Surname, Name, ID number: ______________________

Autonomous and Mobile Robotics M

23 December 2021 - Theory

Some questions may have more than one correct answers: for each question, indicate all the correct answers.

1. The configuration space of a unicycle mobile robot is:
   - ⊗ $[x\ y\ \theta]^T \in \mathbb{R}^2 \times \mathbb{S}$
   - ○ $[x\ y\ \theta]^T \in \mathbb{R}^3$
   - ○ $[x\ y\ \theta\ \gamma]^T \in \mathbb{R}^2 \times \mathbb{S}^2$
2. A constraint is said *non-holonomic* if:
   - ⊗ the differential relation between the coordinates is not reducible to finite form
   - ○ finite relations between the coordinates of the system are present
   - ○ if differentiable/integrable relations between the coordinates of the system are present
3. Given the constraints matrix equation in Pfaffian form $A(q)\dot{q} = 0$, the admissible robot speed:
   - ⊗ is generated by a matrix $G(q)$ such that $\mathrm{Im}(G(q)) = \mathrm{Ker}(A(q)), \forall q$
   - ○ is generated by a matrix $G(q)$ such that $\mathrm{Ker}(G(q)) = \mathrm{Im}(A(q)), \forall q$
   - ○ is generated by a matrix $G(q)$ such that $G(q) = A(q)^{-1}, \forall q$
4. Consider an obstacle avoidance algorithm based on potential fields
   - ⊗ the overall potential is the sum of an attractive one (generated by the goal) and a repulsive one, generated by the obstacles
   - ○ a concave shape of the obstacle can in many cases avoid the problem of local minima
   - ○ the control consists in setting the velocity of the robot equal to the gradient of the potential
5. In Reinforcement Learning algorithms, the reward:
   - ○ must be a function of the agent state
   - ⊗ can be a function of the environment state
   - ○ depends on time
6. In Reinforcement Learning algorithms, the agent:
   - ⊗ selects actions to maximize total expected future reward
   - ⊗ may require require balancing immediate and long term rewards
   - ○ selects actions to minimize the task execution time
7. A process satisfies the Markov property if:
   - ○ the agent state is the same as the environment state
   - ⊗ one can make predictions for the future of the process based solely on its present state
   - ○ one can make predictions for the future of the process only based on the process full history
8. In Reinforcement Learning, the policy:
   - ○ is the learning process that will maximize the sum of future rewards
   - ○ is a deterministic function of the state
   - ⊗ the strategy that the agent employs to decide the action to take on the current state
9. The *action value function* is defined as:
   - ○ $q_\pi(s,a) = \mathbb{E}_\pi[R_{t+1}|S_t = s, A_t = a]$
   - ○ $q_\pi(s,a) = \mathbb{E}_\pi[G_t|S_t = s]$
   - ⊗ $q_\pi(s,a) = \mathbb{E}_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s, A_t = a]$
10. The Bellman optimality equation for the state value function can be written as
    - ○ $v_*(s) = \max v_\pi(s)$
    - ⊗ $v_*(s) = \max_{a\in\mathcal{A}} q_*(s,a)$
    - ⊗ $v_*(s) = \max_{a\in\mathcal{A}} \mathbb{E}_*[G_t|S_t = s, A_t = a]$

Autonomous and Mobile Robotics M

23 December 2021 - Exercise

The student is asked to solve the following problem.

Let us consider a fully observable and deterministic environment with 5 states providing the following rewards

$R(s_1)$	$R(s_2)$	$R(s_3)$	$R(s_4)$	$R(s_5)$
-1	0	1	0	-1

The set of possible actions is {MoveLeft, MoveRight}, proving with probability 1 the transition of the state to the left one or to the right one respectively. The environment is initially at state s_1, and first the policy $\pi_1(\cdot) = \text{MoveRight}$ is applied for 4 time steps. Then, the policy $\pi_2(\cdot) = \text{MoveLeft}$ is applied for additional 4 time steps.

Starting from an arbitrary initialisation of the state value function and assuming a discount factor $\gamma = 1$ and a weight $\alpha = 0.5$, compute the state value function provided by a TD algorithm after the execution of π_1 and π_2 in the following two tables.

$v_{\pi_1}(s_1)$	$v_{\pi_1}(s_2)$	$v_{\pi_1}(s_3)$	$v_{\pi_1}(s_4)$	$v_{\pi_1}(s_5)$

$v_{\pi_2}(s_1)$	$v_{\pi_2}(s_2)$	$v_{\pi_2}(s_3)$	$v_{\pi_2}(s_4)$	$v_{\pi_2}(s_5)$

Solution:

The state value function is initialized to 0 for all the states.

$v_{\pi_1}(s_1)$	$v_{\pi_1}(s_2)$	$v_{\pi_1}(s_3)$	$v_{\pi_1}(s_4)$	$v_{\pi_1}(s_5)$
0	0.5	0	-0.5	0

$v_{\pi_2}(s_1)$	$v_{\pi_2}(s_2)$	$v_{\pi_2}(s_3)$	$v_{\pi_2}(s_4)$	$v_{\pi_2}(s_5)$
0	-0.25	0.25	0.25	-0.25