A2) (a) At the starting of optimistic greedy, we constantly get disappointed and Explore for some untials turns. This happen for all 2000 indivisual 10-asimed bandits. This causes accumulation of bod majority taking good of bad step together. This causes spikes in carly stages.

(b) As we can infer from the graph optimistic, is not good method for non- stationary methods , because after some time this method in the same as greedy which causes no enploration even when the optimal might have changed in non-stationary problem.

A3) As mentioned in the question, taking step sing

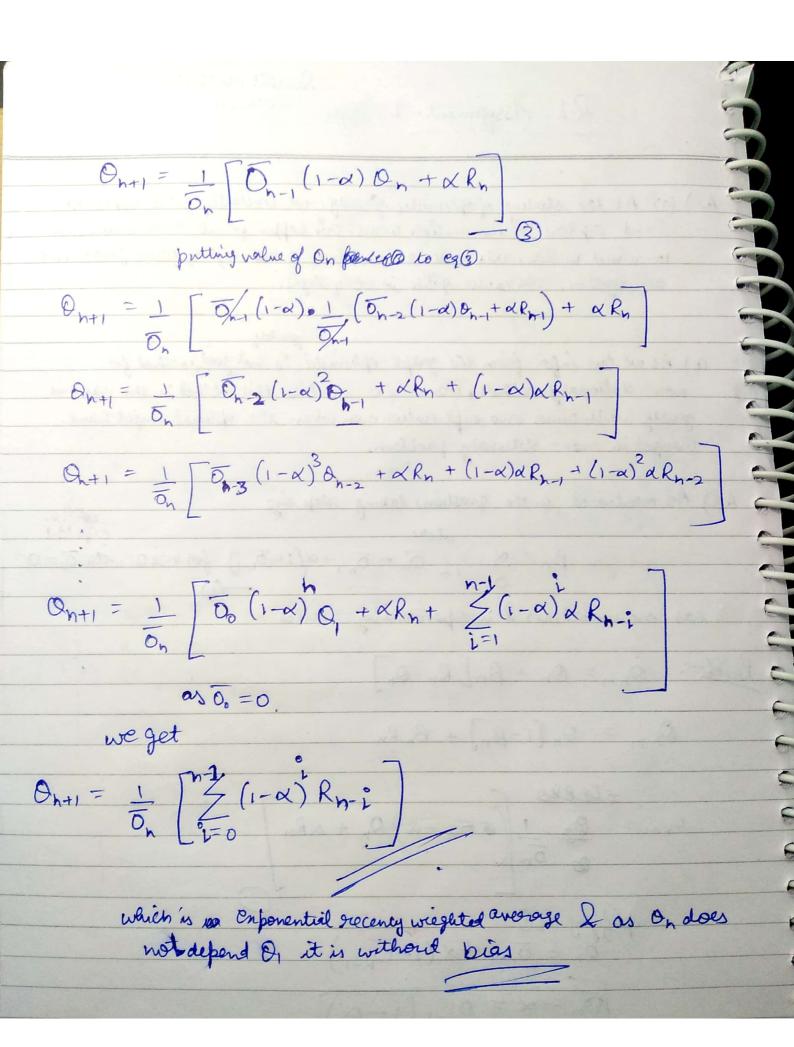
$$\beta_n = \frac{\alpha}{\sigma_n} + \frac{1}{\sigma_n} = \frac{1}{\sigma_{n-1} + \alpha(1 - \sigma_{n-1})}$$
 for  $n \ge 0$  with  $\sigma_0 = 0$  has no initial bias  $\ell$  is exponentially waited

 $O_{n+1} = O_n \left( 1 - \beta_n \right) + \beta_n R_n$ 

$$O_{n+1} = O_{n} - \alpha O_{n} + \alpha R_{n}$$
 $O_{n} + \alpha O_{n} + \alpha O_{n} + \alpha O_{n}$ 

simplifying eq 0  $\overline{O}_{n} = \overline{O}_{n-1} + \alpha \left(1 - \overline{O}_{n-1}\right)$ 

$$\begin{array}{lll}
\overline{On} - \alpha &= \overline{O_{n-1}(1-\alpha)} \\
\text{putting time in } & & & \\
\end{array}$$



A4) In stationary, for 1000 iterations UCB and optimistic perform almost equally a. But UCB performs slightly better teran assessage C-grandy because of better method of emploration

In Non-stationary, UB UCB perform much better than optimistic. Since After some time optimistic becomes normal greedy but assess information gained changes with time hence it is of no-use is after some iterations.

UCB performs slightly better town C-greedy because of better method of Emploration.