goal: derive à analyze Bellman operators for MDP/SOCP

Neuro-Dynamic Programming

Dimitri P. Bertsekas and John N. Tsitsiklis

chapter 2

o given policies $\pi: X \to \Delta(\mathcal{U})$, know that value $\nu^{\pi}: X \to \mathbb{R}$ defined by $\nu^{\pi}(x) = \mathbb{E}[c(x,u) \mid x, \nu x, u_{t} \nu \pi(x_{t})]$ satisfies Bellman equation

 $\forall x \in X : V^{T}(x) = \sum_{u \in \mathcal{U}} T(u|x) \sum_{x \neq e X} P(x^{+}|u,x) \cdot \left(\mathcal{X}(x_{t},u_{t}) + \nabla \cdot V^{T}(x^{+}) \right)$

idea: use this equation to define operator $T_{\pi}: \mathbb{R}^{X} \to \mathbb{R}^{X}$, $\mathbb{R}^{X} = \{v: X \to \mathbb{R}\}$ $: v \to T_{\pi}v = v^{+}$ "value-like" functions,

learning Page

```
· N -> TIN= V' "value - like"
functions,
  by \forall x \in X : (T_{\pi} v)(x) = \sum_{u \in \mathcal{U}} \pi(u|x) \sum_{v \neq v} P(x^{+}|u_{i}x) \cdot (\mathcal{X}(x_{t_{i}}u_{t}) + \nabla \cdot v(x^{+}))
o given optimal policies \pi^*: X \to \Delta(u), the optimal value v^*: X \to \mathbb{R}
  satisfies \forall x \in X: V^*(x) = \sum_{u \in \mathcal{U}} \tau^*(u|x) \sum_{v \neq u, x} P(x^+|u,x) \cdot (\mathcal{X}(x_t, u_t) + \nabla \cdot N^*(x^+))
      optimal Bellman = \min \sum_{u \in \mathcal{U}} P(x^{+}|u,x) \cdot (Z(x_{t},u_{t}) + Y \cdot v^{*}(x^{+})) equation
   so we can define T: \mathbb{R}^X \to \mathbb{R}^X : v \longmapsto Tv = v^t
   by \forall x \in X: (T \cup X)(x) = \min \sum_{u \in U} P(x + |u, x) \cdot (Z(x_{t}, u_{t}) + Y \cdot v(x^{t}))
-> what kind of operator is TT? T? (give a simple expression for TT)
 - To is affine ?
 - letting [A_{T}]_{X_{1}X^{+}} = \sum_{u \in \mathcal{U}} \tau(u|x) \cdot \sum_{x \neq cx} P(x^{+}|x_{1}u),
                    [b_{\pi}]_{x} = \sum_{u \in \mathcal{U}} \pi(u|x) \cdot \sum_{v \neq u \times} P(x^{+}|x_{i}u) \cdot \mathcal{L}(x_{i}u)
       we have T_{\Pi}v = V \cdot A_{\Pi} \cdot V + b_{\Pi} \longrightarrow \text{Bellman equation is } T_{\Pi}V^{\Pi} = V^{\Pi}
 -T is nonlinear (and non-affine)
old T_{\pi}^{k} = T_{\pi} \circ T_{\pi} \circ \cdots \circ T_{\pi}, T_{\pi}^{k} = T_{\pi} \circ \cdots \circ T_{\pi}
                           & timec
                                                                 & times
   and endow RX with the max norm \|v\|_{\infty} = \max_{x \in X} |v(x)|:
thm: (Bellman operators are confractions - 2.5 in BT96)
```

 $\forall v, w \in \mathbb{R}^{\times}, \ \pi: X \rightarrow \Delta(u): \ \| Tv - Tw \|_{\infty} \leq \gamma \cdot \| v - w \|_{\infty}$

learning Page

 $\forall v_1w \in \mathbb{R}^X$, $\forall x_1 : X \to \Delta(u)$: $\| \forall v_1 - \forall v_1 \|_{\infty} \leq x_1 \cdot \| v_1 - \forall v_1 \|_{\infty}$ $\| \forall v_1 - \forall v_1 \|_{\infty} \leq x_1 \cdot \| v_1 - \forall v_1 \|_{\infty}$ $\| \forall v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_1 - \forall v_2 \|_{\infty}$ $\| \forall v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty}$ $\| \forall v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty}$ $\| \forall v_1 - \forall v_2 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_1 \cdot \| v_1 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_1 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_2 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_2 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_2 \cdot \| v_2 \|_{\infty} \leq x_2 \cdot \| v_1 \|_{\infty} \leq x_2 \cdot \| v_2 \|_{\infty} \leq x$

3°. T is optimal \Leftrightarrow T_{tt} $v^* = Tv^* = v^*$

-> use these facts to derive algorithms to (approximately) solve MDP/SOCP