Assignment 2

1. Transition Table

S ⟶ state function A ⟶ Action P ⟶ Probability

Transition function ⟶ $T: S \times A \longrightarrow \{S \times P\}$

State	Action	Resulting state	Probability
A	R	B	0.8
A	R	A	0.2
A	U	C	0.8
A	U	A	0.2
B	L	A	0.8
B	L	B	0.2
B	U	R	0.8
B	U	B	0.2
C	R	R	0.25
C	R	C	0.75
C	D	A	0.8
C	D	C	0.2

2. The best path would be A ⟶ B ⟶ R

- The other path is A ⟶ B ⟶ R which even though has a better best case reward. It is unlikely due to the low probability the path will have due to:

$$T(C, R, R) = 0.25$$

Which makes the original path better on an average.

3. Initial state:

0	16.5
0	0

1st iteration:

A: $R \rightarrow 0.8(-1 + 0.2 \cdot 0) + 0.2(-1 + 0.2 \cdot 0)$
$= -1$

$U \rightarrow 0.8(-1 + 0.2 \cdot 0) + 0.2(-1 + 0.2 \cdot 0)$
$= -1$

B: $L \rightarrow 0.8(-1 + 0.2 \cdot 0) + 0.2(-1 + 0.2 \cdot 0)$
$= -1$

$U \rightarrow 0.8(-4 + 0.2 \cdot 16.5) + 0.2(-1 + 0.2 \cdot 0)$
$= -0.76$

C: $R \rightarrow 0.25(-3 + 0.2 \cdot 16.5) + 0.75(-1 + 0.2 \cdot 0)$
$= 0.675$

$D \rightarrow 0.8(-1 + 0.2 \cdot 0) + 0.2(-1 + 0.2 \cdot 0)$
$= -1$

A: -1, B $= -0.76$, C $= 0.675$

2nd iteration

A: R → $0.8(-1 + 0.2(-0.76)) + 0.2(-1 + 0.2(-1)) = -1.1616$

U → $0.8(-1 + 0.2(-0.675)) + 0.2(-1 + 0.2(-1)) = -1.148$

B: L → $0.8(-1 + 0.2(-1)) + 0.2(-1 + 0.2(-0.75)) = -1.190$

U → $0.8(-4 + 0.2(16.5)) + 0.2(-1 + 0.2(-0.675)) = -0.790$

C: R → $0.25(-3 + 0.2(16.5)) + 0.75(-1 + 0.2(-0.675)) = -0.776$

D → $0.8(-1 + 0.2(-1)) + 0.2(-1 + 0.2(-0.675)) = -1.187$

A = -1.148 B = -0.790 C = -0.776

3rd iteration

A: R → $0.8(-1 + 0.2(-0.790)) + 0.2(-1 + 0.2(-1.148)) = -1.172$

U → $0.8(-1 + 0.2(-0.776)) + 0.2(-1 + 0.2(-1.148)) = -1.170$

B: L $\longrightarrow$ $0.8(-1+0.2(-1.148))+0.2(-1+0.2(-0.790))$
$= -1.215$

U $\longrightarrow$ $0.8(-4+0.2(16.5))+0.2(-1+0.2(-0.790))$
$= -0.791$

C: R $\longrightarrow$ $0.25(-3+0.2(16.5))+0.75(-1+0.2(-0.776))$
$= -0.791$

D $\longrightarrow$ $0.8(-1+0.2(-1.148))+0.2(-1+0.2(-0.776))$
$= -1.214$

$A = -1.170 \quad B = -0.791 \quad C = -0.791$

4th iteration

A: R $\longrightarrow$ $0.8(-1+0.2(-0.791))+0.2(-1+0.2(-1.170))$
$= -1.17346$

U $\longrightarrow$ $0.8(-1+0.2(-0.791))+0.2(-1+0.2(-1.170))$
$= -1.17343$

B: L $\longrightarrow$ $0.8(-1+0.2(-1.170))+0.2(-1+0.2(-0.791))$
$= -1.218$

U $\longrightarrow$ $0.8(-4+0.2(16.5))+0.2(-1+0.2(-0.791))$
$= -0.791$

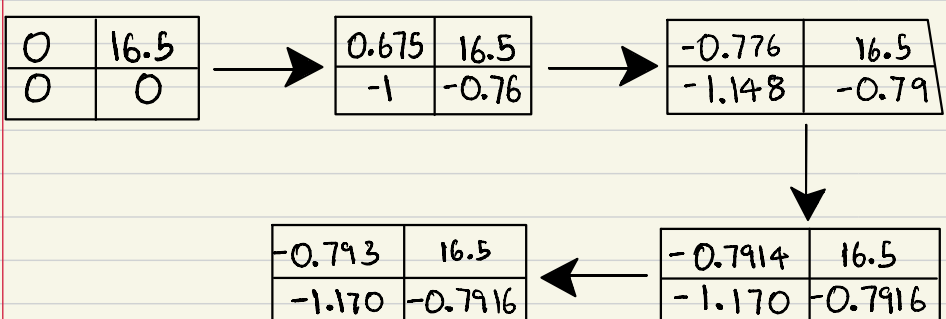
C: $R \longrightarrow 0.25(-3 + 0.2(16.5)) + 0.75(-1 + 0.2(-0.791))$
$= -0.793$

$D \longrightarrow 0.8(-1 + 0.2(-1.17)) + 0.2(-1 + 0.2(-0.791))$
$= -1.218$

$A = -1.172 \quad B = -0.791 \quad C = -0.847$

Since the new value differs in less than the bell error, we can stop the iteration.

Transition



4. The optimal path will be based on the values that have been marked in blue.

So the optimal path will be $A \longrightarrow C \longrightarrow R$

Does not match the predicted path due to the effect of discount which faded the reward (terminal).

Impact

- Changing probability for C & making it lower will change the optimal path to

$$A \longrightarrow B \longrightarrow R$$

- Reducing the cost of $B \longrightarrow R$ will do the same
- Increasing the discount will diminish the discount effect which again can change the optimal path.

Insights

- It can be noticed that $V(B) > V(C)$ even though the optimum path is from C. this is the case because of the discount factor & the terminal reward value.

$\longrightarrow$ Bringing changes to either can divert the path