MAE 288A
Assignment 4
Due 9pm, 30 May

Problems to hand in. (Not all problems may be graded.)
All work and explanations must be included for full credit.
In the case of problems where code was used, include the code.
We can only grade matlab and python.

1. (5) Create a transition matrix for a three-state, irreducible, aperiodic Markov chain. At least one element of the matrix must be 1/2. Prove that it is both irreducible and aperiodic. Generate a stationary distribution for the Markov chain, without taking a limit to obtain it, and show that it is stationary.

2. (5) Consider a Markov chain with transition probability matrix given by
$$\mathcal{P} = \begin{bmatrix} 0.5 & 0.25 & 0 & 0.25 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0.75 \\ 0 & 0 & 0 & 0.75 & 0.25 \end{bmatrix}.$$
Draw a graph indicating the probability flow between the states. Also indicate which states are recurrent and which are transient? Prove the claims.

3. (5) Let
$$C_B(\mathbb{R}^n;\mathbb{R}) \doteq \{\phi \in C(\mathbb{R}^n;\mathbb{R}) \,|\, \exists K_\phi < \infty \text{ s.t. } |\phi(x)| \le K_\phi \;\forall x \in \mathbb{R}^n\}$$
Let $C_B(\mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^m;\mathbb{R}^n)$ be defined similarly. Let $U \subset \mathbb{R}^k$ be compact, and let $w \sim \mathcal{N}(0;Q)$ taking values in $\mathbb{R}^m$. Let $f \in C_B(\mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^m;\mathbb{R}^n)$, Define the operator, $\mathcal{G} : C_B(\mathbb{R}^n) \to C_B(\mathbb{R}^n)$ by
$$\mathcal{G}[\phi](x) \doteq \min_{v\in U} \mathbf{E}\big[\phi(f(x,v,w))\big].$$
Is $\mathcal{G}$ monotonic? If so, prove it; if not, provide a counterexample.

4. (10) Consider the gambling problem discussed in class with the probability of winning each coin toss being $p = 0.47$ (and the probability of losing being $q = 1 - p = 0.53$). Suppose the gambler bets \$1 at each time, that the gambler starts with $g_0 = \$2$, and that the opponent starts with $o_0 = \$4$. If either person goes bankrupt, they stay bankrupt for all time; otherwise the bet remains at \$1 at each time. By repeatedly updating p_t until one gets near convergence, estimate the probabilities of the player eventually going bankrupt and of the opponent eventually going bankrupt. (Three digits of accuracy is sufficient.) What is the expected payoff to the gambler? Obtain the same information in the case where the gambler bets the maximum allowable (such that neither player can ever have a negative balance) at each step. Again estimate the probabilities of each player going bankrupt and the expected payoff to the gambler. Should the gambler bet \$1 each time, or the maximum allowable amount?

   Now, vary o_0 from \$1 to \$50, and compare the expected payoffs to the gambler for each of the two control policy options.

5. (10) Consider a Markov chain control problem where $\mathcal{X} = \{1, 2, 3\}$ and $U = \{a, b, c, d\}$. Let the transition probability matrices be given by

$$P(a) = \begin{bmatrix} 3/4 & 1/8 & 1/8 \\ 1/2 & 1/2 & 0 \\ 0 & 3/4 & 1/4 \end{bmatrix}, \quad P(b) = \begin{bmatrix} 3/4 & 0 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 3/4 \end{bmatrix},$$
$$P(c) = \begin{bmatrix} 2/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}, \quad P(d) = \begin{bmatrix} 3/4 & 1/8 & 1/8 \\ 1/3 & 1/6 & 1/2 \\ 1/8 & 3/8 & 1/2 \end{bmatrix}.$$

   For simplicity of exposition here, let the running cost be given in vector form, where the three components correspond to the three possible states. That is, $l_i(v)$ will denote the cost for control v when the state is i. (N.B.: this is different from the notation form we're using in lecture/analysis.) We take

$$l(a) = \begin{pmatrix} 3 \\ 2.5 \\ 5 \end{pmatrix}, \quad l(b) = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \quad l(c) = \begin{pmatrix} 0.8 \\ 2.8 \\ 4.2 \end{pmatrix}, \quad l(d) = \begin{pmatrix} 9 \\ 2 \\ 4.1 \end{pmatrix}.$$

   Let the discount factor be $\alpha = 0.6$. Solve the problem by value iteration, starting with $V_i^0 = 0$ for all $i \in \mathcal{X}$. Run the iteration until

$\|V^n - V^{n-1}\| < 0.1$, where $\|\cdot\|$ indicates Euclidean norm here. Indicate the value function approximation at each step. Also, indicate the minimizing control generated by the algorithm at each step, i.e., the argmin values that are obtained during the computation of each V^n.

6. (10) Solve this same problem (i.e., problem 5) using policy iteration. You may begin policy iteration by starting either with a policy guess or a value-function guess. For consistency over all submissions, please begin with policy $\bar{u}^0$ given by $\bar{u}^0(i) = c$ for all $i \in \mathcal{X}$.