
Model Predictive Control and Reinforcement Learning

– Summer School 2021 –
Joschka Boedecker and Moritz Diehl

For the multiple choice questions, which give exactly one point, tick exactly one box for the right answer.

1. Which of the following functions $f(x)$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, is NOT convex ($c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$)?

(a) <input type="checkbox"/> $\ Ax\ _2^2 + \log(c^\top x)$	(b) <input type="checkbox"/> $\ Ax\ _2^2 + \exp(c^\top x)$	(c) <input type="checkbox"/> $c^\top x + \exp(\ Ax\ _2^2)$	(d) <input type="checkbox"/> $-c^\top x + \ Ax\ _2^2$		
<table border="1"><tr><td>1</td><td></td></tr></table>				1	
1					

2. The local convergence rate of Newton's method is:

(a) <input type="checkbox"/> superlinear	(b) <input type="checkbox"/> quadratic	(c) <input type="checkbox"/> linear	(d) <input type="checkbox"/> sublinear		
<table border="1"><tr><td>1</td><td></td></tr></table>				1	
1					

3. A point in the feasible set of an NLP that satisfies the KKT optimality conditions is

(a) <input type="checkbox"/> the global minimum	(b) <input type="checkbox"/> a boundary point		
(c) <input type="checkbox"/> a local minimum	(d) <input type="checkbox"/> a candidate for local minimum		
<table border="1"><tr><td>1</td><td></td></tr></table>		1	
1			

4. What is the most general (unconstrained) problem type to which the Gauss-Newton Hessian approximation is applicable?

(a) <input type="checkbox"/> non-linear least squares objective	(b) <input type="checkbox"/> linear least-squares objective		
(c) <input type="checkbox"/> linear objective	(d) <input type="checkbox"/> any convex objective		
<table border="1"><tr><td>1</td><td></td></tr></table>		1	
1			

5. How does CasADi compute derivatives?

(a) <input type="checkbox"/> Finite differences	(b) <input type="checkbox"/> Imaginary trick		
(c) <input type="checkbox"/> Algorithmic Differentiation	(d) <input type="checkbox"/> Symbolic Differentiation		
<table border="1"><tr><td>1</td><td></td></tr></table>		1	
1			

6. How many optimization variables does the NLP arising in the direct **multiple shooting** method have, if the system has n_x states, n_u controls, the initial value is fixed, and the time horizon is divided into N control intervals (piecewise-constant)?

(a) <input type="checkbox"/> Nn_u	(b) <input type="checkbox"/> $Nn_x^3 + Nn_u^2$	(c) <input type="checkbox"/> $(N + 1)n_x + Nn_u$	(d) <input type="checkbox"/> $\frac{1}{3}N^3n_u^3$
1 <input style="width: 20px;" type="text"/>			

7. Regard an MPC optimization problem for $N = 10$ steps of the discrete time system $s^* = 2s + a$ with continuous state $s \in \mathbb{R}$ and continuous bounded control action $a \in [-1, 1]$. The stage cost is given by $c(s, a) = a^2$ and the terminal cost by $E(s) = 100s^2$. The initial state is \bar{s}_0 . To which optimization problem class does the problem belong?

(a) <input type="checkbox"/> Linear Programming (LP)	(b) <input type="checkbox"/> Mixed Integer Programming (MIP) but not LP
(c) <input type="checkbox"/> Quadratic Programming (QP) but not LP	(d) <input type="checkbox"/> Nonlinear Programming (NLP) but not QP
1 <input style="width: 20px;" type="text"/>	

8. Regard an MPC optimization problem for $N = 10$ steps of the discrete time system $s^+ = 2s^2 + a$ with continuous state $s \in \mathbb{R}$ and continuous bounded control action $a \in [-1, 1]$. The stage cost is given by $c(s, a) = a^2$ and the terminal cost by $E(s) = 100s^2$. The initial state is \bar{s}_0 . To which optimization problem class does the problem belong?

(a) <input type="checkbox"/> Linear Programming (LP)	(b) <input type="checkbox"/> Mixed Integer Programming (MIP) but not LP
(c) <input type="checkbox"/> Quadratic Programming (QP) but not LP	(d) <input type="checkbox"/> Nonlinear Programming (NLP) but not QP
1 <input style="width: 20px;" type="text"/>	

9. Regard dynamic programming for the discrete time system $s^+ = s + a$ with continuous state $s \in \mathbb{R}$ and continuous bounded control action $a \in [-1, 1]$, with zero stage cost $c(s, a) = 0$. We apply one step of dynamic programming (with operator T) to the value function $J_1(s) = \max(0, s)$. What is the resulting function $J_0 = TJ_1$?

