CS-747 Assignment 4 Report
Arun Verma, 154190002

NOTE about given MDP: For given configuration of MDP, the value for each state is ***zero***. Hence, the solution will be any scalar multiple (C) of [w_1 w_2 w_3 w_4 w_5 w_6 w_7] = **C * [1 1 1 1 1 4 -2]**. The given problem is linear function approximation as value of state is linear combination of features.

ANSWER 1: A graph between estimated expected value of all states v/s number of updates is shown in ***Figure 1*** for ***Experiment 1*** for given parameters i.e., learning rate, α = 0.001, discounting factor, γ = 0.99, Number of updates, N = 1000000, for state *i,* initial $V(s_i)$ = 1. As, we know TD(0) algorithm converges for linear function approximation problems. Therefore, for given MDP with increasing number of updates, TD(0) is converging the value of each state to ***zero*** (as the actual value).

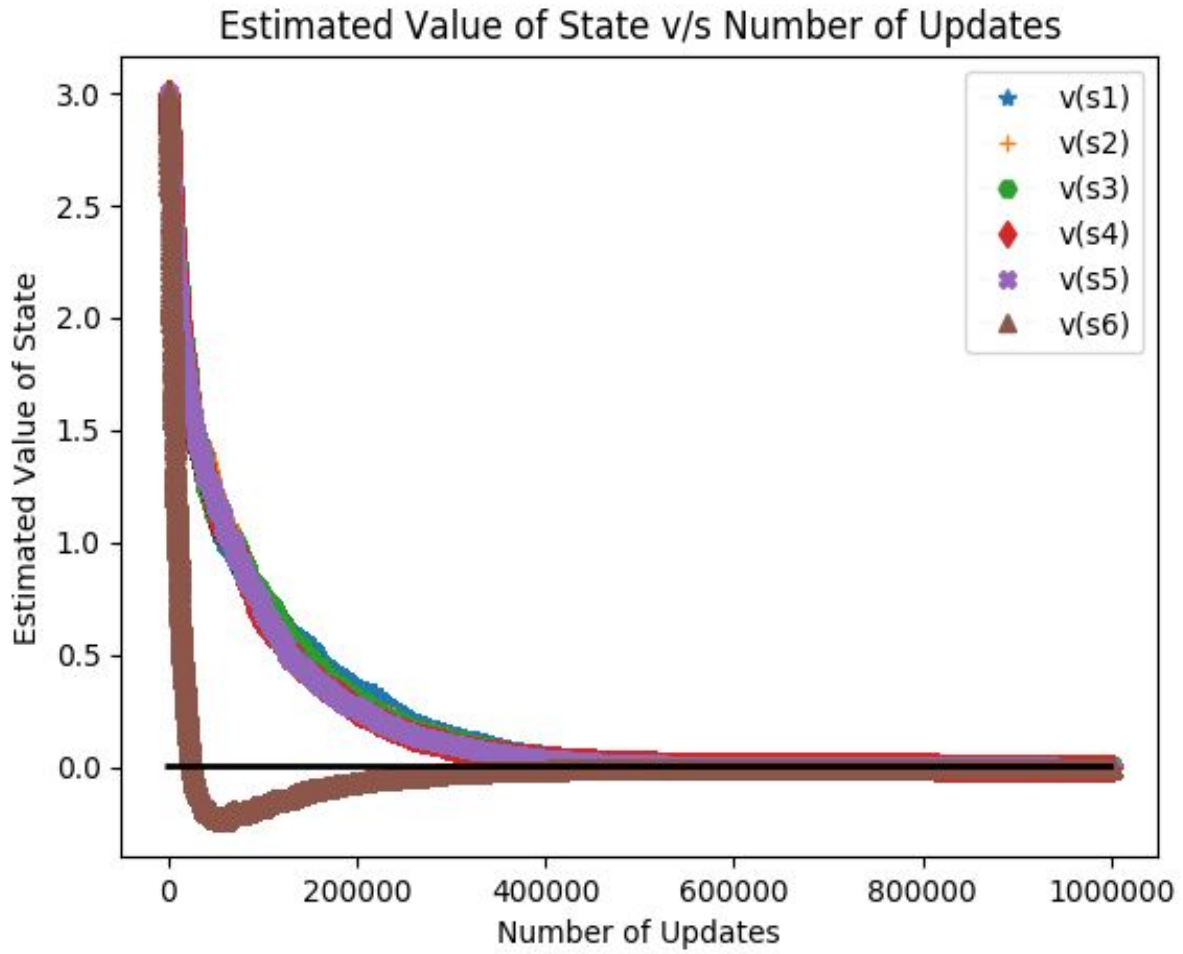

Figure 1

ANSWER 2: A graph between estimated expected value of all states v/s number of updates for different λ values {0, 0.2, 0.4, 0.6, 0.8, 1.0} is shown in ***Figure 2*** for ***Experiment 2*** with