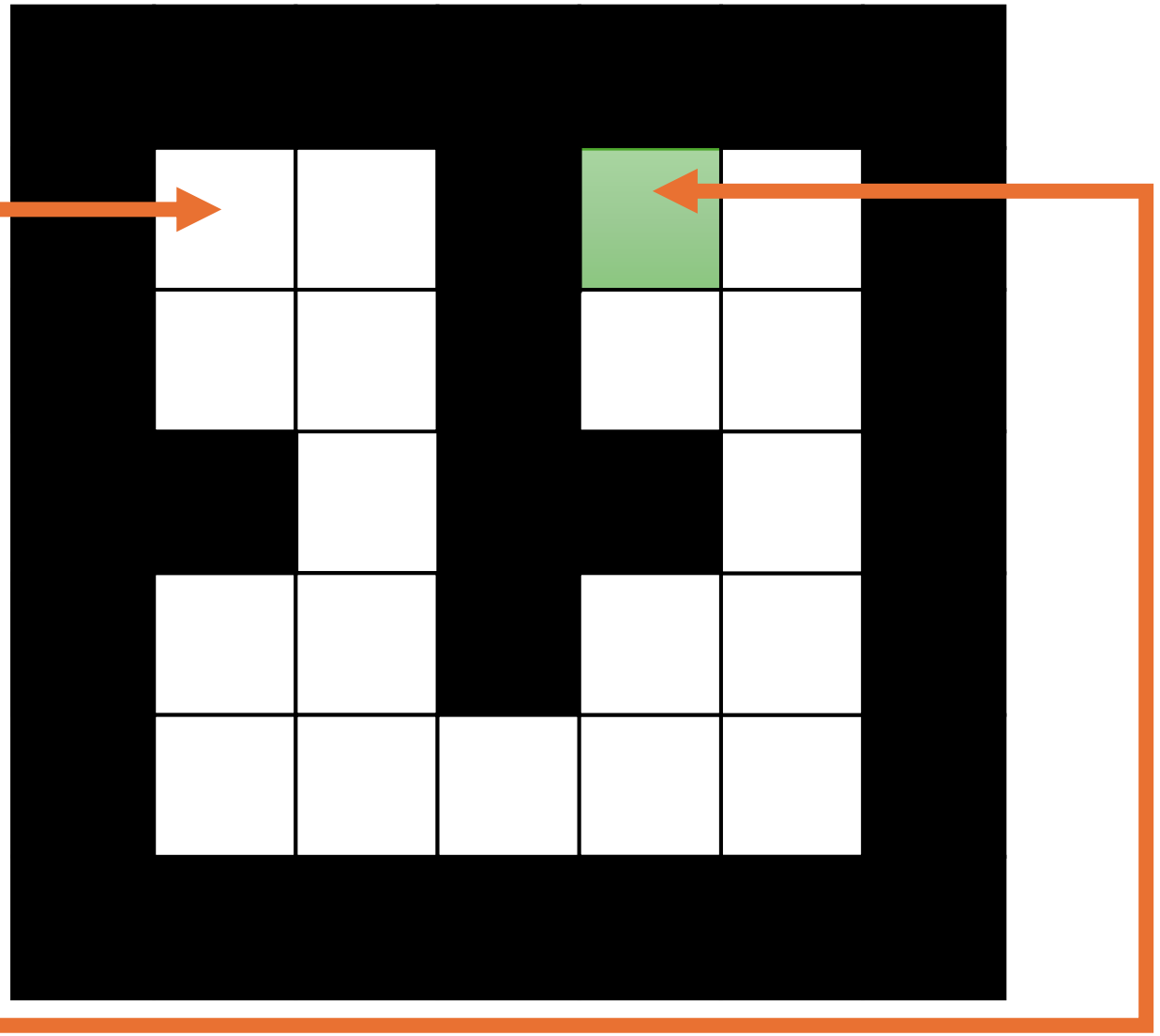
SwiftVI:
Time-Efficient Planning and Learning with MDPs

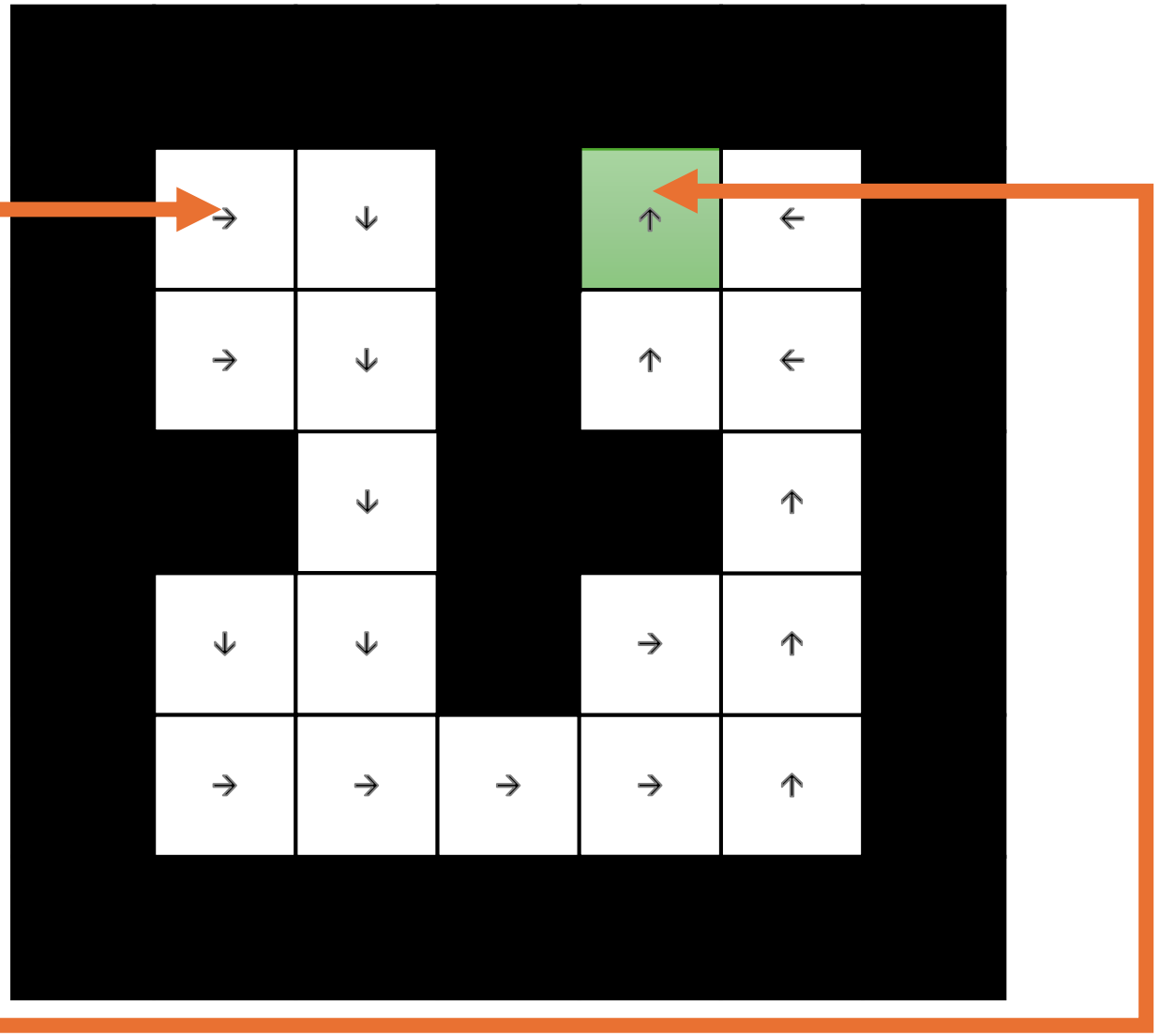
Kasper Overgaard Mortensen · Konstantinos Skitsas · Emil Morre Christensen · Mohammad Sadegh Talebi · Andreas Pavlogiannis · Davide Mottin · Panagiotis Karras

