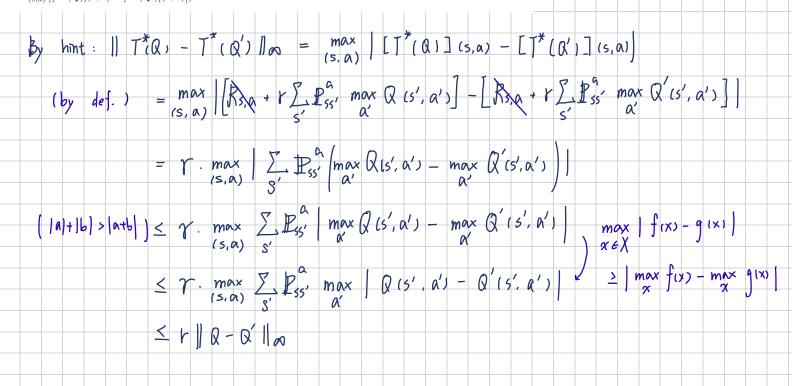


 $Q^*(s,a) = R_{s,a} + \gamma \sum_{a'} P_{ss'}^a \left(\max_{a'} Q^*(s',a') \right)$ (3)

Similar to the standard Value Iteration, we can also study the Q-Value Iteration by defining the Bellman optimality operator $T^*: \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \to \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ for the action-value function: for every state-action pair (s, a)

$$[T^*(Q)](s,a) := R_{s,a} + \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q(s',a')$$
(4)

Show that the operator T^* is a γ -contraction operator in terms of ∞ -norm. Please carefully justify every step of your proof. (Hint: For any two action-value functions Q, Q', we have $\|T^*(Q) - T^*(Q')\|_{\infty} = \max_{(s,a)} \left| [T^*(Q)](s,a) - [T^*(Q')](s,a) \right|$)



Problem 2 (Regularized MDPs)

(10+10=20 points)

In Lecture 4, we formally describe the regularized MDP, which is a direct extension of the classic MDP with a regularizer Ω . In this problem, for simplicity, suppose we use the Shannon entropy as our regularizer, i.e., $\Omega(\pi(\cdot|s)) \equiv H(\pi(\cdot|s)) := -\sum_{a \in \mathcal{A}} \pi(a|s) \ln \pi(a|s)$. Let us verify a few important properties mentioned in Lecture 4 as follows

(a) Recall that we introduce the "regularized Bellman expectation operator" T_{Ω}^{π} as

$$[T_{\Omega}^{\pi}V](s) := R_s^{\pi} + \Omega(\pi(\cdot|s)) + \gamma P_{ss'}^{\pi}V. \tag{5}$$

Please verify that T_{Ω}^{π} is a contraction operator in L_{∞} norm. (Hint: Try to extend the proof procedure of the contraction property of T^{π} in Lecture 3)

$$\| T_{\Delta_{1}}^{\pi}(V) - T_{\Delta_{2}}^{\pi}(V') \|_{\partial S} = \max_{S} \left[T_{\Delta_{3}}^{\pi}(V') \right](s) - \left[T_{\Delta_{3}}^{\pi}(V') \right](s) \right]$$

$$= \max_{S} \left[T_{SS}^{\pi}(V) + T_{SS}^{\pi}(V') \right](s) + T_{SS}^{\pi}(V') + T_{SS}^{$$

