

# Problem Set 2

Group 14 - Ragnhild Klette, Maya Evensen, Halvor Håvardssruud

February 11, 2026

## Part 1

### Exercise 1 - Markov Processes

1. (\*) The model does not depend on history to find the future states, it only relies on the current state. This is the Markov property, and it is satisfied in this model.
2. (\*)

We model the behavior of the Mars Rover as a discrete-time Markov process  $(\mathcal{S}, P)$ , where  $\mathcal{S}$  denotes the set of states and  $P$  the transition probability matrix.

The state space of the Markov process is defined as

$$\mathcal{S} = \{E, P, S, R, M\},$$

where:

- $E$ : Normal exploration mode
- $P$ : Taking pictures of the sky
- $S$ : Taking samples from the ground
- $R$ : Recharging
- $M$ : Malfunction (absorbing state)

	$E$	$P$	$S$	$R$	$M$
$E$	0.7	0.1	0.1	0.1	0
$P$	0.6	0	0.4	0	0
$S$	0.35	0	0.30	0.30	0.05
$R$	1	0	0	0	0
$M$	0	0	0	0	1

3. No, all states are not reachable from the normal operative mode. It can not reach malfunction (or shutdown). If the model has reached malfunction, it can not reach any other state except shutdown.

Figure 1: Mars rover Markov process graph representation

4. If the robot is in normal operative mode, the probability of malfunctioning within the next 10 minutes is 0. Within the next 20 minutes, there is only one path to malfunction, which is through "take samples". The probability of this happening is  $0.1 \cdot 0.05 = 0.005$ . Within the next 30 minutes, there are multiple paths to malfunction, starting from Explore,  $E \rightarrow$ :

- $E \rightarrow S \rightarrow M$ ,  $0.7 \cdot 0.1 \cdot 0.05 = 0.0035$
- $S \rightarrow S \rightarrow M$ ,  $0.1 \cdot 0.3 \cdot 0.05 = 0.002$
- $P \rightarrow S \rightarrow M$ ,  $0.1 \cdot 0.4 \cdot 0.05 = 0.0015$
- Plus the probability of malfunctioning after 20 minutes, 0.005
- $0.0035 + 0.002 + 0.0015 + 0.005 = 0.012$

There is a 1.2% chance of malfunction after 30 minutes.

5. (\*) If the probability of malfunctioning while in the sampling (digging) state depends on the number of times the robot has previously extracted samples, then the process no longer satisfies the Markov property in its current formulation. Since you no longer know the transition probabilities based on the current state. If the malfunction probability changes with the number of past digging actions, then knowing state  $S$  is no longer sufficient to determine the next state distribution.

To recover Markovianity all information needed to determine future transitions need to be present in the current state. Adding the number of times the rover has sampled in the state definition.  $S_k =$  "sampling for the  $k$ -th time" would fix this issue and the process becomes Markovian again.

6. Yes. Model the board with the configuration of all Xs and Os and whose turn it is, then the state for the game at time  $t+1$  is only dependent on the current state at time  $t$ .
7. There are  $3^9$  possible configurations of the tic-tac-toe board, since each of the 9 squares can be either empty, contain an X, or contain an O. Many of these configurations are not valid game states, for instance boards with five Xs and one O. Removing such illegal and unreachable configurations significantly reduces the effective size of the state space.

The number of distinct states can be even more reduced by exploiting symmetries of the board, since multiple configurations are equivalent under rotations and reflections. Even after accounting for these reductions, the number of states remains large and grows exponentially with the size of the board.

This exponential growth becomes evident when increasing the board size. For a  $4 \times 4$  board, the total number of configurations is  $3^{16}$ , which is 2187 times larger than the  $3^9$  configurations of the  $3 \times 3$  board. This illustrates the rapid increase of the state space and shows scalability limitation of Markov process models for larger games.

## Exercise 2 - Markov Reward Processes

1. (\*)

Figure 2: Mars rover MRP (Markov reward process)

2.  $E(10 \text{ minutes}) = 0.7 \cdot 0 + 0.1 \cdot 5 + 0.1 \cdot 20 + 0.1 \cdot 0 = 2.5$ . The probability of getting a negative reward in twenty minutes is the same as the probability of getting a malfunction in twenty minutes, which is 0.5%

## Exercise 3 - Markov Decision Processes

1. (\*)

**States** The state space remains defined as:

$$\mathcal{S} = \{E, P, S, R, M\},$$

where:

- $E$ : Normal exploration mode
- $P$ : Taking pictures of the sky
- $S$ : Taking samples from the ground
- $R$ : Recharging
- $M$ : Malfunction (absorbing state)

### Actions

- $A(E) = \{\text{explore again, take pictures, take samples, recharge}\}$
- $A(P) = \{\text{explore, take samples}\}$
- $A(S) = \{\text{explore, take samples again, recharge, malfunction}\}$
- $A(R) = \{\text{explore}\}$
- $A(M) = \{\phi\}$

### Transition

Moving to state  $s'$  after taking action  $a$  in state  $s$  is:

$$P(s' \mid s, a)$$

and since we are choosing the action and not randomly selecting ...

### Policy

A policy  $\pi$  specifies the rover's behaviour.