A project based upon Emil Morre Christensen's master's thesis

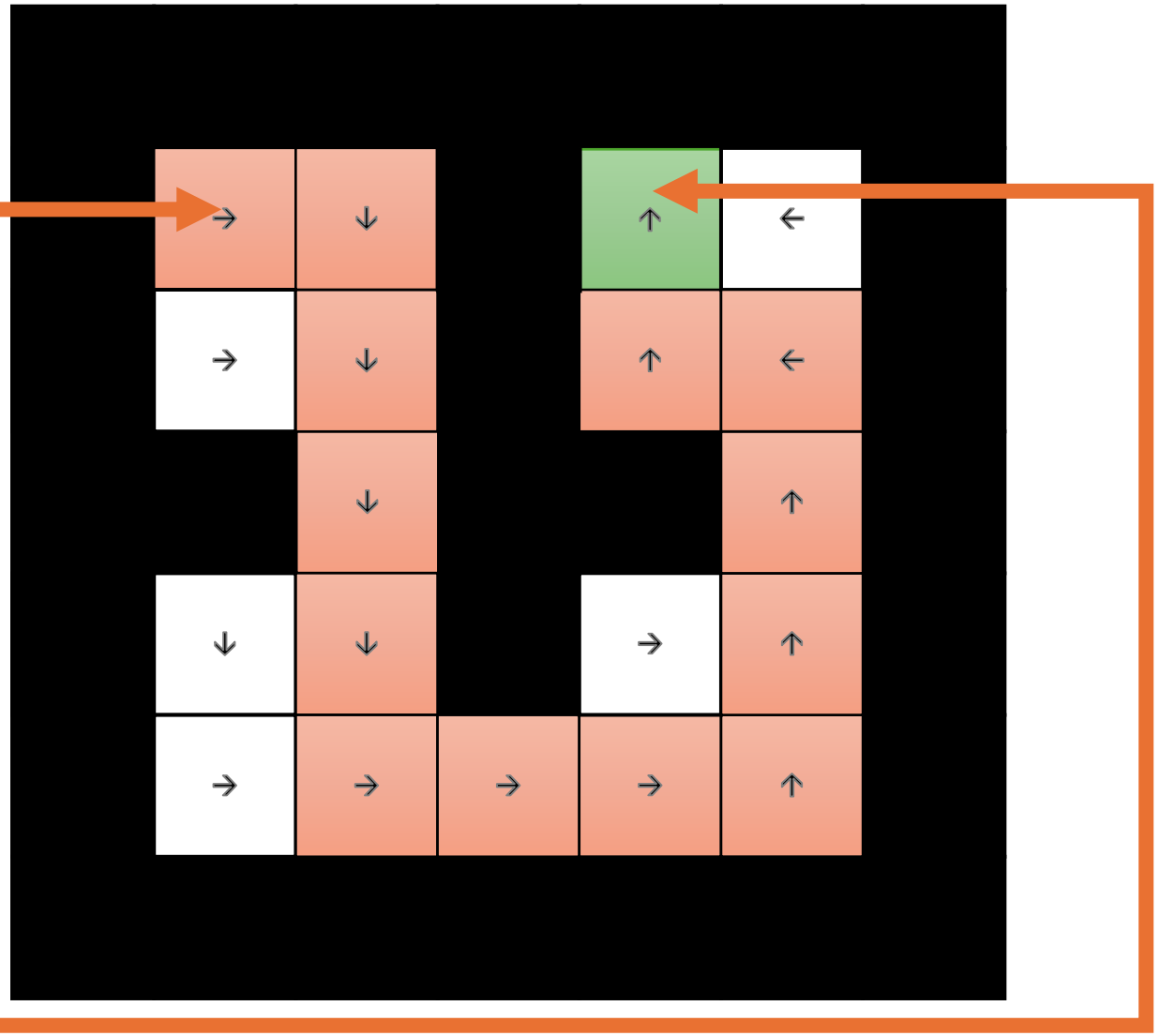
Planning



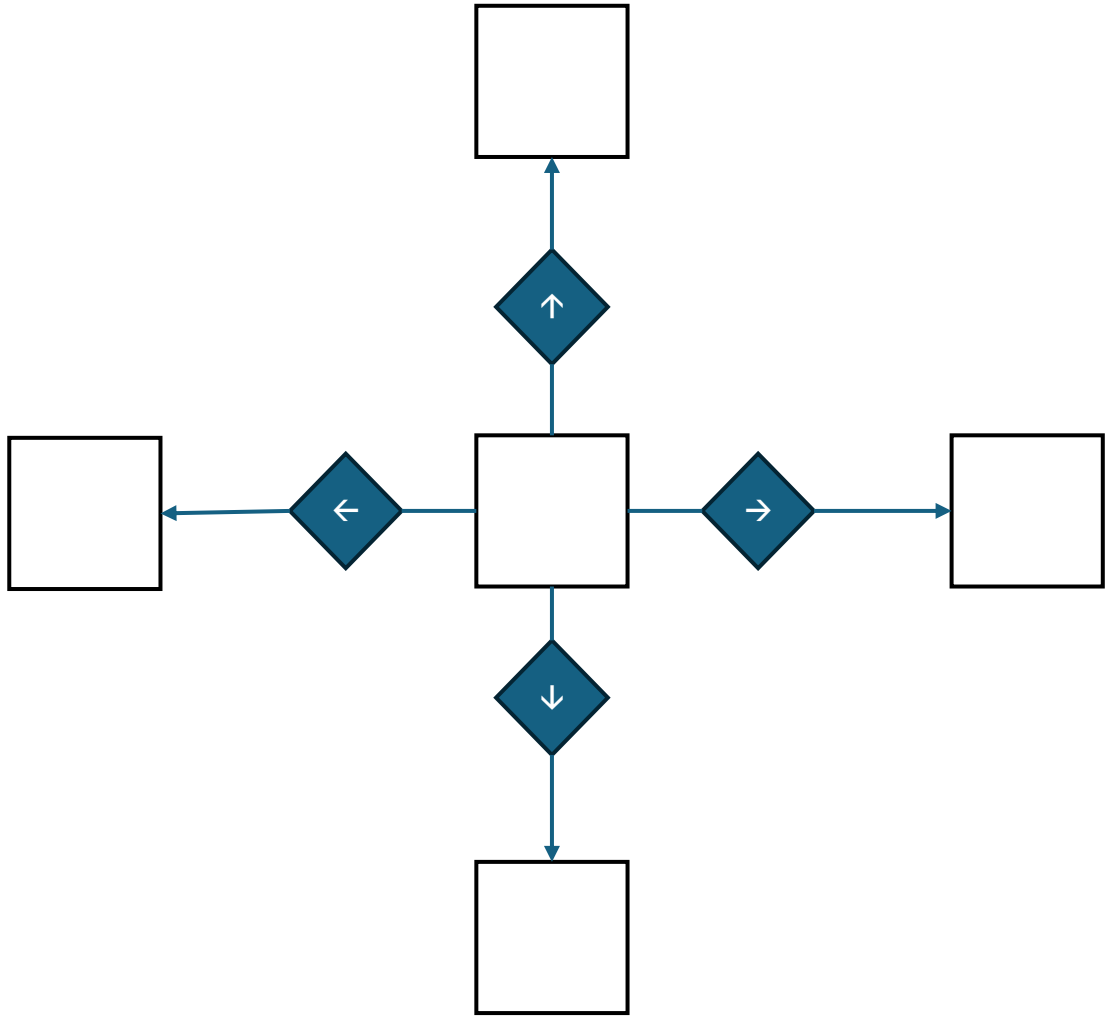
Planning



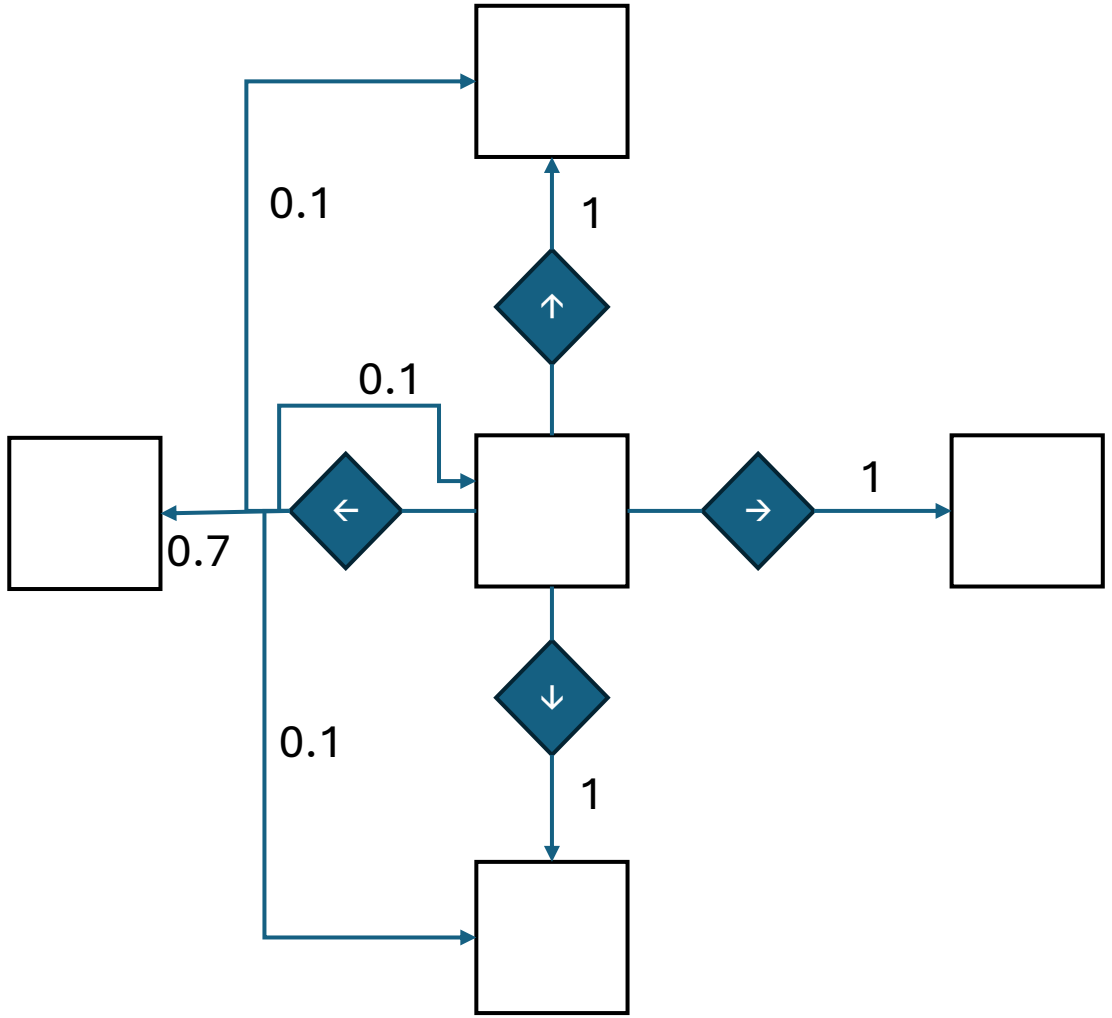
Planning



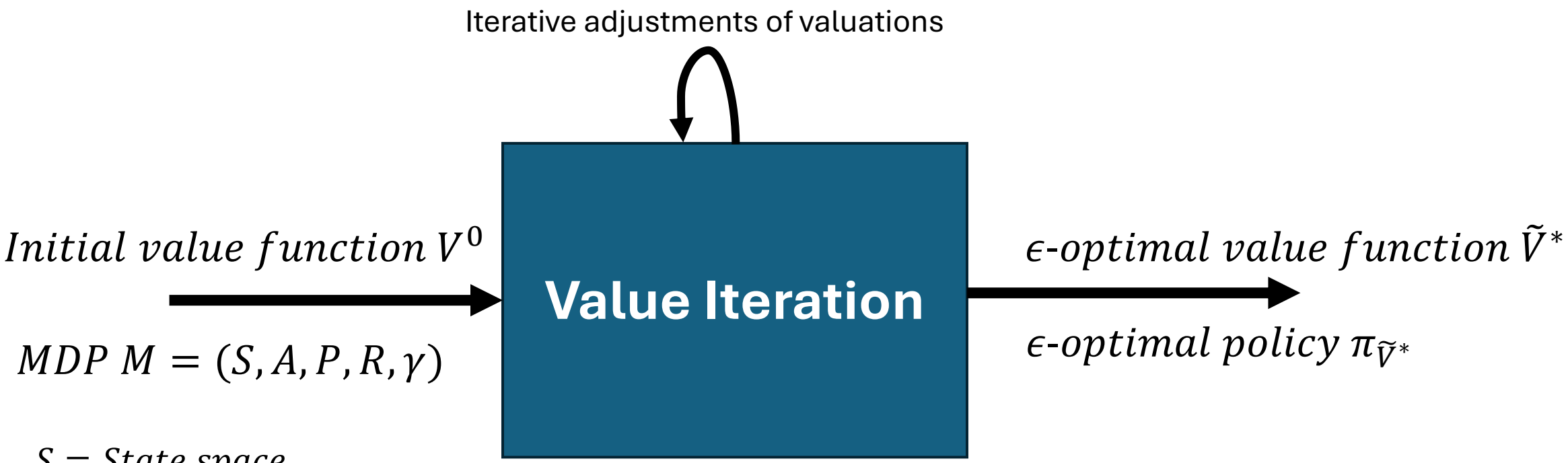
