Dynamic programming equations

Michel DE LARA
CERMICS, École des Ponts ParisTech
Université Paris-Est
France

February 1, 2011

Outline of the presentation

Outline of the presentation

Discrete-time nonlinear control system

$$\begin{cases} x(t+1) = \texttt{Dyn}\big(t, x(t), u(t)\big), \quad t = t_0, t_0+1, \ldots, T-1 \\ x(t_0) \quad \text{given,} \end{cases}$$

Trajectories

- Control trajectory

$$u(\cdot) = \underbrace{\big(u(t_0), u(t_0+1), \ldots, u(T-1)\big)}_{\text{decision path}}$$

- State trajectory

$$x(\cdot) = \big(x(t_0), x(t_0+1), \ldots, x(T-1), x(T)\big)$$

- Control-state trajectory

$$\big(x(\cdot), u(\cdot)\big) = \big(x(t_0), \ldots, x(T), u(t_0), \ldots, u(T-1)\big)$$

Criterion

- Set of trajectories

$$\big(x(\cdot), u(\cdot)\big) \in \underbrace{\mathbb{X}^{T+1-t_0} \times \mathbb{U}^{T-t_0}}_{\text{set of trajectories}}$$

- A criterion $\texttt{Crit}$ is a function

$$\begin{aligned} \texttt{Crit} : \quad \mathbb{X}^{T+1-t_0} \times \mathbb{U}^{T-t_0} & \to \mathbb{R} \\ \big(x(\cdot), u(\cdot)\big) & \mapsto \texttt{Crit}\big(x(\cdot), u(\cdot)\big) \end{aligned}$$

  - which assigns a scalar value $\texttt{Crit}\big(x(\cdot), u(\cdot)\big)$
  - to a state and control trajectory $\big(x(\cdot), u(\cdot)\big)$.

General additive criterion

$$\texttt{Crit}\big(x(\cdot), u(\cdot)\big) = \sum_{t=t_0}^{T-1} \underbrace{\texttt{Util}\big(t, x(t), u(t)\big)}_{\text{instantaneous gain}} + \underbrace{\texttt{UtilFin}\big(T, x(T)\big)}_{\text{final gain}} .$$

The optimization problem

$$\mathtt{Crit}^\star(t_0, x_0) = \max_{\big(x(\cdot),u(\cdot)\big)\in\mathcal{T}_{\mathrm{ad}}(t_0,x_0)} \sum_{t=t_0}^{T-1} \underbrace{\mathtt{Util}\big(t, x(t), u(t)\big)}_{\text{instantaneous gain}} + \underbrace{\mathtt{UtilFin}\big(T, x(T)\big)}_{\text{final gain}} .$$

Additive value function

For $t = t_0, \dots, T-1$,

$$V(t,x) := \max_{\left(x(\cdot),u(\cdot)\right) \in \mathcal{T}_{\mathrm{ad}}(t,x)} \sum_{s=t}^{T-1} \mathtt{Util}\big(s, x(s), u(s)\big) + \mathtt{UtilFin}\big(T, x(T)\big)$$

is the optimal performance starting from state x at time t.

Dynamic programming equation

Proposition

In the case without state constraints, the value function is the solution of the following backward dynamic programming equation (or Bellman equation)

$$\begin{cases} V(T,x) &= \texttt{UtilFin}(T,x)\ , \\ V(t,x) &= \max_{u \in \mathbb{B}(t,x)} \Big(\texttt{Util}(t,x,u) + V\big(t+1, \texttt{Dyn}(t,x,u)\big) \Big)\ , \end{cases}$$

where $t = T-1,\ T-2,\ \ldots,\ t_0+1,\ t_0$.

