

2.5. Bynanie Programming
(1) Stochastic:
(1) Stochastic: using egn 4.4, rul can say:
$ \frac{1}{\sqrt{(s)}} = \underbrace{E_{\chi} \left[G_{1} + \left[S_{1} + S_{2}\right]}_{= \chi(a s)} \underbrace{E_{\chi} \left[G_{1} + \left[S_{2} + S_{3}\right]}_{= \chi(a s)} \underbrace{E_{\chi} \left[G_$
a 5, y y (3)
Using can 4.6, we can say
9,7 (3,a) = E[R+++ YVZ(S++) St=8,
$= \sum_{S,N} p(S,y S,q) \left[N + y \sqrt{\chi(S)} \right]$
Using the about two egrs, we get:
VX(3) = \(\frac{2}{a} \) \(\frac{1}{a} \) \(\



(2) Deterministic :-

 $v^{\chi}(s) = Q_{\chi}(s,a) + a \in A$: $\chi(a|s) = 1$

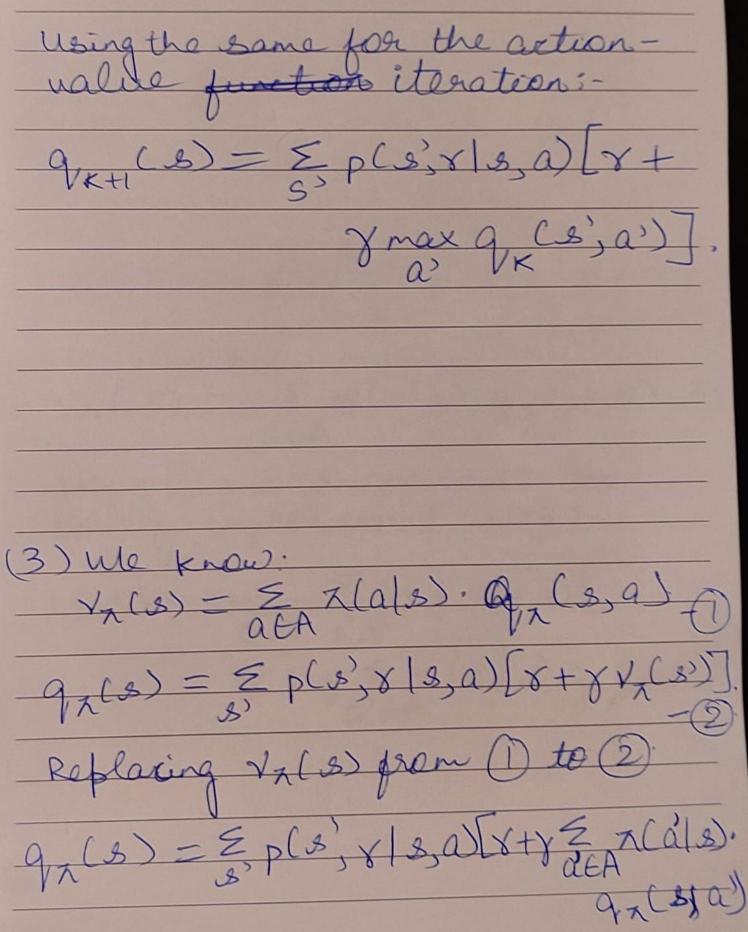
(2) The Bollman oftimality egn for the action-value gentlion:

9,x(s,a)= \(\frac{1}{2}\)p(s,x|s,a)[x+

γ max q (s', a')].

The value - iteration for the state value quartion was obtained by turning its Bellman egn to an update (xule-EBarto Section)







the war of 49 was an ear
F
(4) Guiner:
T(S) = argmax & p(s', x s,a)[x+
whe also know that: Vals) = Ex(als). qx(s,a)-0 aca
Using (1) we can rewrite the policy improvement step:-
7(3) < aggrax E' p(3', x 3,a) 7 + y \(\frac{2}{2} \) \(\frac{2}{3} \) \(\frac
a'EA VX



(5) The policy evaluation stop on
page 75 assumes that 7(3) has
a probability distribution over
a probability distribution over la EA. That's why a summation
ouer & + a E A is taken on page
75.
The evaluation step on page 80
assures a deterministic or a
Liste a sale soli in with goodent
to the actions. Honce, there is
to the actions. Home, increase
only one action that should be
chober for a state.