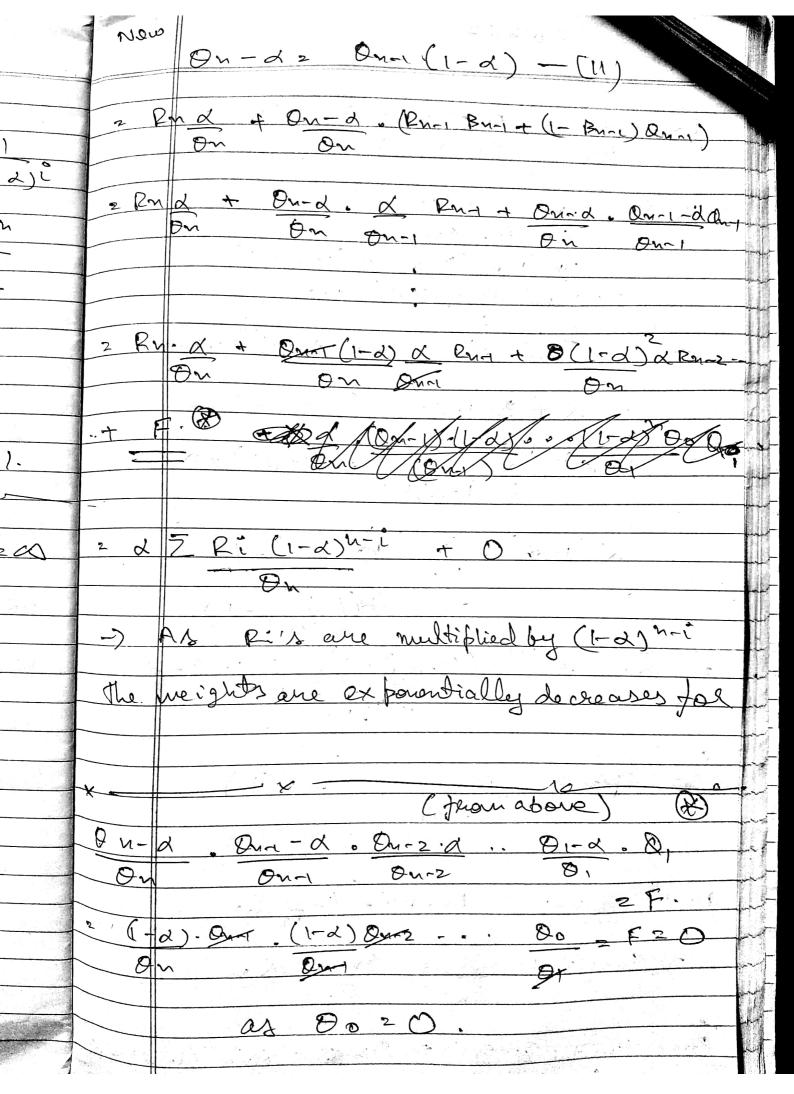
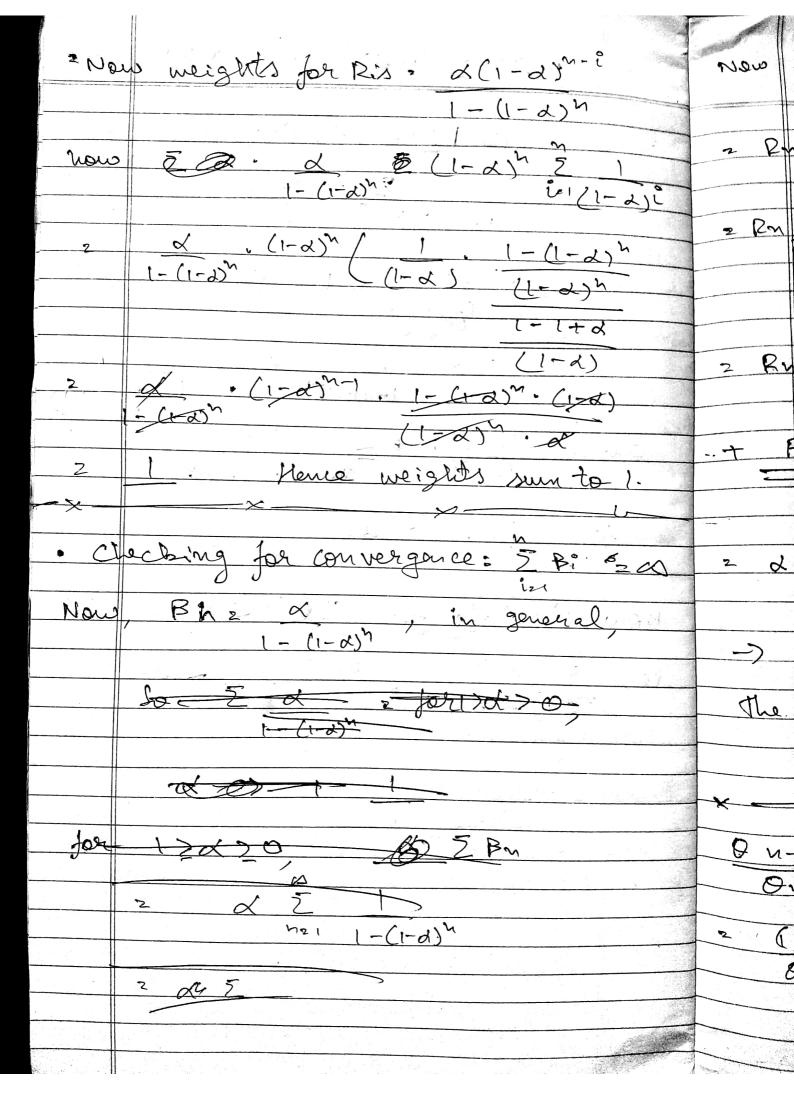
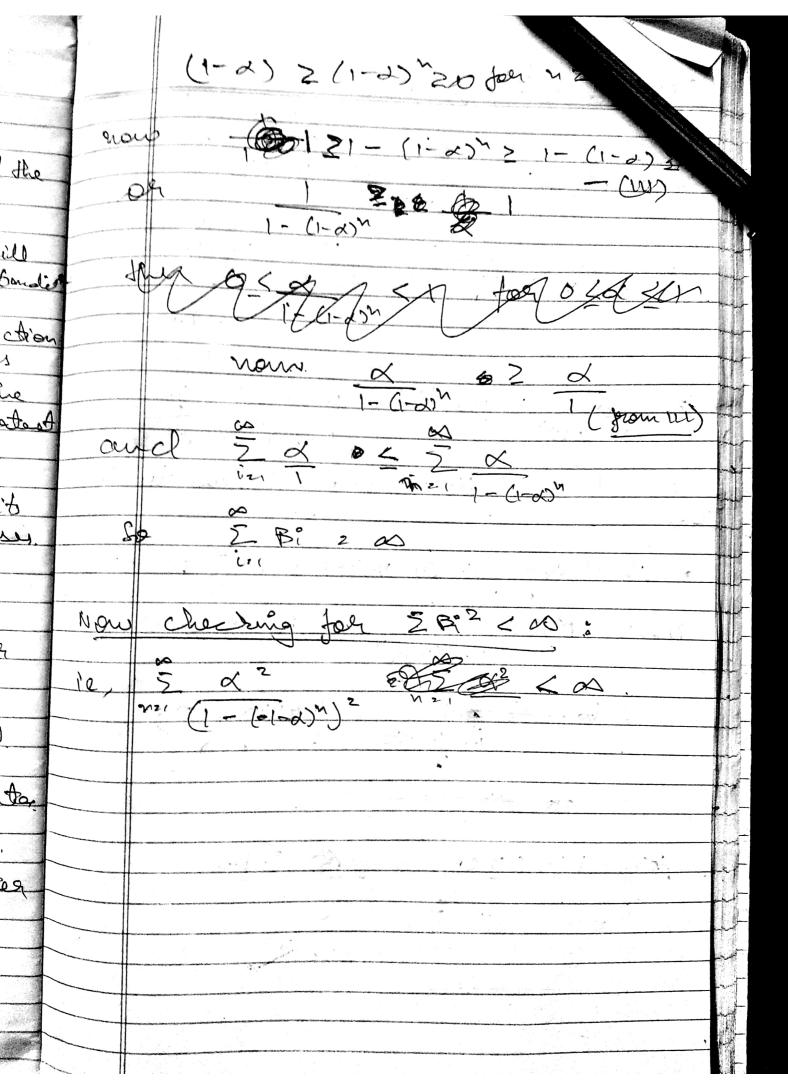
	Diogram Anshumaan RL-M·W·I	
23.	2017147	Qni
Bn	= d, on 2 on + d (1 - on-1)	Qna
	Ph 8020, h710.	