The Maximin: Rawls criterion

$$\texttt{Crit}\big(x(\cdot), u(\cdot)\big) = \min_{t=t_0,\dots,T-1} \texttt{Util}\big(t, x(t), u(t)\big)$$

- John Rawls, *A Theory of Justice*, 1971
- The utility level of the least advantaged generation

$$\texttt{Crit}^\star(t_0, x_0) = \max_{\big(x(\cdot),u(\cdot)\big)\in\mathcal{T}_{\mathrm{ad}}(t_0,x_0)} \min_{t=t_0,\dots,T-1} \texttt{Util}\big(t, x(t), u(t)\big)$$

Maximin dynamic programming equation

Proposition

$$V(t,x) := \max_{\big(x(\cdot),u(\cdot)\big)\in\mathcal{T}_{\mathrm{ad}}(t,x)} \Big(\min_{s=t,\dots,T-1} \mathtt{Util}\big(s, x(s), u(s)\big) \Big)$$

is the solution of

$$\begin{cases} V(T,x) &= +\infty \, , \\ V(t,x) &= \max_{u\in\mathbb{B}(t,x)} \min \Big(\mathtt{Util}(t,x,u), V\big(t+1, \mathtt{Dyn}(t,x,u)\big) \Big) . \end{cases}$$

Outline of the presentation

Discrete-time control dynamical system with uncertainty

$$\begin{cases} x(t+1) = \texttt{Dyn}\big(t, x(t), u(t), w(t)\big), & t = t_0, \ldots, T-1 \\ x(t_0) = x_0 \end{cases}$$

Scenarios

We assume that

$$w(t) \in \mathbb{S}(t) \subset \mathbb{W} \; ,$$

so that the sequences

$$w(\cdot) := \big(w(t_0), w(t_0+1), \ldots, w(T-1), w(T)\big)$$

belonging to

$$\Omega := \mathbb{S}(t_0) \times \cdots \times \mathbb{S}(T) \subset \mathbb{W}^{T+1-t_0}$$

capture the idea of possible *scenarios* for the problem.

Criterion

A criterion $\texttt{Crit}$ is a function

$$\texttt{Crit} : \mathbb{X}^{T+1-t_0} \times \mathbb{U}^{T-t_0} \times \mathbb{W}^{T+1-t_0} \to \mathbb{R}$$

which assigns a real number $\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big)$
to a state, control and uncertainty trajectory $\big(x(\cdot), u(\cdot), w(\cdot)\big)$.

General additive criterion

$$\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \sum_{t=t_0}^{T-1} \underbrace{\texttt{Util}\big(t, x(t), u(t), w(t)\big)}_{\text{instantaneous gain}} + \underbrace{\texttt{UtilFin}\big(T, x(T), w(T)\big)}_{\text{final gain}} .$$

General multiplicative criterion

$$\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \prod_{t=t_0}^{T-1} \texttt{Util}\big(t, x(t), u(t), w(t)\big) \times \texttt{UtilFin}\big(T, x(T), w(T)\big)$$

$$\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \prod_{t=t_0}^{T} \mathbf{1}_{\mathbb{A}(t)}\big(x(t)\big) .$$

The Maximin

$$\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \min_{t=t_0,\ldots,T-1} \texttt{Util}\big(t, x(t), u(t), w(t)\big) ,$$

Fear operator

Consider a general set Ω.
The so-called fear operator $\mathbb{F}_\Omega$ on Ω is defined on the set of functions $A : \Omega \to \overline{\mathbb{R}}$ by:

$$\mathbb{F}_\Omega[A] = \mathbb{F}_w\left[A(w)\right] := \min_{w \in \Omega} A(w) \; .$$

When $\Omega = \Omega_1 \times \Omega_2$, we have the formula:

$$\mathbb{F}_\Omega[A] = \mathbb{F}_{(w_1,w_2)}\left[A(w_1,w_2)\right] = \mathbb{F}_{w_1}\left[\mathbb{F}_{w_2}\left[A(w_1,w_2)\right]\right] \; .$$

Robust additive dynamic programming

$$\min_{w(\cdot)\in\Omega} \sum_{t=t_0}^{T-1} \underbrace{\texttt{Util}\big(t, x(t), u(t), w(t)\big)}_{\text{instantaneous gain}} + \underbrace{\texttt{UtilFin}\big(T, x(T), w(T)\big)}_{\text{final gain}} .$$

$$\begin{cases} V(T,x) & := \min_{w\in\mathbb{S}(T)} \texttt{UtilFin}(T,x,w) , \\ V(t,x) & := \max_{u\in\mathbb{B}(t,x)} \min_{w\in\mathbb{S}(t)} \Big[\texttt{Util}(t,x,u,w) \\ & \qquad + V\big(t+1, \texttt{Dyn}(t,x,u,w)\big)\Big] . \end{cases}$$