Planning with MDPs



Planning with MDPs

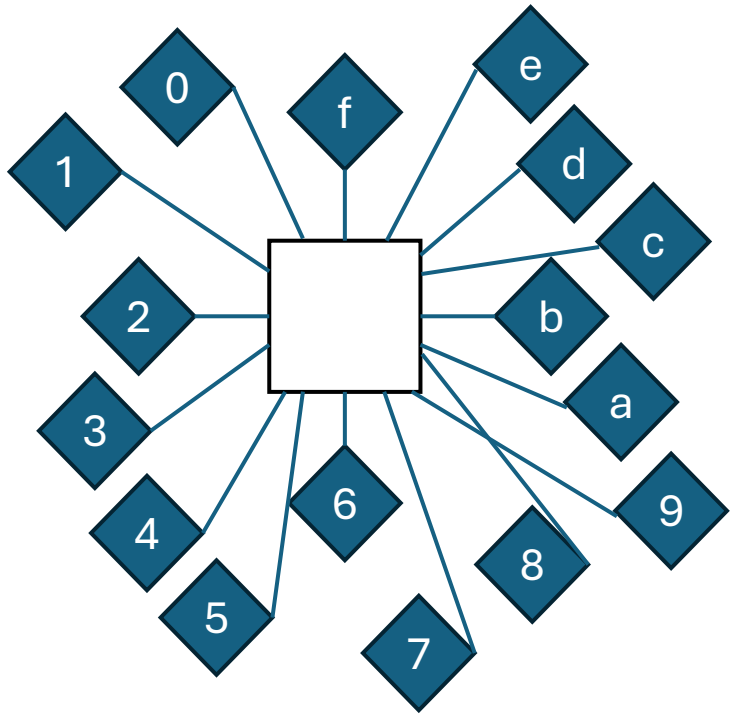


Planning with MDPs using Value Iteration

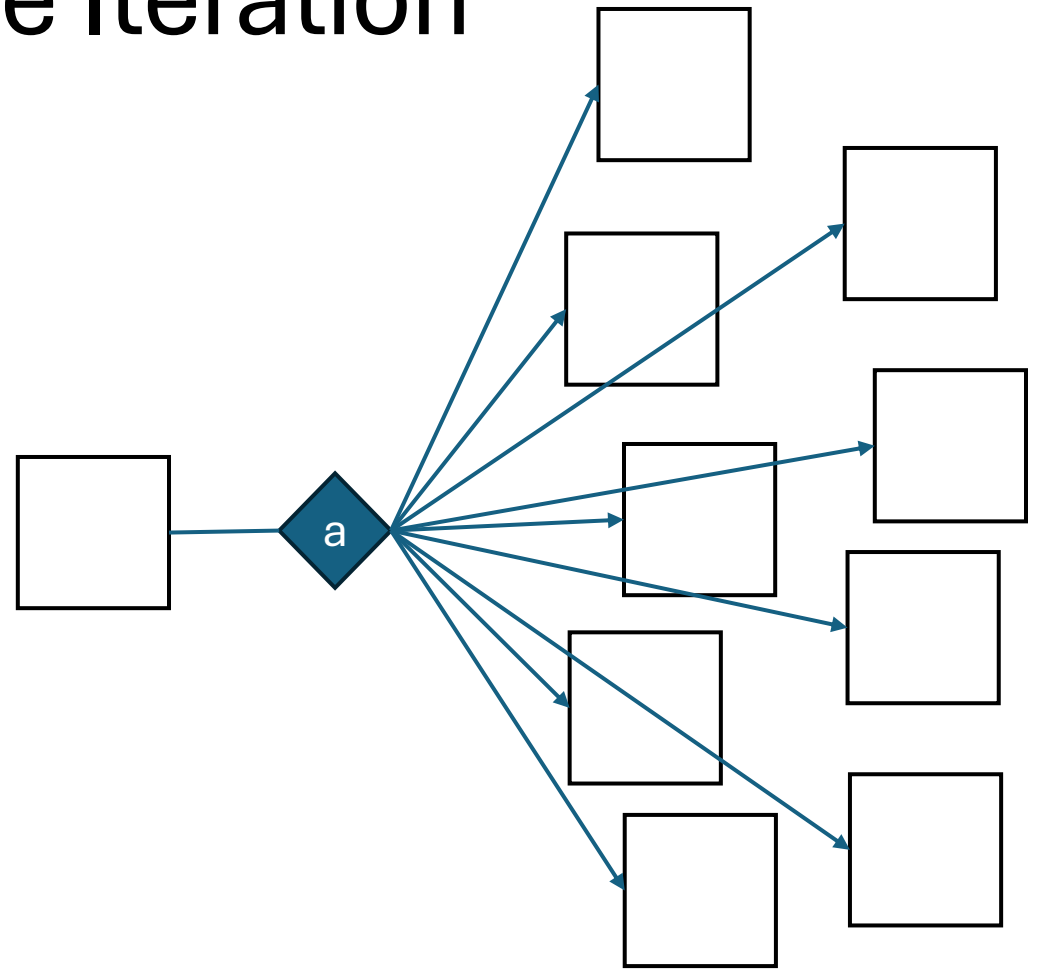


$S = State\ space$
$A = Action\ space$
$P = Transition\ probability\ funtion$
