	20	2019101056				Asyan Jain Page:				
				MDL	Assgn 2 Part I					
H	ichu!	At either of the state (A/B/C/R) we can have 4 possible moves which are up (v), Down (								
_		4	possible	moves	wh	ch.	are	Up (U)	, Down (2)	
_		Lef	t(L),	Right (F	2).		V081 31	STEACH C		
-	A 15. 14		1	The same of	3 .3h	The sale	2.000	Tayle Janks	1	
+	Tol	<u>s))</u> =	•	A	C TO	1 B	Wastin.	C	R	
-	From (	).	LR	UD	L	R	UP	LRUD	LRUD	
-	4	1						41 3000	1	
-	A		0 0.2	0.2 0	0	0.8	00	0 0 080	0000	
-	Jan all	3	0.8	0 0	0.2	0	0.20	0000	00080	
1		C	00	0 0.8	0	0	00	0 0 75 0 0 2	0 0.25 0 0	
1	Here each value suppresents P(s' s,a) i.e. probability of suaching a state s' forom initial									
1										
	given initial state & and given action a.									
1	TO TOTAL STATE OF	0	· IT Y			Mary 1	0	man makes	1	

has different possibilities from A + R druiz (noing A -> B -> R A + C>R ( these will add on some A + B - C -> R I extera cost than A - C - B - R above two. path will be choosen from A+B+R :. Best Given P(B - R)= Step Cost (B + R) - - 4 P(B - R) = 0.8 P(c→R)=0.25 Step(cos+(C→R)=-3 :. Expected utility (B-R)= 0.8x (-4)= -3.2 Expected utility (C+R)= 0.25x(-3)= -0.75 : Optimal sol's follows MEU principle By MEU principle (Marimum Expected Utility)
we can say A - C = R is the optimal path

Page:	RWRMRH
Date:	11

un							
	R= Arr [ 2019101056 1.15]= . Arr [1] = 9						
GLEEN,	· P Roward Value (R)=9.						
	To perform the value iteration algorithm  we need to apply Bellman equation viz.  Util [I] = max[R[I,A] + X \( \frac{1}{2} \) P[J] I,A] (1)						
in the same of	we need to apply Bellman equation viz.						
\s_\s_\s_\s_\s_\s_\s_\s_\s_\s_\s_\s_\s_\	U, (I) = max[ R/IA) + x < P[J] I, A] (*(J))						
	Also four $t = 0$ $V_0(A) = 0$ $V_0(B) = 0$ $V_0(c) = 0$ $V_0(B) = 0$						
	0 9 - Uo (c)=0 Uo (g)=09						
<u> </u>	Let's call this as Destination D						
t=0	a) For state 'A': R, U are only possible rection U, (A) - R(A,R) + & [P(B A,R). Uo(B)+P(A A,R).Uo(B) + P(A A						
1st ites	rection U, (A) T R(A,R) + & LP(B A,R). Uo(B)+P(A A,R).Uo(A)						
	R(A, U) +8 PC, A(1) = U0(C)+P(A A(1).U0(A)						
93.	Tallabolo meta sti make salk measurat of						
VIXIV.	U, (A) T -1+0.2[0.8×0+0.2×0] · mou1+0.2[0.8×0+0.2×0]						
(E) 1U.1	-1+ 0.2 [0.8 X O + 0.2 X O]						
	$\therefore U_1(A) = -1$						
	1 ) - I O V O V O V O V						
	b.) For state B': 1, U alle 2 possible						
Lisex	actions.						
	1 10 110 1 10 10 10 10 10 10 10 10 10 10						
	U, (B) = R(B, L) + 8 LP(A B, L). U. (A) + P(B B, L). U. (B)						
	U, (B) = R(B, L) + 8[P(A B,L). U. (A) + P(B B, L). U. (B)] man: R(B, U) + 8[P(D B,U). U. (D) + P(B B, U). U. (B)]						
	To calculate R(B, U): The reward associated with Action = "U' and initial state = 'B' is						
	Fo calculate R(B,U): The reward associated						
	with Action = "I'V' and initial state= 'B'is						
	also dependent on the destination state. : $R(B, U) = \sum_{i} R(B, U, J) \cdot P(J B, U)$						
	:. R(B,U) = ER(B,U,J). P(J B,U)						
	J						

Page:

Date: / /

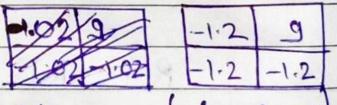
Date:

