CS 294-40 Problem Set 1
Due Tuesday September 23 *before* lecture

Please refer to the class webpage (http://inst.eecs.berkeley.edu/~cs294-40/fa08/) for the homework policy.

1 Approximate Bellman back-ups

Suppose F is a γ-contraction with respect to the norm $\|\cdot\|$ and has fixed point V^*. Suppose $\bar{F}$ satisfies $\|FV - \bar{F}V\| \leq \varepsilon$ for all V and $\bar{F}^k V \to \bar{V}$. Show that

$$\|V^* - \bar{V}\| \leq \frac{\epsilon}{1-\gamma}.$$

2 Gauss-Seidel back-ups

Define the operator T_s such that

$$(T_s V)(s') = \begin{cases} (TV)(s') & \text{if } s' = s \\ V(s') & \text{otherwise.} \end{cases}$$

Consider $V_{k+1} = T_n T_{n-1} \ldots T_2 T_1 V_k$. Prove that $V_k \to V^*$.

3 Policy performance when using an approximate value function

Assume we have an approximate value function $\hat{V}$ which satisfies: $\|\hat{V} - V^*\|_\infty \leq \varepsilon$. Consider the policy $\pi = (\mu, \mu, \ldots)$ which chooses actions according to the following greedy strategy with respect to $\hat{V}$:

$$\mu(s) \in \arg\max_{a \in \mathcal{A}} R(s) + \gamma \sum_{s'} P(s'|s,a)\hat{V}(s').$$

(a) Show that when executing this policy π our performance loss relative to the optimal policy can be bounded. In particular, show that

$$\|V_\pi - V^*\|_\infty \leq \frac{2\varepsilon\gamma}{1-\gamma}. \quad (1)$$

(b) Give an example where the bound is tight. (Hint: there is an example with a two-state MDP.)

4 Receding horizon policy

In the previous question, we considered the greedy policy with respect to the approximate value function $\hat{V}$. Now let's consider what happens when we use a policy π that is greedy with respect to an $H+1$-step lookahead. In particular, we have $\pi = (\mu, \mu, \ldots)$ such that:

$$\mu(s) \in \arg\max_{a \in \mathcal{A}} R(s) + \gamma \sum_{s'} P(s'|s,a)(T^H \hat{V})(s').$$

Again assume $\|\hat{V} - V^*\|_\infty \leq \varepsilon$. Show a stronger performance guarantee for V_π than the one given in Eqn. (1).

5 Finite number of iterations in value iteration

Consider we run value iteration as follows:

1. Initialize $V = 0$.
2. $\bar{V} \leftarrow TV$.
3. If $\|\bar{V} - V\|_\infty \leq \varepsilon$ then exit, otherwise set $V = \bar{V}$ and go back to Step 2.

(a) Provide a bound on the number of iterations in terms of $\gamma, \epsilon, \|R\|_\infty$.

(b) Provide a bound on $\|V - V^*\|_\infty$ for V when the algorithm exits.