EC 710, HOMEWORK 1

1. Consider a discounted Markov Decision Process (MDP) with finite state space $\mathcal{S}$, finite action space $\mathcal{A}$, transition probabilities $P(s' \mid s, a)$, reward function $r(s, a)$, and discount factor $\gamma \in (0, 1)$. Let π be any policy with corresponding value function V^π. Define a new policy π' that, for every state s, satisfies the following policy improvement condition:

$$r\big(s, \pi'(s)\big) + \gamma \sum_{s' \in \mathcal{S}} P\big(s' \mid s, \pi'(s)\big) V^\pi(s') \geq r\big(s, \pi(s)\big) + \gamma \sum_{s' \in \mathcal{S}} P\big(s' \mid s, \pi(s)\big) V^\pi(s').$$

Prove that for all $s \in \mathcal{S}$, the value functions satisfy

$$V^{\pi'}(s) \geq V^\pi(s).$$

2. Consider inexact policy iteration where at step k you compute a vector $\tilde{V}^{\pi_k}$ which satisfies

$$||V^{\pi_k} - \tilde{V}^{\pi_k}||_\infty \leq \epsilon_k.$$

Prove that if $\sum_k \epsilon_k < \infty$, then V^{π_k} converges to V^*, the true optimal value vector. You may need to look up the notion of a Cauchy Sequence.

3. In class, we considered fixing the time k and letting the time horizon N go to infinity for the noiseless system $x_{k+1} = Ax + Bu_k$. We showed that the optimal policy approaches $u_k = -Lx_k$. Show that this implies that the policy $u_k = -Lx_k$ is optimal for the infinite-horizon problem.

4. Go the lecture slides, and under the slide ”Noiseless Case” in the LQR lectures, you will see a claim that K_k converges. Prove that it converges exponentially, i.e., $||K_k - K||_2 \leq C\rho^{N-k}$ for some $\rho \in (0, 1)$.