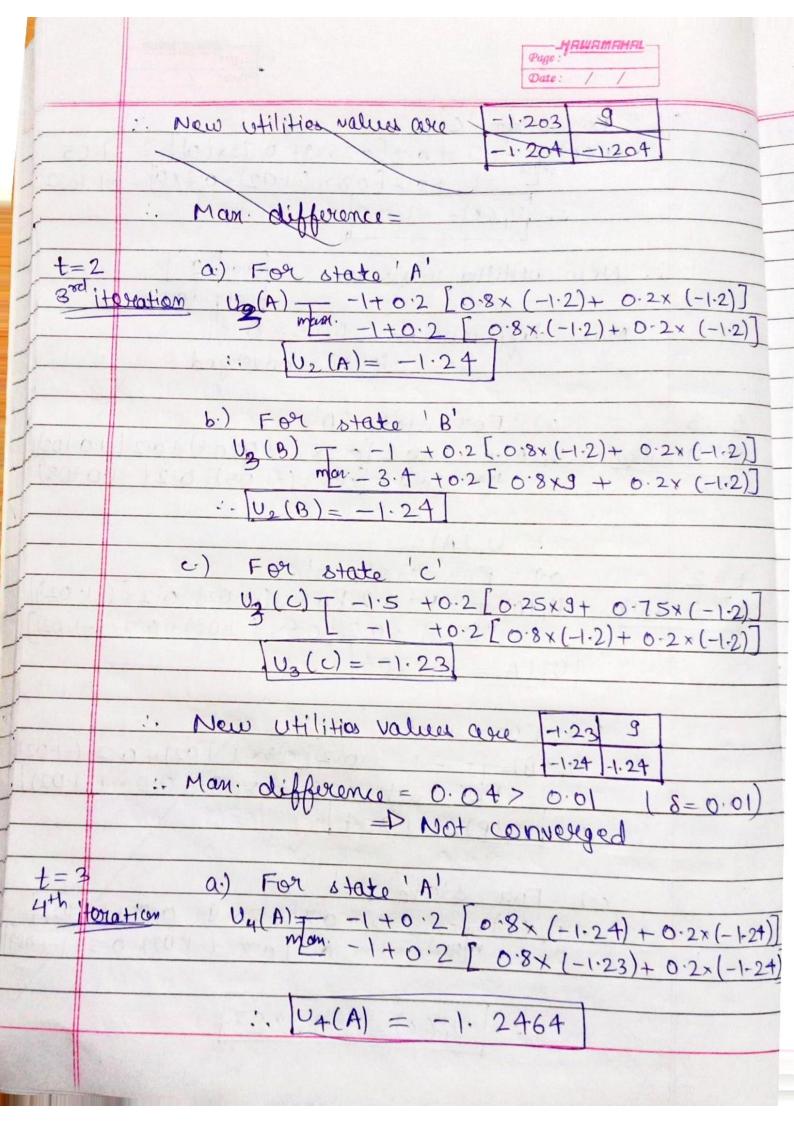
b) For	state	'B'	1	40.819	-1.1
U2 (B) T	-1+0.2	[0.8x(	-1)+ 6	3.2x (-1)	= -
me	-1+0.2 -3.4+	0.2 0.	8×9+	0.2× (-1)	J = -2.00
Tour-	1, (8)	1.02	102 (B	)=-1.2	
		1	4.18 h		
c) For	atata	161			1.0

c) For state 'c'

$$U_{2}(c) = -1.5 + 0.2 [0.25 \times 9 + 0.75 \times (-1)] = 0$$
 $0.2 \times (-1) + 0.2 \times (-1) + 0.2 \times (-1) = 0.2 \times (-1$ 

utilities values are New





Page:

Date: / /

For blate B' + 0.2 [0.25x0 + 0.75x(-1.24)] For state 'B' U4 (B) T -1+ 0.2 [0.8x (-1.24) + 0.2x (-1.24)]. U4 (B)=-1.2+8  $U_{4}(c) = -1.5 + 0.2 \left[0.25 \times 9 + 0.75 \times (-1.23)\right]$   $V_{4}(c) = -1.2345$ c.) For state 'c' :. New utility values are -1.2345 9 .. Man. difference = 0.0064 < 0.01 -> Converged (At 4th iteration) And I Now, we seen the algorithm, one last time to determine the policy. Clearly Clearly Us (A UP Right

(De > D From A you should take an UP action

(By comparison) 0=-1.249536 0=-1.247376 U5(B) -> calculating it will be useless.

Page:

Date: / /

Us(c) T -1.5+0.2[0.25xg+0.75x(-1.2345) - 1+0.2[0.8x(-1.2464)+0.2x Us(c) T - 1.235175 - Right 1.235175 - 1.248804 + Douen - From C u should move Right Us(c)=-1.235175

DIA -> C-> R is optimal bath (Which we also gressed in Ques.2) And 5 As the sew and value in creases the optimal path  $A \rightarrow C \rightarrow R$  starts changing to  $A \rightarrow B \rightarrow R$ . This is because now we have a high reward so the algorithm will tend to more risk seeking methodology.

(Means more risk can be taken).