given parameters i.e., learning rate, α = 0.001, discounting factor, γ = 0.99, Number of updates, N = 1000000, for state *i,* initial $V(s_i)$ = 1. As, we know TD(λ) algorithm converges for linear function approximation problems. Therefore, for given MDP with increasing number of updates, TD(λ) is converging the value of each state to ***zero*** (as the actual value).

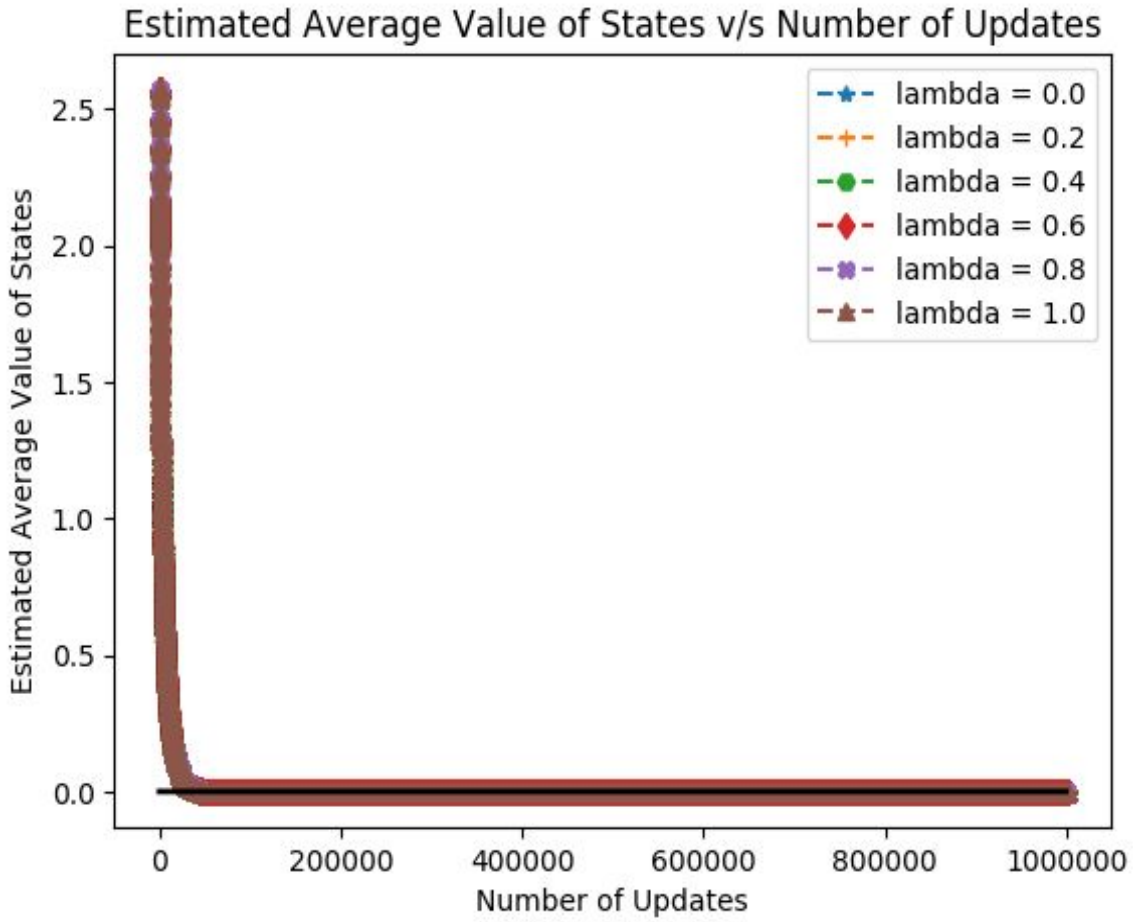

Figure 2

ANSWER 3: I have run ***Experiment 2*** with 10 different initial random values for the weights keeping λ = 0 and other parameters same as given in *ANSWER 2.* The second column of **Table 1** represents initial random weights, third column represents final updated weights and fourth column represents estimated value of the states. From the ***Table 1*** results, following observations are made:

1. Each time process converges to different weights but they are nearly scalar multiple of each others.
2. Similarly, value function of states also has different values but some of them nearly close to each other. For given MDP, all values will converge to zeros as N→∞ (from theoretical results).

