CLASS WILL START AT 5PM

Spring 2020

Lecture 5
Sections 4:
Dynamic Programming for Discrete-Time Stochastic Systems
Problem Definition

- Let $\{w_t\} = \{w_t\}_{t=0}^{T-1}$ be an IID sequence of r.v.s.
- Let $\xi_t$ denote the system state at time, $t$, and let $u_t$ denote the control input at time, $t$.
- Dynamics:
  \begin{align*}
  \xi_{t+1} &= f(\xi_t, u_t, w_t) \quad \forall \ t \in ]0, T - 1[, \\
  \xi_s &= x.
  \end{align*}
\hspace{1cm} (D) \hspace{1cm} (IC)

- $s \in ]0, T[$ is the initial time, and $x \in \mathcal{X} \doteq \mathbb{R}^n$ is the initial state.
- Suppose $u_t \in U \subseteq \mathbb{R}^m$ (possible control bounds)
- Suppose the $w_t$ take values in $\mathbb{R}^k$.
- For $s \in ]0, T[$, the payoff or cost criterion will be $J'(s, \cdot, \cdot) : \mathcal{X} \times U^{T-s} \to \mathbb{R}$, given by
  \begin{align*}
  J'(s, x, u.) &= \mathbb{E}\{\xi_s = x \left\{ \sum_{t=s}^{T-1} L(\xi_t, u_t) + \Psi(\xi_T) \right\}
  \mathbb{E}\{\sum_{t=s}^{T-1} L(\xi_t, u_t) + \Psi(\xi_T) | \xi_s = x \}. 
  \end{align*}
\hspace{1cm} (P')

- Notation here: $U^{T-s} \doteq U \times U \times ... U$.
- Sometimes may write the following to note what we’re taking expectation over (notation abuse, obviously).
  \begin{align*}
  J'(s, x, u.) &= \mathbb{E} \{w_t\}_{t=s}^{T-1} \left\{ \sum_{t=s}^{T-1} L(\xi_t, u_t) + \Psi(\xi_T) \right\}.
  \end{align*}
Problem Definition

- Need to make sure $u$ doesn't know the future.
- For $s \leq T - 1$, let
  \[
  \tilde{\mathcal{M}}_{s,T-1} = \{ \mu : \mathcal{X}^{T-s} \times ]s, T-1[ \rightarrow U^{T-s} \mid \text{for all } t \in ]s, T-1[, \text{ if } \xi_r, \hat{\xi}_r \in \mathcal{X}^{T-s} \text{ is s.t. } \xi_r = \hat{\xi}_r \forall r \leq t, \text{ then } \mu(\hat{\xi}_r, t) = \mu(\xi_r, t) \}. \]
- Elements are denoted by $\mu_t(\xi_r)$.
- This is another condition that guarantees the controls are non-anticipative.
- $s \in ]0, T[$ is the initial time, and $x \in \mathbb{R}^n$ is the initial state.
- Adjust the payoff/cost criterion definition to $J(s, \cdot, \cdot) : \mathcal{X} \times \tilde{\mathcal{M}}_{s,T-1} \rightarrow \mathbb{R}$, given by
  \[
  J(s, x, \mu) = \mathbb{E} \bigg|_{\xi_s = x} \left\{ \sum_{t=s}^{T-1} L(\xi_t, \mu_t(\xi_r)) + \Psi(\xi_T) \right\}, \quad (P)
  \]
  where $\xi$ satisfies $(D), (IC)$ (Alt: $\cdots \mathbb{E} \big|_{\{w_k\}_{k=s}^{T-1}} \cdots$)
- The value function is $V : ]0, T[ \times \mathcal{X} \rightarrow \mathbb{R}$ given by
  \[
  V(s, x) = \inf_{\mu \in \tilde{\mathcal{M}}_{s,T-1}} J(s, x, \mu) \quad \forall s \in ]0, T[, \ x \in \mathcal{X}.
  \]
Aside on Infimum and Supremum

Motivational example for control over time interval $t \in [0, 1]$:

\[
\dot{\xi}_t = ut; \quad \xi_0 = x = (0, 0)^T; \\
U = \{ v \in \mathbb{R}^2 \mid |v| = 1 \}; \quad \mathcal{U} \equiv L_2((0, 1); U);
\]

Minimize the payoff:

\[
F(0, x, u) = \int_0^1 |\xi_t|^2 \, dt.
\]

Try $u_t = (\cos(\omega t), \sin(\omega t))$...
Aside on Infimum and Supremum

- Definitions and examples
Dynamic Programming Concept

DPP (Finite Time-Horizon (FTH))

- A “principle” needs to be proven for specific classes of problems.

**Theorem**

Let \(0 \leq s \leq t \leq T, \; x \in \mathcal{X}\). We have

\[
V(s, x) = \inf_{\mu \in \tilde{\mathcal{M}}_{s, t-1}} E_{\xi_s = x} \left\{ \sum_{r=s}^{t-1} L(\xi_r, \mu_r(\xi_r)) + V(t, \xi_t) \right\}, \quad (DPP)
\]

\[
V(T, x) = \psi(x). \quad (TC)
\]

**Proof sketch:**