Attempted Proof Sketch to show Convergence of the Value Iteration Algorithm

Dhruv Malik

January 6th 2017

Consider the Value Update Equation:

$$V_{k+1}^{*}(s) = \max_{a} \Big\{ \sum_{s'} T(s,a,s') \times [R(s,a,s') + \gamma V_{k}^{*}(s')] \Big\} \tag{1}$$

We first note that there for any particular state s, $V_k^*(s)$ denotes the value of that state s given that we ran Expectimax search from a depth k to compute this value. We make the key observation that an Expectimax search from a depth k is the same as an Expectimax search from depth $k+1$ where the $k+1$ level of the tree is populated with zeros. This because at depth k, the value for a state s is exactly the max reward that one could have achieved by taking a particular action a (in one time step) from that state, so having zeros at the $k+1$ layer (or assigning 0 to the values of the s' states that one gets to from s), does not affect the values at depth k, due to (1).

Now note that if we write the value of $V_k^*(s)$ non-recursively, then:

$$V_{k+1}^{*}(s) = r_1 + \gamma \times (r_2 + \gamma^2 \times (r_3 + \dots (r_k + \gamma^k \times (r_k + 1))))) \tag{2}$$

and since $r_{k+1} = 0$ in the following case:

$$\begin{aligned} V_{k}^{*}(s) &= r_1 + \gamma(r_2 + \gamma^2 \times (r_3 + \dots (r_k) + \gamma^k \times (r_{k+1})))) \\ &= r_1 + \gamma \times (r_2 + \gamma^2 \times (r_3 + \dots (r_k)))) \end{aligned} \tag{3}$$

Now, assume, that at each level of the tree, we get the maximum value that a reward can be, call this R_{max}. So for all i, $1 <= i <= k$, $r_i = R_{max}$. Then, we have the following:

$$V_k + 1^{*}(s) = R_{max} + \gamma \times (R_{max} + \gamma^2 \times (R_{max} + \dots (R_{max} + \gamma^k \times (R_{max}))))) \tag{4}$$

$$V_{k}^{*}(s) = R_{max} + \gamma(R_{max} + \gamma^2 \times (R_{max} + \cdots + \gamma^{k-1} \times (R_{max})))) \tag{5}$$

Note, this is the interesting case, because we want to ensure that our value iteration algorithm still converges even if we are receiving maximum value of rewards for each state. If we show convergence for this case, then we need not worry about any other case.

So, we now perform the following computation, by observing that all the R_{max} terms cancel, except the innermost R_{max} term inside $V^*_{k+1}(s)$:

$$\begin{aligned} |V_k + 1^*(s) - V^*_k(s)| &= |\gamma \times \gamma^2 \times \gamma^3 \times \ldots \gamma^k \times R_{max}| \\ &= |\gamma^{(1+2+3+\cdots+(k-1)+k)} \times R_{max}| \\ &= |\gamma^{\frac{k\cdot(k+1)}{2}} \times R_{max}| \end{aligned} \tag{6}$$

Since $\gamma < 1$, there exists some large positive integer k such that $\gamma^{\frac{k\cdot(k+1)}{2}} < \epsilon$, for any small real $\epsilon > 0$. Now, define $< V_k(s) >$ to be the sequence of values for a particular state s, assuming we update the value of s at each time step using Eq. (1) according to the Value Iteration Algorithm. We now establish the following crucial bound, to demonstrate $< V_k(s) >$ is actually a Cauchy sequence. Begin by picking a large integer k such that $\gamma^k \times R_{max} \times \frac{1}{1-\gamma} < \epsilon$, for some small real $\epsilon > 0$. Then, for any integers $m, n > k$:

$$\begin{aligned} |V_n - V_m| &= |V_n - V_{n-1} + V_{n-1} - V_{n-2} + V_{n-2} \ldots V_{m+1} - V_m| \\ &<= |V_n - V_{n-1}| + |V_{n-1} - V_{n-2}| + \cdots + |V_{m+1} - V_m| \quad \text{(by Triangle Inequality)} \\ &= R_{max} \times [\gamma^{\frac{(n-1)\cdot(n)}{2}} + \gamma^{\frac{(n-2)\cdot(n-1)}{2}} \cdots + \gamma^{\frac{(m+1)\cdot(m+2)}{2}} + \gamma^{\frac{m\cdot(m+1)}{2}}] \quad \text{(by Eq. (6))} \\ &< R_{max} \times [\gamma^{\frac{(n-1)^2}{2}} + \gamma^{\frac{(n-2)^2}{2}} + \cdots + \gamma^{\frac{(m+1)^2}{2}} + \gamma^{\frac{m^2}{2}}] \\ &< R_{max} \times [\gamma^{n-1} + \gamma^{n-2} + \cdots + \gamma^{m+1} + \gamma^m] \quad (\text{since } x < \frac{x^2}{2} \text{ for large } x) \\ &< R_{max} \times \gamma^m \times [\gamma^{n-m-1} + \gamma^{n-m-2} + \cdots + \gamma^2 + \gamma^1 + 1] \\ &< R_{max} \times \gamma^k \times [\gamma^{n-k-1} + \gamma^{n-k-2} + \cdots + \gamma^2 + \gamma^1 + 1] \quad (\text{since } m > k) \\ &< R_{max} \times \gamma^k \times \frac{1}{1-\gamma} \quad \text{(by sum of geometric series)} \\ &< \epsilon \quad (\text{by our initial choice for } k) \end{aligned}$$

Since our choices for the values of m, n and ϵ were arbitrary, the sequence $< V_k(s) >$ is thus shown to be a Cauchy sequence, which is equivalent to demonstrating that it is convergent. We are thus done.

For the sake of completeness, one notes that to show the convergence of the Value Iteration Algorithm, we need to technically show that the sequence of vectors of state values converges, where each entry in a vector after time step k is the value of a particular state after time step k. We assume that the state space is finite, which means that the vector has finitely many entries. We know that for each entry s in the vector, the individual sequence formed by extracting the values of that particular state s from the sequence of vectors, must converge. This is true for each state s in the vector, and since the vector is finite, the sequence of vectors itself must also converge.