$R = Reward\ function$
$\gamma = Future\ reward\ discount\ factor$

Scalability issues of Value Iteration



Per-state complexity scales with number of actions



And scales with non-zero transition probabilities

A simple solution

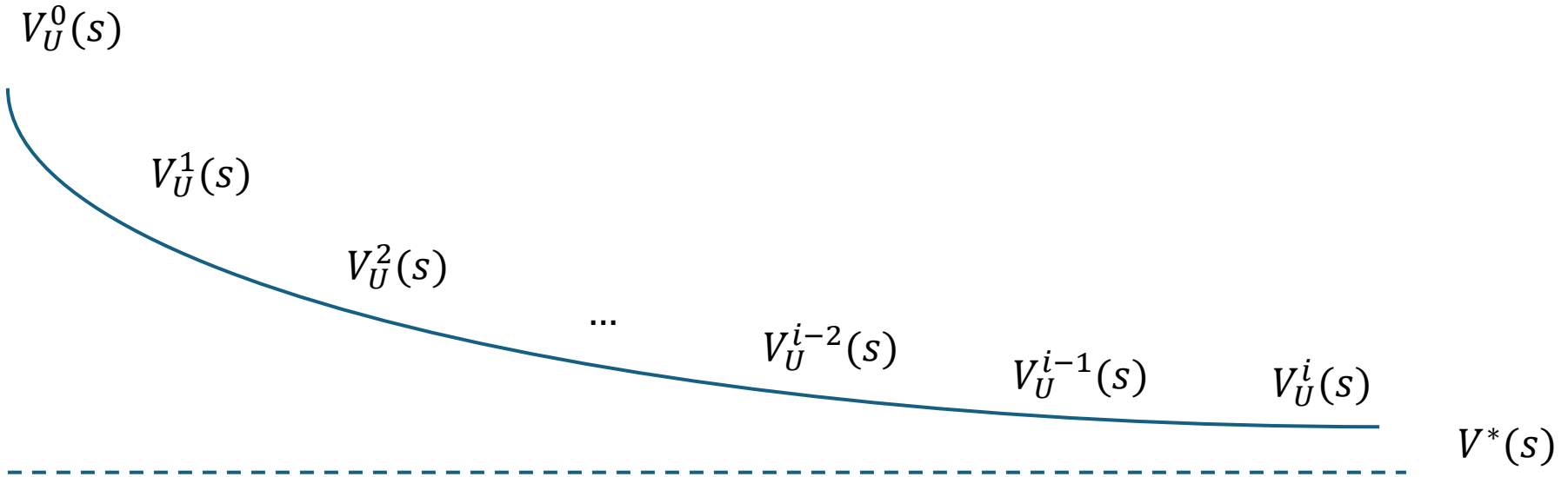
Beneficial to avoid updating action values!

But how?

An upper bound exists…

…such that value's updates are monotonic decreasing.

For state $s \in S$, $V^*(s) \leq V_U(s)$ ← Upper bound conveniently not defined here.



