




# Reinforcement Learning 2015: Tutorial 5 (week from 2. 3. 2014)

1. Recall the three types of errors in RL (hint: value, policy, exploration). How are they represented in the definition of the global utility measure (or global reward average). Is it possible to trade-in the reduction of error with respect to one criterion for an increase of any other?
2. What differences can be expected from the choice of different basis functions in function-approximation-based RL? What are the advantages or disadvantages of (non-linear) function approximation?
3. In what way does function-approximation cope with the "curse of dimensionality" problem in reinforcement learning?
4. Describe and discuss the effects and side-effects of eligibility traces in RL with function-approximation.
5. Go back to tutorial 2, problem 5 (the list of potential RL applications) and discuss whether function-approximation may be useful in these cases. If so, check whether value function and/or policy are to be approximated and whether any other properties of the approximation (complexity, on-line, local basis functions etc.) can be specified in more detail using domain knowledge or after availability of the results of a few test runs .
6. Discuss the aliased gridworld example (from David Silver's lecture 7), where the agent cannot distinguish between the two grey states, i.e. for both states the same action has to be used. Actions are: N, W, S, E. Rewards and states as shown in the figure. Compare the optimal deterministic policy for the example with the optimal stochastic policy. How could an algorithm find the stochastic policy?

				
7. If a policy is given by a probability distribution of actions (conditioned on the state) from an exponential family, what is the score function (see lecture 11)?
8. Prove that gradient descent still works if the gradient of a deterministic function is multiplied by a positive matrix. What issues need to be considered in the stochastic case?
9. Use one of the Matlab programs from the first two tutorials to test RL with function-approximation. First, approximately express the states in the grid world in terms of local basis functions and, using an algorithm with function-approximation, reproduce the behaviour that was produced by the discrete algorithm for the example. Second, scale the example to larger sizes and check whether how the discrete and the continuous algorithms react to changes of the problem size.

10. Consider the figure below from a paper by Amari and Douglas "Why natural gradient?" This is actually a counterexample against natural gradient. Why? What predictions can we make about the behaviour of a stochastic algorithm. See also: Jan Peters (2010) Policy gradient methods. *Scholarpedia*, 5:11, 3698.

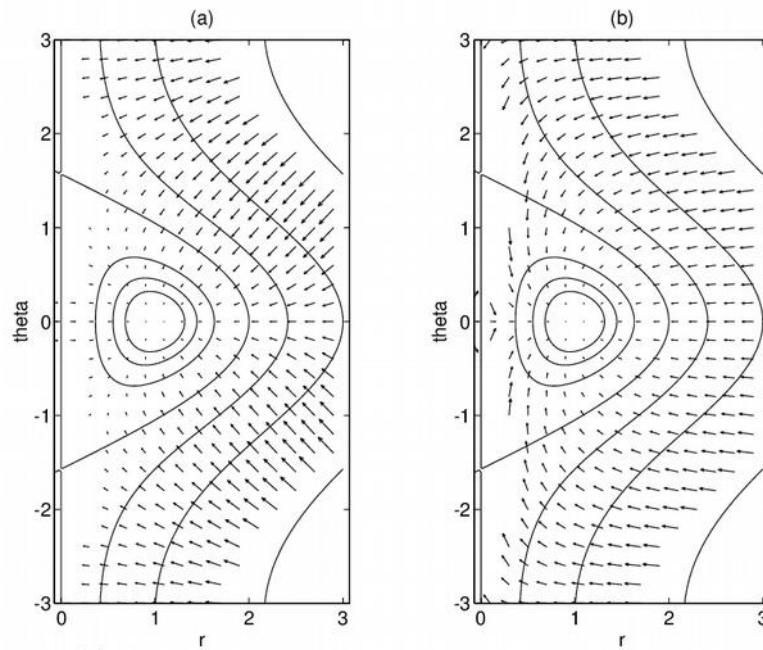


Fig. 1: (a) Standard gradients of the cost function  $\mathcal{J}_P(\mathbf{w})$ .  
 (b) Natural gradients of the cost function  $\mathcal{J}_P(\mathbf{w})$ .