# Reinforcement Learning, Tutorial 03

Philipp Kratzer

Machine Learning and Robotics Lab

MAX-PLANCK-GESELLSCHAFT

University of Stuttgart
Germany

May 13th, 2020

# Outline

# Announcements

- Points for exercise sheet in ilias
- Next exercise sheet is available

# Admin

- From now on, submissions in the wrong file format will give 0 points. Allowed is: one single pdf and one single .py. If you want to compress upload as a single .zip file! (not allowed is anything else: .doc, .txt, .odt, .rar, .ipynb, ...)
- If you send me files after the deadline they will not be graded
- Advice: Upload early and verify that your files are the correct ones

- The exercises do not influence your final grade, only the final exam does. But: you need $50\%$ of the points in the exercises to participate in the exam.

# Outline

# 1a

**Task:** Show that the Bellman optimality operator $\mathcal{T}$ is a $\gamma$-contraction

$$
\begin{aligned}
\left\|\mathcal{T}v - \mathcal{T}v'\right\|_\infty &= \left\| \max_a \sum_{s',r} p(s',r \mid s,a)\Big[r + \gamma v(s')\Big] - \max_a \sum_{s',r} p(s',r \mid s,a)\Big[r + \gamma v'(s')\Big] \right\|_\infty \\
&\leq \left\| \max_a \left( \sum_{s',r} p(s',r \mid s,a)\Big[r + \gamma v(s')\Big] - \sum_{s',r} p(s',r \mid s,a)\Big[r + \gamma v'(s')\Big] \right) \right\|_\infty \\
&= \left\| \max_a \left( \gamma \sum_{s',r} p(s',r \mid s,a)\Big[v(s') - v'(s')\Big] \right) \right\|_\infty \\
&\leq \left\| \max_a \left( \gamma \sum_{s',r} p(s',r \mid s,a) \|v - v'\|_\infty \right) \right\|_\infty \\
&= \left\| \gamma \|v - v'\|_\infty \max_a \sum_{s',r} p(s',r \mid s,a) \right\|_\infty \\
&= \gamma \|v - v'\|_\infty
\end{aligned}
$$

- Even if all inequalities are equalities it converges for $\gamma < 1$
- This proof shows convergence of value iteration, in the lecture it was for policy evaluation!

# 1b

**Task:** With bounded rewards, show: $\frac{r_{\text{min}}}{1-\gamma} \leq v(s) \leq \frac{r_{\text{max}}}{1-\gamma}$

$$
\begin{aligned}
v(s) &= \mathbb{E}\left[G_t \mid S_t = s\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i R_{t+i+1} \mid S_t = s\right] \\
&\leq \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r_{\text{max}} \mid S_t = s\right] \\
&= \sum_{i=0}^{\infty} \gamma^i r_{\text{max}} \\
&= \frac{r_{\text{max}}}{1-\gamma}
\end{aligned}
$$

- Similar for $r_{min}$

# 1b cont.

**Task:** Show: $|v(s) - v(s')| \leq \frac{r_{\text{max}} - r_{\text{min}}}{1-\gamma}$

- Use equation previous slide:

$$|v(s) - v(s')| \leq v_{\text{max}} - v_{\text{min}} = \frac{r_{\text{max}}}{1-\gamma} - \frac{r_{\text{min}}}{1-\gamma} = \frac{r_{\text{max}} - r_{\text{min}}}{1-\gamma}$$

# 2a

**Task:** Compute optimal value function with value iteration.

- number iterations: 43

| | | | |
|---|---|---|---|
| 0.01543432 | 0.01559069 | 0.02744009 | 0.01568004 |
| 0.02685371 | 0. | 0.05978021 | 0. |
| 0.0584134 | 0.13378315 | 0.1967357 | 0. |
| 0. | 0.2465377 | 0.54419553 | 0. |

# 2b

**Task:** Compute optimal policy

- Multiple optimal policies:

| | | | |
|---|---|---|---|
| Down/Right | Up | Right | Up |
| Left | H | Left/Right | H |
| Up | Down | Left | H |
| H | Right | Down | G |

- Why action "Right" in state 2?

# Outline

1. Announcements
2. Solutions
3. Outlook

# Next exercise sheet

- Next exercise sheet available
- It is about Monte-Carlo Methods
- Programming part is on the blackjack environment that we also had in the lecture
- Sourcecode on github
  `https://github.com/humans-to-robots-motion/rl-course`