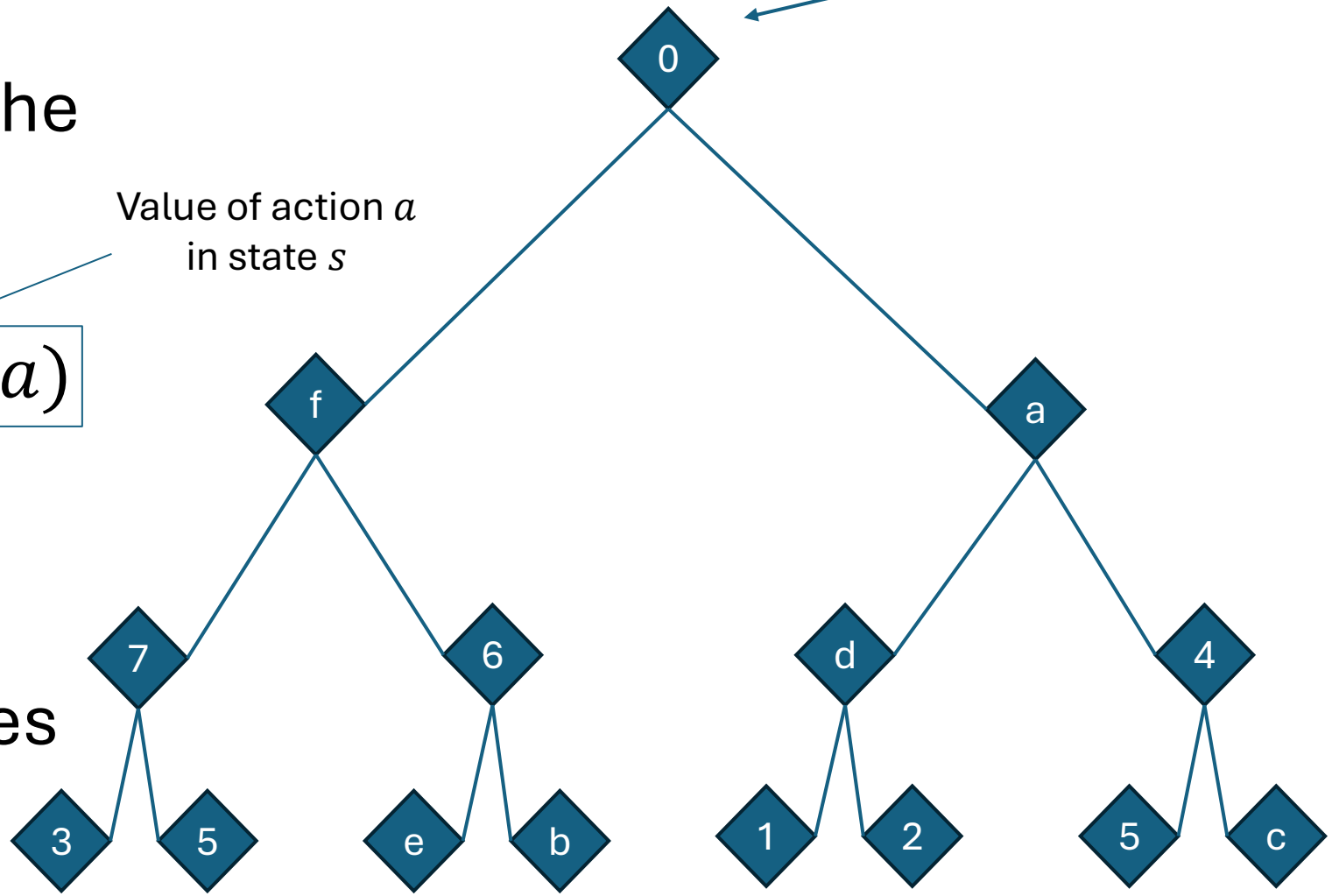
Update only the best action

… until one remains the best.

$$V(s) = \max_{a \in A(s)} Q(s,a)$$

Value of action a in state s

Highest valued action for state $s \in S$



Maintain action-values in max heaps.

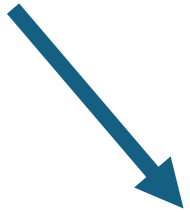
Update only the best action

Overhead: Build the heaps.

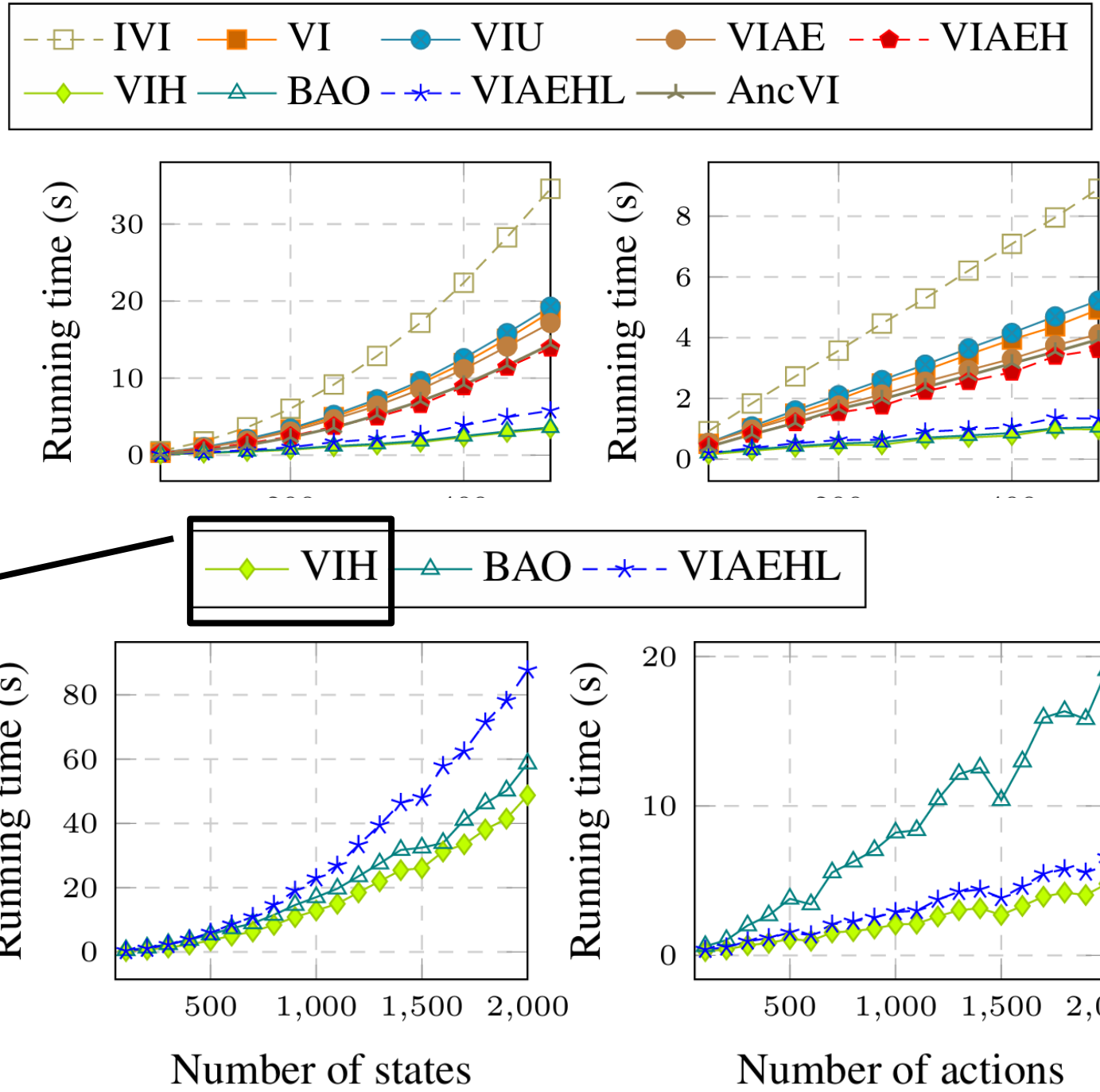
Per iteration of Value Iteration:

