Homework 3
Nishad Gothoskar

1 Bellman Convergence

We will represent a value iteration update as applying I to produce $V_{t+1} = I(V_t)$. This update is as follows:

$$I(V_t)(s) = R(s) + \gamma \max_{a \in A} \sum_{s' \in S} P(s'|a,s)V(s')$$

Let us assume there is some optimal value function V^* such that $V^* = I(V^*)$ because it has already converged (TODO prove existence of this function). First we will prove a lemma regarding the difference between value functions after applying a value iteration.

Lemma 1 For arbitrary value functions V_1, V_2, $\max_{s \in S} |I(V_1)(s) - I(V_2)(s)| \leq \gamma \max_{s \in S} |V_1(s) - V_2(s)|$
Proof

$$\max_{s \in S} |I(V_1)(s) - I(V_2)(s)| \tag{1}$$
$$= \max_{s \in S} |R(s) + \gamma \max_{a \in A} \sum_{s' \in S} P(s'|a,s)V_1(s') - R(s) - \gamma \max_{a \in A} \sum_{s' \in S} P(s'|a,s)V_2(s')| \tag{2}$$
$$= \gamma \max_{s \in S} |\max_{a \in A} \sum_{s' \in S} P(s'|a,s)V_1(s') - \max_{a \in A} \sum_{s' \in S} P(s'|a,s)V_2(s')| \tag{3}$$
$$\leq \gamma \max_{s \in S} \max_{a \in A} |\sum_{s' \in S} P(s'|a,s)(V_1(s') - V_2(s'))| \tag{4}$$
$$\leq \gamma \max_{s \in S} \max_{s' \in S} |(V_1(s') - V_2(s'))| = \max_{s' \in S} |(V_1(s') - V_2(s'))| \tag{5}$$

To clarify some of the steps taken above: 4 to 5 can be done because taking the absolute difference between the max of two functions individually is bounded by the max of the differences. The proof of this fact derives from $f(x) \leq |f(x) - g(x)| + g(x)$. Step 5 to 6 can be taken because all the probabilities will sum to exactly 1 (with any choice of a) therefore the sum outputs a weighted average of $|(V_1(s') - V_2(s'))|$ across all the states. This weighted average is clearly bounded by the maximum value the function can take.

Now, with the above Lemma we see that since the iterator does not effect the optimal value function:

$$\max_{s \in S} |I(V_1)(s) - V^*(s)| \leq \gamma \max_{s \in S} |I(V_1)(s) - V^*(s)|$$

This implies that with each step of value iteration we reduce the maximum difference between our value function and the optimum by a factor of γ. Let us call our series of functions (by applying the iterator) $f_1, f_2, f_3, ...$ Then we can prove pointwise convergence to the optimal function since $0 < \gamma < 1$.

$$\lim_{n \to \infty} |f_n - V^*| \leq \lim_{n \to \infty} \gamma^n |V_1(s) - V^*(s)| = 0$$

2 k-step look ahead vs. 1-step look ahead

The proof follows from the above result. I_k, the iterator looking k steps ahead, is simply applying the operate I k times. Looking at I_2 makes this clear

$$I_2(V)(s) = R(s) + \gamma \max_{a \in A} \sum_{s' \in S} \big(P(s'|a,s)\big(R(s') + \gamma \max_{a' \in A} \sum_{s'' \in S} P(s''|a',s')V(s'')\big)\big)$$

Since in Part 1 we proved that applying the operator will reduce the max difference between the current value function and the optimum by a factor of γ, looking k steps ahead will reduce this difference by a factor of γ^k at each iteration allowing use to converge to the optimal value function faster.