* 1. The TD(0) algorithm-

The TD(0) algorithm converges to the solution of the linear equation -

The online TD(h) algorithm will be described similarly -

And converges to the solution of the linear equation-

* 1. We shall calculate the estimation of based on the data-a sequence of N state pairs . We choose by distribution and receive by the transition probability

The projected h-step bellman equation-

The solution (minimizing the Euclidean norm)-

We recall -

The estimators are-

As seen in the result above, we would need to observe the rewards and state before using for estimation.

* 1. The error bound for policy evaluation in the 1-step case proved in lecture-

The error bound for policy evaluation in this case-

Since we get a tighter error bound for the h-step case.

For , as seen from result above, we get and-

Which is reached when *.*

However, online estimation is not possible for very large h, since observation time (observing the rewards and state before using for estimation) will massively increase the algorithm's running time.