5.5 Simulation-Based Value Iteration

Saturday, February 13, 2021 12:58 PM

General updates will be of the form,

$$J(i_k) = \min_{u \in U(i_k)} \sum_{j=0}^{n} p_{i_k j}(u)\left(g(i_k, u, j) + J(j)\right).$$

Simulation based on a Greedy Policy -

Assume that state trajectories are generated at $i_0 = 1$, and by following a greedy policy w.r.t. the current available cost-to-go function J,

$$J_{k+1}(i_k) = \min_{u \in U(i_k)} \sum_{j=0}^{n} p_{i_k j}(u)\left(g(i_k, u, j) + J_k(j)\right).$$

and $J_{k+1}(i) = J_k(i)$, if $i \neq i_k$.

The method presents lack of exploration!

Suppose we initialize each $J(i)$ to $J_0(i)$, s.t.,

$$J_0(i) \leq J^*(i).$$

We will later show that the result converges to an optimal policy starting from state 1.

Proposition 3: Assume that there exists a proper policy and all improper policies have infinite cost at some state. Assuming $J_0 \leq J^*$, then -

(a) The sequence J_k converges to some J_∞.

(b) With I being the set of states that are visited infinitely often, then $\forall i \in I$,

$$J^\mu(i) = J_\infty(i) = J^*(i), \quad \forall i \in I.$$

Proof:

We will first prove that $J_k \leq J^* \ \forall i$;

Using induction,

$$J_{k+1}(i_k) = \min_{u \in U(i_k)} \sum_{j=0}^{n} p_{i_k j}(u)\left(g(i_k, u, j) + J_k(j)\right).$$

$$\leq \min_{u \in U(i_k)} \sum_{j=0}^{n} p_{i_k j}(u)\left(g(i_k, u, j) + J^*(j)\right).$$

$$= J^*(i_k).$$

Now, we will assume that $J_k(i) = J_\infty(i) \ \forall i \notin I$ and $\forall k$. This is a special case of asynchronous VI wherein all states $i \notin I$ are treated as the final states with terminal cost $J_\infty(i)$.

We can now use result from Chapter 2 to prove that J_k converges to some J_∞. Furthermore, $\forall i \in I$, J_∞ is equivalent to the optimal cost-to-go function that we have introduced.

For each time $i \in I$, let $\mu(i)$ be a decision that is applied infinitely many times. Then we have,

$$J_{k+1}(i) = \sum_{j=0}^{n} p_{ij}(\mu(i))\left(g(i, \mu(i), j) + J_k(j)\right),$$

$$= \min_{u \in U(i)} \sum_{j=0}^{n} p_{ij}(u)\left(g(i, u, j) + J_k(j)\right).$$

Taking limit as $k \to \infty$,

$$J_\infty(i) = \sum_{j=0}^{n} p_{ij}(\mu(i))\left(g(i, \mu(i), j) + J_\infty(j)\right).$$

$$= \min_{u \in U(i)} \sum_{j=0}^{n} p_{ij}(u)\left(g(i, u, j) + J_\infty(j)\right).$$

This implies that $\bar{\mu} = \{\mu(i) \mid i \in I\}$ is an optimal policy for the modified problem.

The cost J_∞ is equivalent $J^\mu(i)$ since the states outside I are never reached.

$$J_\infty(i) = J^\mu(i) \geq J^*(i)$$

On the other hand, we showed in step 1 that $J_k \leq J^* \ \forall k \Rightarrow J_\infty = \lim_{k \to \infty} J_k \leq J^*$, hence we conclude that $J_\infty(i) = J^*(i), \ \forall i \in I$. ∎

Intuition: The algorithm, in the limit, provides us with an optimal action and with value $J^*(i)$ for all states i that are visited infinitely.