solution of finite MP / SDE

goal: determine dynamics of finite MP/SDE

refs:

Neuro-Dynamic Programming

Dimitri P. Bertsekas and John N. Tsitsiklis

Chapter 2

Reinforcement Learning

An Introduction

Richard S. Sutton and Andrew G. Barto

Chapter 3

- consider a finite MP/SDE $(X, \mathcal{U}, P)$, i.e. $|X|, |\mathcal{U}| < \infty$

$$x^+ \sim P(x, u), \quad P : X \times \mathcal{U} \to \Delta(X)$$

→ starting from state $p \in \Delta(X)$ and applying policy $\pi : X \to \Delta(\mathcal{U})$
determine next state $p^+ \in \Delta(X)$
(it can help to regard $p : X \to [0,1]$ — determine $p^+(x^+)$, all $x^+ \in X$)

– we can compute $p^+(x^+)$ for each $x^+ \in X$ using rules of probability:

$$p^+(x^+) = \sum_{x \in X} \left(p(x) \left[\sum_{u \in \mathcal{U}} \pi(u \mid x) P(x^+ \mid x, u) \right] \right)$$

conditional probability $= [\pi(x)](u)$

* this determines a <u>deterministic</u> (!) difference equation:

* this determines a <u>deterministic</u> (!) difference equation:

$$p^+ = F(p), \quad (\underbrace{p : X}\underbrace{\to [0,1]}) \in \Delta(X) \subset [0,1]^X \subset \mathbb{R}^X = \mathbb{R}^N$$

$p \in [0,1]^X$ — more generally we let

$\Rightarrow p \in \mathbb{R}^X = \mathbb{R}^N$ $\quad B^A = \{ f : A \to B \}$

e.g. $\mathbb{R}^n = \{ v : \{1, \dots, n\} \to \mathbb{R} \}$

→ show that F is <u>linear</u> (!)

(i.e. find $\Gamma \in \mathbb{R}^{N \times N}$, $N = |X|$, s.t. $F(p) = \Gamma p = p^+$)

– $[\Gamma]_{x^+, x} = \sum_{u \in \mathcal{U}} \pi(u|x) P(x^+ | x, u)$ yields $p^+ = \Gamma p$

* we can use linear systems theory to characterize "solutions" / trajectories including their asymptotic properties !

→ we're studying discrete-time linear time-invariant DE $p^+ = \Gamma p$

◦ first of all, we know trajectories $p_t = \Gamma^t p_0$, any $t \in \mathbb{N}$

↑ t-fold matrix multiplication

◦ furthermore, Γ has special properties: $1^T \Gamma = 1^T \Rightarrow 1 \in \operatorname{spec} \Gamma$

↑ $1^T = (1, \dots, 1) \in \mathbb{R}^N$

i.e. Γ is "left-stochastic" $\Rightarrow \forall \lambda \in \operatorname{spec}\Gamma : |\lambda| \leq 1$

* if $\underbrace{[\Gamma]_{x^+, x} > 0}$ then $\lim_{t \to \infty} p_t = \bar{p}$ where $\Gamma \bar{p} = \bar{p}$ is unique

↑ more generally, if Γ is <u>irreducible</u> and <u>aperiodic</u>)

so all trajectories asymptotically converge to unique $\bar{p}$!