Worst case: Update all actions.

Best case: One action update per state.

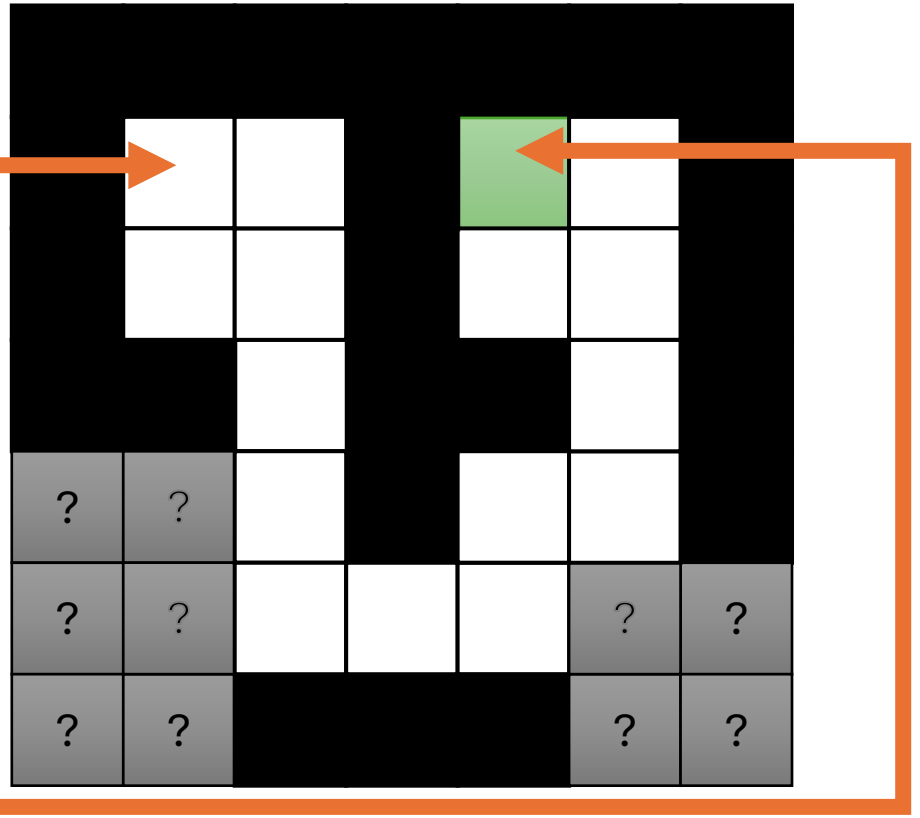


Approaches the best case
as Value function converge.


IVI
VI
VIU
VIAE
VIAEH
VIH
BAO
VIAEHL
AncVI
Running time (s)
0
10
20
30
Running time (s)
0
2
4
6
8
VIH
BAO
VIAEHL
Method presented
Running time (s)
0
20
40
60
80
500
1,000
1,500
2,000
Number of states
Running time (s)
0
10
20
500
1,000
1,500
2,000
Number of actions

