Model-based RL \implies know the knowledge of p(s;r|s,a)Model-free RL \implies p(s;r|s,a) is unknown

(1) Model-based RL => work with { Bellman Equa. exactly.

Bellman Optimality Equa.

Recep: a) Bellman Optimality Equation

Non-linearity
$$Q_*(s, \alpha) = \max_{\alpha} Q_*(s, \alpha)$$

$$Q_*(s, \alpha) = \sum_{s', r} p(s', r \mid s, \alpha) (r + \gamma) V_{\pi^*}(s')$$

iterative method

 $V_{\pi}(s) = \sum_{\alpha} \pi(\alpha(s) q_{\pi}(s, \alpha))$

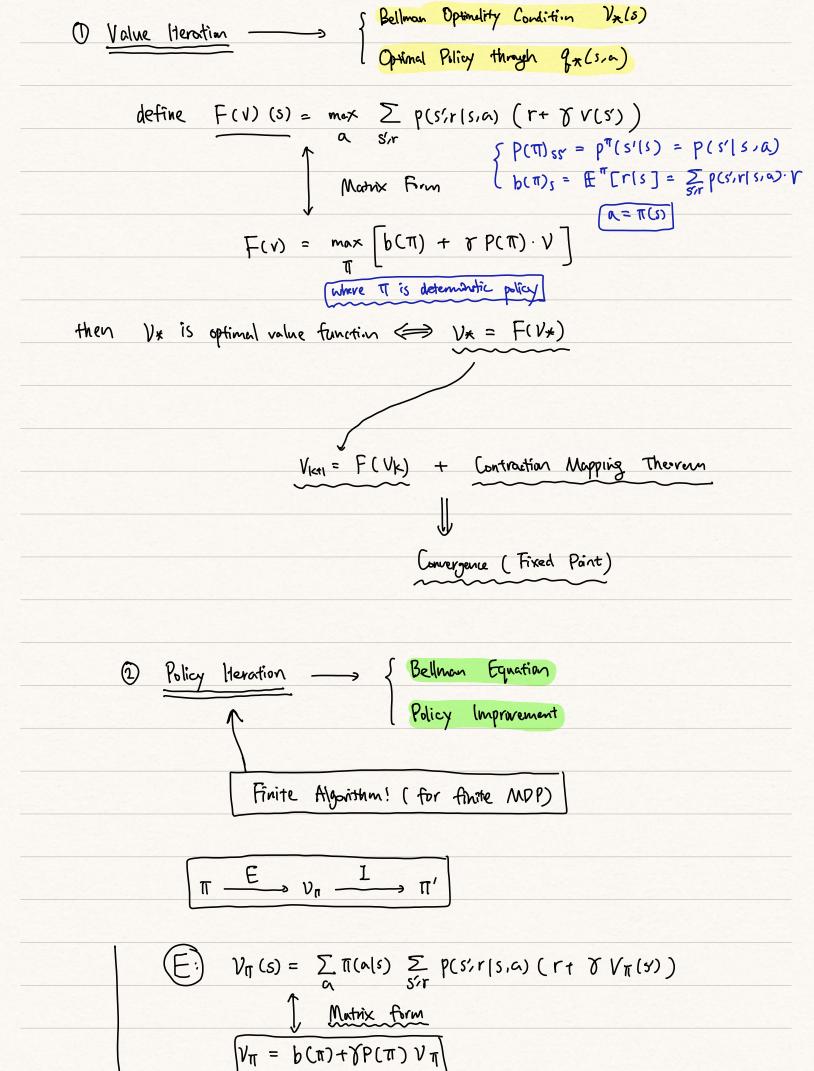
$$g_{\pi}(s,a) = \sum_{s'r} p(s',r|s,a) (r + \gamma v_{\pi}(s'))$$

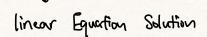
Policy I tempion

c) Openal Policy

$$\pi^*(s) \in argmax q_*(s,a)$$
 $\alpha \in A$

d) Policy Improvement





I:
$$T(s) = argmax \theta_T(s,a)$$

 $\pi' = \underset{\widetilde{\pi} \in \text{ determinitie}}{\operatorname{argmax}} \left[b(\widetilde{\pi}) + \gamma P(\widetilde{\pi}) \cdot V_{\pi} \right]$

(policy improvement: $T' \geqslant \pi \iff V_{\pi'}(s) \geqslant V_{\pi}(s) \quad \forall s \in S$)

-> estimate Expertation from Sample Mean

guarantee by { LLN CLT

 $\mathbb{E}_{X\sim M} f(x) = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \quad X_i \sim M \quad (i.i.d)$

- Then P(s,r), a) is our known knowledge, then $V_{\Pi} = b(\Pi) + \mathcal{T} P(\Pi) V_{\Pi}$
- -> When PCs/r/s,a) is unknown,

then we go back to the defn of Va (5).

Mounte - Carlo Comes in:

Sample N trajectories (episodes) through Policy
$$TI$$
 $S_{t}^{(n)}$, $R_{t}^{(n)}$,

Generalized Policy Iteration Framework

Not solve for Un exactly

Short-term Monte Carlo Fashian

