An analysis of Temporal-Difference Learning with Function Approximation

Pierre Thodoroff

March 20, 2017

Outline

1. Proof of convergence
   - Assumptions
   - Theorem
2. TD with non linear function approximator
   - Divergent example

Outline

1. Proof of convergence
   - Assumptions
   - Theorem
2. TD with non linear function approximator
   - Divergent example

Underlying markov chain assumptions

- Underlying Markov chain is irreducible and aperiodic
- There exists a unique distribution π such that $\pi^{'} P = \pi^{'}$
- Variance of transition costs is finite

Assumption on the stability of the markov chain

- The function approximator ϕ has full rank, all the ϕ_k are linearly independent
- For every k, $E[\phi_k^2(x)] < \infty$
- Those assumptions guarante a stable markov chain. They are always satisfied when the state space is finite

Outline

1. Proof of convergence
   - Assumptions
   - Theorem
2. TD with non linear function approximator
   - Divergent example

Convergence Theorem

- For any $\lambda \in [0, 1]$, the $TD(\lambda)$ algorithm with linear function approximators converges with probability one

- The limit of convergence r^* is the unique solution of the equation

$$\pi T^{\lambda}(\phi r^*) = \phi r^*$$

Outline

1. Proof of convergence
   - Assumptions
   - Theorem
2. TD with non linear function approximator
   - Divergent example

Divergence with a non-linear approximator

- Markov chains with 3 states (1,2,3)
- All transition costs equal to zero and a discount factor $\alpha \in (0, 1)$
- The optimal value function is $J^* = (0, 0, 0)$
- Let the function approximator be

$$J(r) = (J(1, r), J(2, r), J(3, r))$$

Constraint on J

- By defining $J(R)$ to be the unique solution to the linear differential equation
$$\frac{\partial J}{\partial r}(r) = (Q + \epsilon I)J(r)$$

Divergence

- We can obtain a divergence shown in the picture below

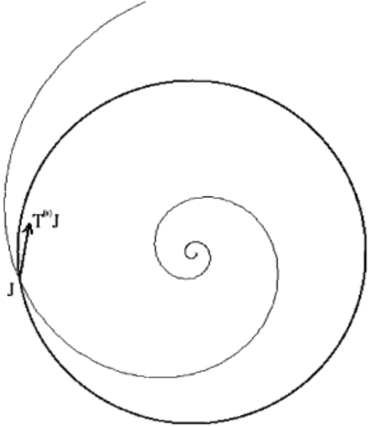

Fig. 1. Example of divergence with a nonlinear function approximator. The

References I

John N. Tsitsiklis
An Analysis of Temporal-Difference Learning with Function Approximation
IEEE VOL42.NO5.1997.