(a) <input type="checkbox"/> $J_0(s) = \max(0, s - 1)$	(b) <input type="checkbox"/> $J_0(s) = \max(0, s + 1)$
(c) <input type="checkbox"/> $J_0(s) = 0$	(d) <input type="checkbox"/> $J_0(s) = \max(s - 1, s + 1)$
1 <input style="width: 20px;" type="text"/>	

10. What is meant by monotonicity of DP? Formulate it using T as the DP operator, acting on value functions J and J' .

(a) <input type="checkbox"/> $J' \geq J \Rightarrow TJ' \geq TJ$	(b) <input type="checkbox"/> $J' \leq J \Rightarrow TJ' \geq TJ$
(c) <input type="checkbox"/> $TJ' \geq TJ \Rightarrow J' \leq J$	(d) <input type="checkbox"/> $TJ' \leq TJ \Rightarrow J' \leq J$
1 <input style="width: 20px;" type="text"/>	

11. Which of the following components is *not* part of an MDP specification?

(a) <input type="checkbox"/> Set of states	(b) <input type="checkbox"/> Set of rewards	(c) <input type="checkbox"/> Policy	(d) <input type="checkbox"/> Set of actions
1			

12. Imagine you want to apply the algorithms from this lecture on a real physical system. You get sensor input after each 0.05 seconds, but the execution of actions has a delay of 0.2 seconds. Is the Markov property fulfilled?

(a) <input type="checkbox"/> No	(b) <input type="checkbox"/> Yes	(c) <input type="checkbox"/> Yes, if a history of the last 0.2 seconds is added to the state space	(d) <input type="checkbox"/> Only if a function approximator is used for the value function
1			

13. With what could you derive/calculate the value function $v_\pi(s)$ from the action-value function $q_\pi(s, a)$:

(a) <input type="checkbox"/> With the Markov Decision Process (MDP)	(b) <input type="checkbox"/> With the policy π	(c) <input type="checkbox"/> Not possible	(d) <input type="checkbox"/> Only possible with both the MDP and the policy π
1			

14. Why is Q-learning an off-policy method?

(a) <input type="checkbox"/> Because using an ϵ -greedy policy changes actions randomly	(b) <input type="checkbox"/> Because Q-learning uses a bootstrapped value, instead of a Monte-Carlo rollout
(c) <input type="checkbox"/> Because we learn Q-values instead of a policy	(d) <input type="checkbox"/> Because we learn Q-values for the greedy policy, while using a different policy to interact with the environment.
1	

15. The correct Q-Learning update is:

(a) <input type="checkbox"/> $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a^*} Q(s, a^*) - Q(s', a^*)]$	(b) <input type="checkbox"/> $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a^*} Q(s', a^*) - Q(s, a)]$
(c) <input type="checkbox"/> $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a) - Q(s', a)]$	(d) <input type="checkbox"/> $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a) - Q(s, a)]$
1	

16. What can you control with the ϵ of an ϵ -greedy policy?

(a) <input type="checkbox"/> The update size of a temporal difference method	(b) <input type="checkbox"/> How much the agent emphasizes short term rewards vs long term rewards	(c) <input type="checkbox"/> The randomness of the MDP	(d) <input type="checkbox"/> How much the agent emphasizes exploration
--	--	--	--

1

17. Target networks were introduced in order to:

(a) <input type="checkbox"/> Introduce correlations into the sequence of observations	(b) <input type="checkbox"/> Prevent forgetting past experiences	(c) <input type="checkbox"/> Avoid oscillations during training which slow down learning	(d) <input type="checkbox"/> Make RL problems less like Supervised Learning problems
---	--	--	--

1

18. In policy gradient methods, what should a baseline ideally depend on?

(a) <input type="checkbox"/> On the state	(b) <input type="checkbox"/> On the action	(c) <input type="checkbox"/> On state and action	(d) <input type="checkbox"/> On nothing (i.e. it should be constant)
---	--	--	--

1

19. Which of the following is true for Actor-Critic algorithms:

(a) <input type="checkbox"/> Can be used only in problems with discrete actions	(b) <input type="checkbox"/> They reduce gradient variance usually occurring in vanilla PG methods	(c) <input type="checkbox"/> The usage of baselines is compulsory for variance reduction	(d) <input type="checkbox"/> The actor learns a value function and the critic learns a policy
---	--	--	---

1

20. We use experience replay to:

(a) <input type="checkbox"/> Introduce correlations into the sequence of observations	(b) <input type="checkbox"/> Prevent forgetting past experiences	(c) <input type="checkbox"/> Avoid oscillations during the learning process	(d) <input type="checkbox"/> Break the curse of dimensionality
---	--	---	--

1

Empty page for calculations