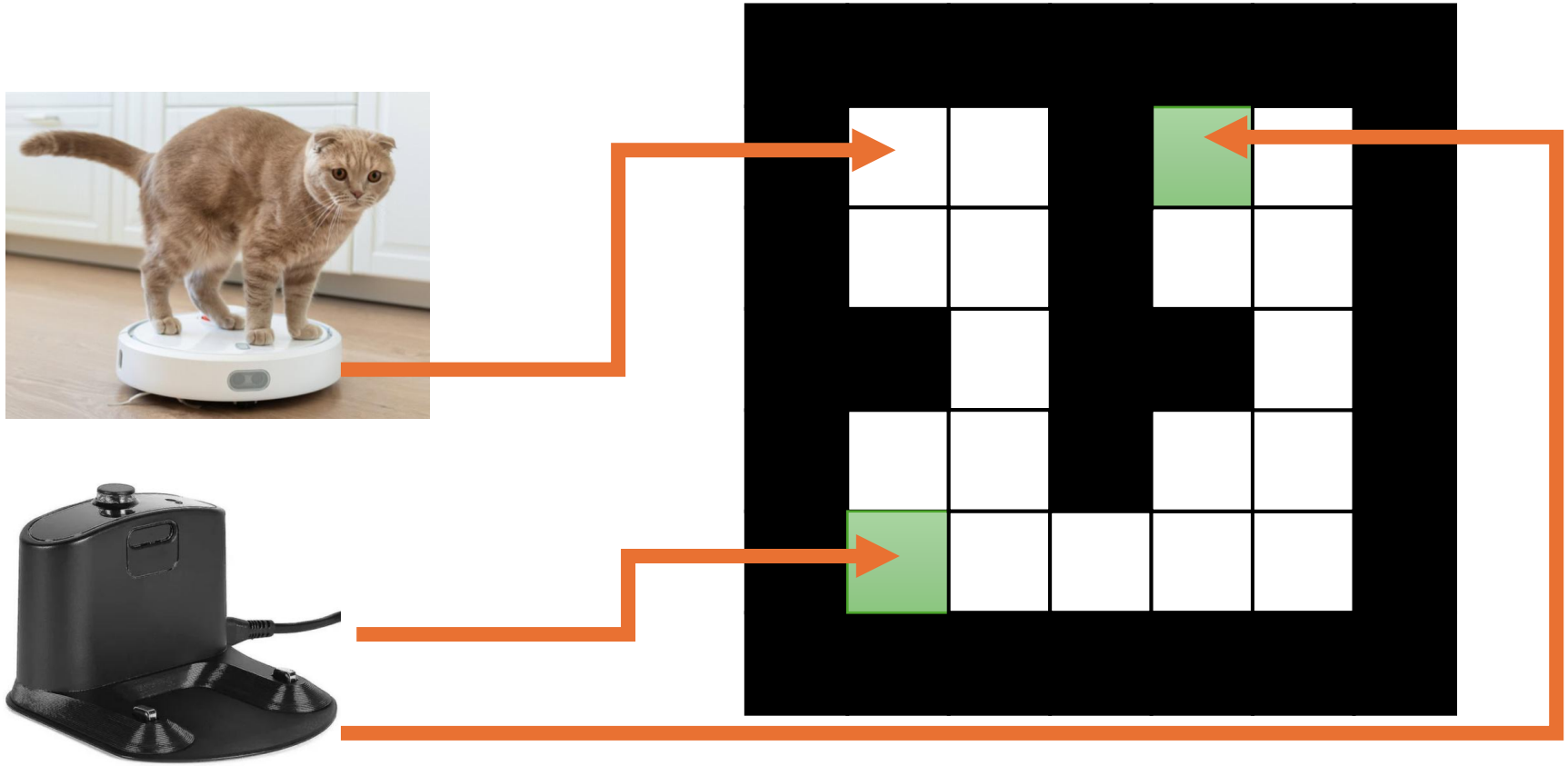
Learning with MDPs

Unknown reward function and transition probabilities.



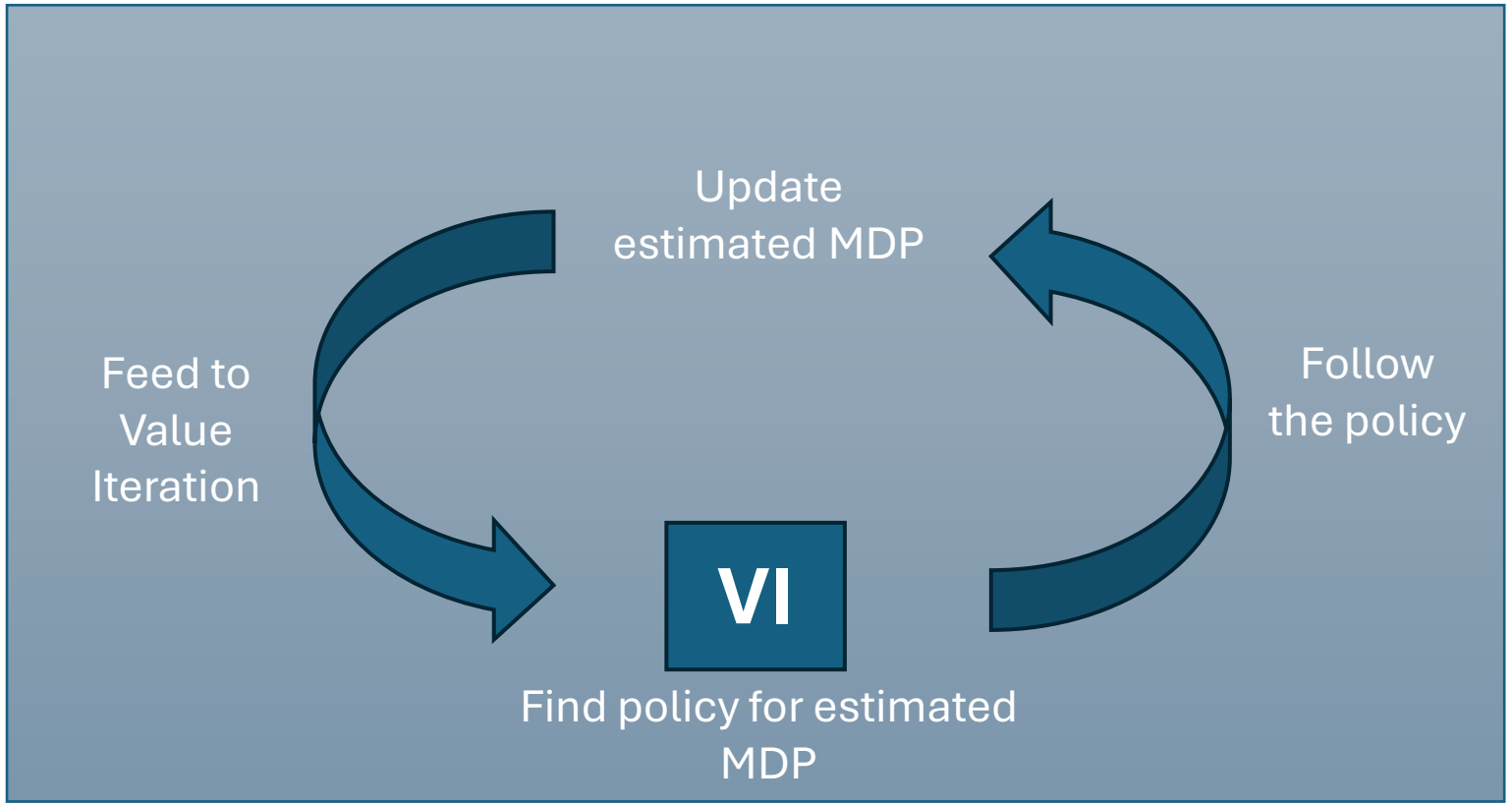
Learning with MDPs

Learned through a balance of exploration and exploitation.



Learning with MDPs using Value Iteration

Substitute VI if it is used as a subroutine.



Thank you

Contact me: km@cs.au.dk

Repository: github.com/constantinosskitsas/SwiftVI

Repository for Emil’s master’s thesis:
github.com/Dugtoud/Time-Efficient-VI-for-MDPs