Thus, $p(s_1) = 0$.

The chain is irreducible and hence, $s_1$ and $s_2$ are recurrent.

The chain is reducible and hence, $s_1$ and $s_2$ are transient.

Solutions to Homework 8
But this contradicts the assumption that all deterministic policies
\[ p_i^*(u) = 0 \quad \text{for all} \quad u \notin T \quad \text{(where } \text{and } \nu \text{ defined as above)} \]
set \( \epsilon = \alpha \) for which \( p(D) = \alpha \). Thus \( \nu(v_i) = 0 \quad \forall u \leq 0 \).
Consider a deterministic decision rule \( \delta \) as follows: For all
\[ \delta(\mu) \text{ and } \delta(S) \text{, thus } \nu(v_i) = 0 \quad \forall u \leq 0 \]
3) The problem could be modelled as a Markov decision process with state space $S = \{s_0, s_1, s_2, s_3\}$ where $T$ denotes termination.

Let $0$ denote the ongoing continuing and $1$ denote the action of leaving.

Let $A_{s} = \begin{bmatrix} 0.0 & 0.3 & 0.2 \end{bmatrix}$ denote the action

Let $b(0) = \begin{bmatrix} 0 \end{bmatrix}$, $b(1) = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$ and $b(2) = \begin{bmatrix} 0.5 & 0 \end{bmatrix}$

Let $P = \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$

Let $v_{0} = \begin{bmatrix} 0 \end{bmatrix}$, $v_{1} = \begin{bmatrix} 0.0 \end{bmatrix}$, $v_{2} = \begin{bmatrix} 0.0 \end{bmatrix}$ and $v_{3} = \begin{bmatrix} 0.0 \end{bmatrix}$

$V_0 = v_0 + P \cdot v_0 \Rightarrow V_0 = 0$

$V_1 = v_1 + P \cdot v_1 \Rightarrow V_1 = 2.1335$

$V_2 = v_2 + P \cdot v_2 \Rightarrow V_2 = 26.5000$

$V_3 = v_3 + P \cdot v_3 \Rightarrow V_3 = 29.2013$

$V_0 = 0$, $V_1 = 2.1335$, $V_2 = 26.5000$, $V_3 = 29.2013$, we have $d_1 = d_0 = 0$

Computing $d_1$, we see $\max\{f_2 + P \cdot v_2\}$, we have $d_1 = d_0 = 0$

Thus, $d = d_1 = \begin{bmatrix} 0 \end{bmatrix}$. The policy iteration terminates.