Sr. #	Initial W = [w(1) w(2) w(3) w(4) w(5) w(6) w(7)]	Final W = [w(1) w(2) w(3) w(4) w(5) w(6) w(7)]	V = v(s1) v(s2) v(s3) v(s4) v(s5) v(s6)
1	[0.88864559 0.32399596 0.1794751 0.77835965 0.67214691 0.90946595 0.17446242]	[0.24534673 0.24516708 0.24509888 0.24534459 0.24531395 0.9810592 -0.49052707]	0.000166396993154 -0.000192911046048 -0.000329301165134 0.00016211093501 0.000100825732048 5.06249837273e-06
2	[0.97393073 0.76768771 0.51612167 0.32484304 0.62793004 0.47154895 0.79683337]	[0.14019683 0.14018102 0.14009779 0.14000851 0.14014602 0.56048225 -0.28024155]	0.000152111007483 0.000120493794295 -4.59728858522e-05 -0.000224518524135 5.04854219611e-05 -8.47364826084e-07

3	[0.74351545 0.57446728 0.64544299 0.12159966 0.15854816 0.89337747 0.44674592]	[0.19710998 0.19699914 0.19699097 0.19680901 0.19685206 0.78776512 -0.39388497]	0.000334997846818 0.000113317250043 9.69782030505e-05 -0.000266944139925 -0.00018085154125 -4.81598870616e-06
4	[0.81673847 0.7478347 0.10602496 0.68996683 0.16534387 0.90769264 0.12822786]	[0.23611816 0.23606939 0.23590478 0.23605474 0.23591814 0.94403142 -0.47201639]	0.000219921571733 0.000122393840285 -0.000206840216199 9.30861812333e-05 -0.000180102665179 -1.36440259435e-06
5	[0.91512479 0.13967678 0.48387124 0.90746538 0.80565439 0.58277344 0.18822044]	[0.20836843 0.20808136 0.20820643 0.20834405 0.20832211 0.83302403 -0.41651348]	0.000223382386371 -0.000350751161662 -0.000100624462599 0.000174624578454 0.000130741868278 -2.92009518366e-06
6	[0.32965819 0.06285186 0.88361639 0.57516421 0.41523716 0.81110344 0.46969968]	[0.18282881 0.18275377 0.18300374 0.18289885 0.18284753 0.73144083 -0.3657231]	-6.54778134106e-05 -0.00021556590711 0.000284370423185 7.4596954151e-05 -2.80465148582e-05 -5.36948132657e-06
7	[0.95717255 0.10526201 0.97916157 0.10379978 0.52334339 0.47990699 0.95923049]	[0.10699747 0.10662825 0.10691897 0.10665044 0.10679736 0.42718176 -0.21359338]	0.000401569085353 -0.000336874247094 0.000244558048294 -0.000292497851599 1.33976511751e-06 -5.00197008124e-06
8	[0.20122079 0.37597952 0.05231067 0.48756858 0.23811868 0.83031414 0.46009052]	[0.15022195 0.15029375 0.15016333 0.15034323 0.15024054 0.60100108 -0.30050333]	-5.94303261305e-05 8.41728114628e-05 -0.00017666496519 0.000183134223925 -2.22368922838e-05 -5.57733981532e-06
9	[0.25623496 0.01652581 0.58971909 0.34304679 0.34439877 0.91836997 0.57501351]	[0.16290968 0.16283299 0.16301232 0.16293911 0.16298394 0.65173925 -0.32587162]	-5.225678682e-05 -0.000205644508068 0.000153026125406 6.6047925808e-06 9.6260111916e-05 -3.99505149629e-06
10	[0.63514875 0.72446466 0.243566 0.10349691 0.18231997 0.37406715 0.19367576]	[0.11999018 0.12006719 0.11988204 0.11982724 0.11984123 0.47966097 -0.23983081]	0.000149550261316 0.000303565836903 -6.67285433436e-05 -0.000176337191585 -0.000148343078613 -6.52961254788e-07

Table 1