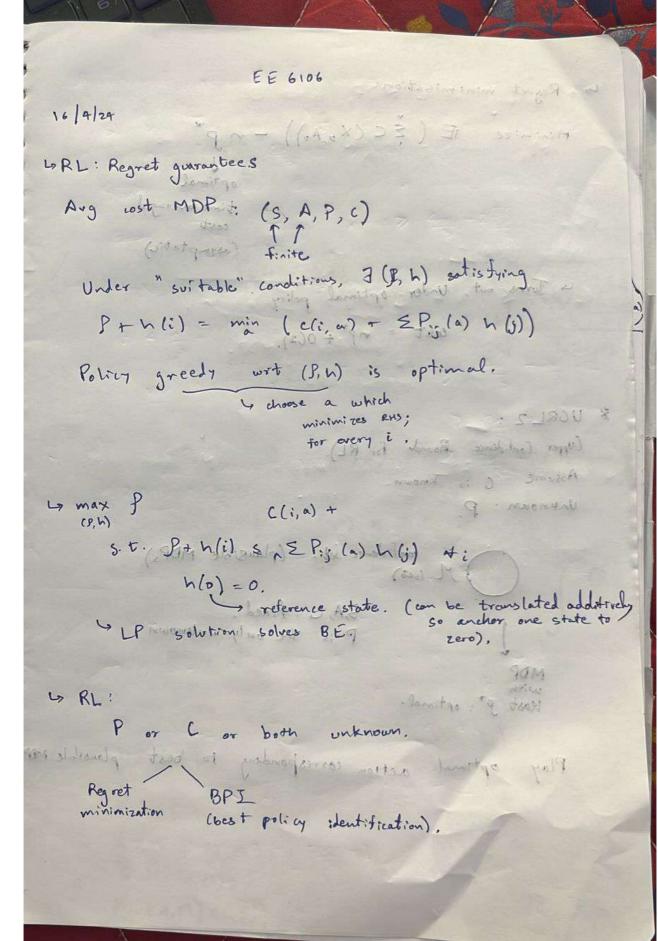


Solving above MDP, solves O. 1 = (0) My: convex polyhedral set in higher dim. Thm: E(Rn) & I + CD(M) SJZAnlogn. diameter of MDP. univ. constant. . 9 am statuely to 7 36 M * in case of BPI, if all maps in M have some action for each state, stop. Sampling: not as trivial. 1999999999 TO THE MEMORY 4 Phase who when Ty (1/4, 40) = 2 Ty (1/4, 1/4). god Holosta for Man Men ich (A(1)) = ((A(1)) 3 13 (MM) = P. (M).



12 Regret minimisation: #5 4 01 Minimize E (EC(Xb,Ab)) - np* optimal legist was the (3 9 A 2) time overaged cost cost (asymptotic) is Turns out, Under optimal policy ((1) ~ (a) east = ng++ oca). Policy greeds with (RW) is optimal. dorder a sound of * UCRL 2: CH2 ex infinite Cupper Confidence Bound For RL) Assume C is known-Unknown: P. + (0,1))) M (set) (plansible MDPs) person of S,A, C. fixed. Pis the only unknown MDP least 9t. optimal. S RIL I or both unknown Play optimal action corresponding to best plansible MD Regard to the policy shoutstration of minimum and the contractions of the contraction of