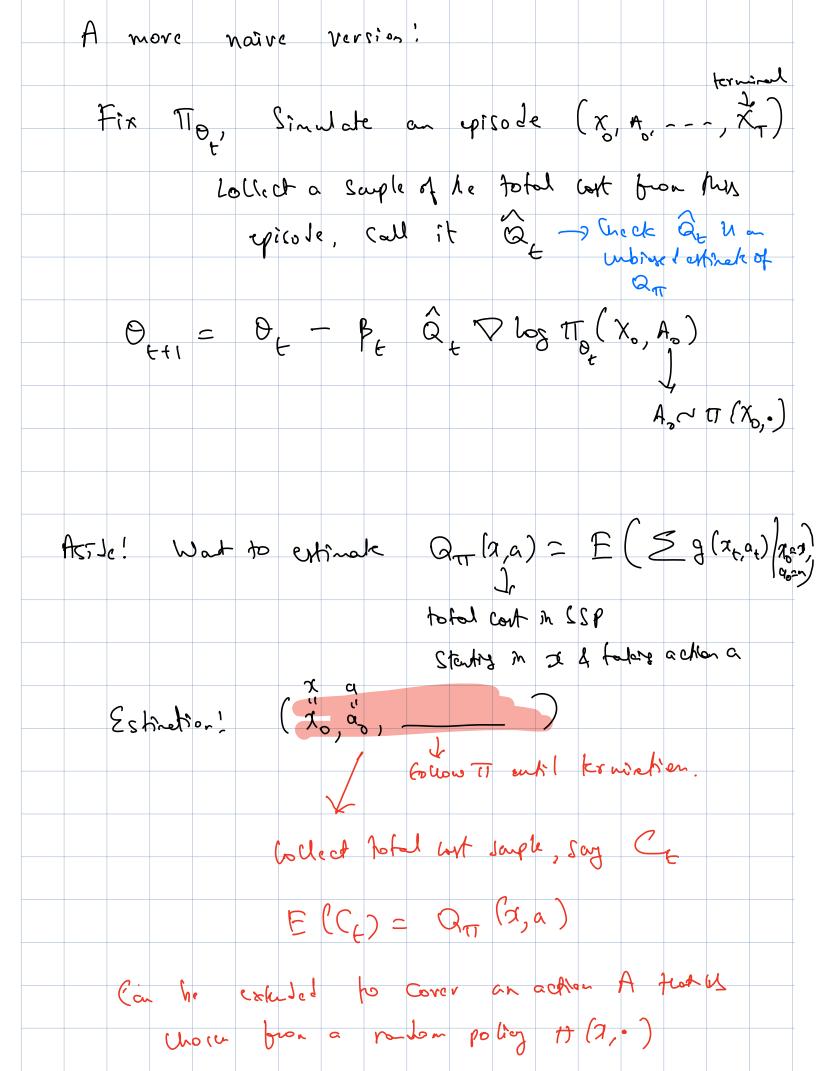


Q<sub>T</sub> > extincte of Q<sub>T</sub>  $\nabla_{\sigma} \mathcal{I}_{\Pi_{\sigma}}(p) \simeq \mathcal{E}_{\Pi_{\sigma}}\left(\underset{\alpha}{\leq} \nabla \Pi(X, \alpha) \otimes_{\Pi}(X, \alpha)\right)$  $= E_{T_{0}} \left( \begin{array}{c} \leq TT(X,a) Q_{TT}(X,a) & TT(X,a) \\ a & TT(X,a) & TT(X,a) \end{array} \right)$  $= E_{\pi} \left( \sum_{a} \operatorname{tr} (X_{,a}) Q_{\pi} (X_{,a}) \nabla \log \operatorname{tr} (X_{,a}) \right)$ Suppose A in - r.v. chosen with the distribution TT  $-) = E_{\pi} \left( Q_{\pi} (X, A) \nabla \log t (X, A) \right)$ both K.A are random = E(h(x))X & chosen from a Distributionary pe occuracy nonne & A & chosen way to (X, .) Suppose E(Q) = Qr. Tron,  $\Theta_{t+i} = \Theta_t - B_t \hat{Q}_t \nabla \log \pi_{o_t}(x_t, A_t)$ REINFORCE algorithm



In *F*finiFor(E, policy evaluation (= Q<sup>To</sup>(·,·)) femant! "Monk Corlo". C Instead, if we use a peranctric approximation for  $\Theta_{TT}$ , Say  $\Theta_{TT}(x,a) \simeq \gamma^T \phi(x,a)$ linear Func. approx. Con me TD with 2FA to obtain an estimate of Q<sub>(T<sub>B</sub></sub> (.,.) Suppose TD conveyes to 7th.  $\Theta_{\pi}(x,a) \stackrel{\sim}{\sim} \stackrel{\sim}{\sim} \stackrel{\sim}{\sim} \phi(x,a)$  $E(\gamma^{*T}\phi(\chi_{a})) \not\subset Q_{T}(\chi_{a}) \overset{\&?}{\sim}$ "Paranchie approximpion induce a bras" On the other hand, MC methods Windly have large varance. Can de policy gradiat with function approximition Actor - Critic algorithms.

