V4E2 - Numerical Simulation

Sommersemester 2018
Prof. Dr. J. Garcke
Teaching assistant: Biagio Paparella
Tutor: Marko Rajković
(marko.rajkovic@uni-bonn.de)

UNIVERSITÄT BONN

Exercise sheet 12 (Bonus).

To be handed in on **Tuesday, 17.07.2018**.

Theory recap (end of Chapter 3)

An important part in the conclusion of the previous chapter has been the study of the infinite-horizon optimal control problem. We indicate with v the value function that we want to estimate. Recall that our strategy is to firstly discretize time, obtaining $v^\tau \to v$ as $\tau = \Delta t \to 0$ (due to a discrete version of DPP - Thm 38) , then to choose a space discretization (with a corresponding DPP principle) in a way that $v_h^\tau \to v^\tau$ as $h \to 0$. Differently from the finite-horizon case, we opted for a polyhedral finite-element method with leading parameter h (see A7, Thm 40 and 41). In particular the function evaluation is then not limited on nodes only.
The solution to the infinite-horizon problem is finally given by Theorem 42, claiming the full convergence $v_h^\tau \to v$ when $\tau \to 0$ and $h \to 0$ too. After that section we assumed to have $c_1 h \leq \tau \leq c_2 h$ for positive constants, justifying the writing $v_h \to v$ understood to mean the convergence in Thm 42 where now only h needs to go to zero thanks to the coupling with τ.
A practical question arises: how can we concretely generate the values $v_h(x)$?
An opportunity widely used in the field is given by the so-called *Q-values*. For instance, if we set:

- $v_h^0(x) = 0$
- $Q_h^{k+1}(x,a) = \gamma^\tau v_h^k(x + \tau f(x,a)) + \tau l(x,a)$
- $v_h^{k+1} = \min_{a \in A} Q_h^{k+1}(x,a)$

we observe that $v_h^{k+1} \to v_h$ as $k \to \infty$ by using the theorems just mentioned (nothing new). In other words, there exists a limiting Q_h and we have $V_h(x) = \min_{a\in A} Q_h(x,a)$. Note how the computation of Q requires a complete knowledge of f, l, and that - if it can help to clarify - this is the Q used in the remark for proof of theorem 44 added later.
We introduced the Q-values in order to shift to the Reinforcement-Learning case. The RL setting is based exactly on the same principles here stated, but aims to obtain a final convergence to v *without a complete information* of f or l. In other words, we need different Q-values and consequently a different iterative sequence. Note that for the values defined above, one has first the convergence to v_h, and then to v by letting $h \to 0$.
Key idea: it is possible to skip the step in between. For sequences satisfying the weak contraction property (Thm 43) one has *directly* the convergence to v for $k \to \infty$, $h \to 0$, and generally is **not** actually true that $v_h^k \to v_h$ as $k \to \infty$ (this is the meaning of the triangle diagram pictured in class).
According to the way in which Q is defined, we have a model-based or model-free algorithm. We wrote in class both the precise definitions (as well as geometric intuition), but proved the weak contraction only for the model-based case (Thm 44).

Exercise 1. (Convergence of the model-free case)

If v is the exact value function for our optimal problem, explain how you would use the model-free algorithm for approximating it. Point out the structure of the proof and which properties do you need for concluding.

(9* Punkte)

The next exercise is understood to be in the context of Chapter number 4.

Exercise 2. (Monte Carlo in Reinforcement Learning)

Prove that the Monte-Carlo simulation formula

$$J_\mu(i) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} c(i,m)$$

is valid even if a state may be revisited within the same sample trajectory.
Hint: Suppose the M cost samples are generated from N trajectories, and that the $k-th$ trajectory involves n_k visits to state i and generates n_k corresponding cost samples. Denote $m_k = n_1 + ... + n_k$. Write:

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} c(i,m) = \lim_{N\to\infty} \frac{\frac{1}{N}\sum_{k=1}^{N}\sum_{m=m_{k-1}+1}^{m_k} c(i,m)}{\frac{1}{N}(n_1 + \cdot + n_N)} = \frac{\mathbb{E}(\sum_{m=m_{k+1}+1}^{m_k} c(i,m))}{\mathbb{E}(n_k)}$$

and prove that $\mathbb{E}(\sum_{m=m_{k+1}+1}^{m_k} c(i,m)) = \mathbb{E}(n_k)J_\mu(i)$.

(5* Punkte)