now,	On > On-1 + of CI-Drace	
	2 d+ (1-d)0mm1	
_	2 d+ (1-d) (0n-2+ d(1-0n-2))	2 6
	2 d+ d(1-d) + (1-d)2 0 n-z	
-	N N	
	2 Zd(1-d)n-i + (1-d) 90+=1-d	- 1 - 1 - 1
- American Company	= d (= 1-d)	2 K
2	(1-2) (5 1) + (1-2) N V	
Tipleage and an artist and a second	((1-d) (5 1) + (1-d) x 0	<u> </u>
destipations		
2	d(1-d), 1 0 (1-d), -1	
	$\left(\mathcal{O}\left(\frac{1}{1-\alpha}\right)-1\right)$	
•	2.11-xxxxx /11-xxxxx	
	(() () () () () () () () () (
	A(1-d)	
2		
So,	Baz (1)	
	1-(-a)	
Non	P, Onti = Rm. Pu+ (1-8) Qn	
2)	Bu+1 2 Rm - d + On-d . Qn	
	On On	



Scanned by CamScanner





Scanned by CamScanner

<u> </u>	
Chationary Metionistic man la	
Stationary Optimistic gerody:	
1. Spikes are visible for the first pass of the generally action search. This is:	9000
geneedy action search. This is	
· Due to an oftimistic initial value, which will decrease after it is selected for a given bondis	
decrease after it is selected for a given for it	:
2. A large spike (40%) optimistic optimal action	
is selected immediatly after the first pass since the actual oftimal boundit Cuiththe highest expected reward) will have the greatest	0. (1)
since the actual oftheal boundit Cwiththe	
highest expected reward) will have the greatest	
estimate after the first pass.	C.
> to motally lags while exploring all bandits	- 10g A
3. It initially lags while exploring all bandits but eventually settles as exploration decreases.	_ '
Non- Stationary Offinistic geneda:	
Man - Startanay aprilmsell gracing:	£1 0
1. Ohtimistic geready initially bantonne better	100
1. Obtimistic generally initially penforms better Because & O* (a) for bandites has not	<i>)</i> 0
changed much. But after time passes	14
the instial distribution has changed	
and if no non- longer has the coarso	
20 realistic gleedy heeps exploring estimates	
2. Realistic guedy keeps explaining other	
2. Realistic generally keeps exploring other. est actions and annually gets a better estimate.	
estimato.	

24. Coptimal Action) 1. In stationary case, UCB explains all actions first. As hit is bounded the increments will eventually become ne gligible and will have explored all actions with lower estimates and more frequent examples expland lessers this is fetter than a healistic generaly Zerapperoach that explores to leafis explaining suboptimal actions even after a large no of steps. Optimistic guedy explores actions and then immediately stokes due to blas and then immediatly stops exploing due to sample averaging. · The spile in UCB appears after the fiest pass over all bandits, as the actual and will get selected more after.