Robust maximin dynamic programming

$$\min_{w(\cdot)\in\Omega} \min_{t=t_0,\dots,T-1} \texttt{Util}\Big(t, x(t), u(t), w(t)\Big) \ ,$$

$$\begin{cases} V(T,x) & := \displaystyle\min_{w\in\mathbb{S}(T)} \texttt{UtilFin}(T,x,w) \ , \\ V(t,x) & := \displaystyle\max_{u\in\mathbb{B}(t,x)} \min_{w\in\mathbb{S}(t)} \min\Big(\texttt{Util}(t,x,u,w), V\big(t+1, \texttt{Dyn}(t,x,u,w)\big)\Big) \ . \end{cases}$$

Expectation operator

Consider a probability space Ω with σ-field $\mathcal{F}$ and probability $\mathbb{P}$. The so-called expectation operator $\mathbb{E}_{(\Omega,\mathcal{F},\mathbb{P})}$ is defined on the set of measurable and integrable functions $A : \Omega \to \overline{\mathbb{R}}$ by:

$$\mathbb{E}_{(\Omega,\mathcal{F},\mathbb{P})}[A] = \mathbb{E}_w \left[A(w)\right] = \mathbb{E}_{\mathbb{P}} \left[A(w)\right] = \int_\Omega A(w) \, d\mathbb{P}(w) \, .$$

When $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ and $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$, we have the Fubini formula:

$$\mathbb{E}_{(\Omega,\mathcal{F},\mathbb{P})}[A] = \mathbb{E}_{(\Omega_1,\mathcal{F}_1,\mathbb{P}_1)} \left[\mathbb{E}_{(\Omega_2,\mathcal{F}_2,\mathbb{P}_2)} \left[A(w_1, w_2)\right]\right] \, .$$

The primitive random process $w(\cdot)$ is assumed to be a sequence of independent random variables $\big(w(t_0), w(t_0+1), \dots, w(T-1), w(T)\big)$ under probability $\mathbb{P}$ on the domain Ω of scenarios.
The probability $\mathbb{P}$ is the product of its marginals.

Stochastic additive dynamic programming

$$\mathbb{E}_{w(\cdot)}\Big(\sum_{t=t_0}^{T-1} \underbrace{\mathtt{Util}\big(t, x(t), u(t), w(t)\big)}_{\text{instantaneous gain}} + \underbrace{\mathtt{UtilFin}\big(T, x(T), w(T)\big)}_{\text{final gain}} \Big) .$$

$$\begin{cases} V(T, x) & := \mathbb{E}_{w(T)}\Big[\mathtt{UtilFin}\big(T, x, w(T)\big)\Big] , \\ V(t, x) & := \max_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)}\Big[\mathtt{Util}\big(t, x, u, w(t)\big) \\ & \quad + V\big(t+1, \mathtt{Dyn}(t, x, u, w(t))\big)\Big] . \end{cases}$$

Stochastic multiplicative dynamic programming

$$\mathbb{E}_{w(\cdot)}\Big(\prod_{t=t_0}^{T-1} \mathtt{Util}\big(t, x(t), u(t), w(t)\big) \times \mathtt{UtilFin}\big(T, x(T), w(T)\big)\Big)$$

$$\begin{cases} V(T,x) := \mathbb{E}_{w(T)}\Big[\mathtt{UtilFin}\big(T, x, w(T)\big)\Big] , \\ V(t,x) := \max_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)}\Big[\mathtt{Util}\big(t, x, u, w(t)\big) \times V\big(t+1, \mathtt{Dyn}(t, x, u, w(t))\big)\Big] . \end{cases}$$

Stochastic multiplicative dynamic programming

$$\mathbb{E}_{w(\cdot)}\Big(\prod_{t=t_0}^{T}\mathbf{1}_{\mathbb{A}(t)}(x(t))\Big)$$

$$\begin{cases} V(T,x) & := & \mathbf{1}_{\mathbb{A}(T)}(x)\;, \\ V(t,x) & := & \mathbf{1}_{\mathbb{A}(t)}(x)\displaystyle\max_{u\in\mathbb{B}(t,x)}\mathbb{E}_{w(t)}\Big[V\Big(t+1,F\big(t,x,u,w(t)\big)\Big)\Big]\;. \end{cases}$$

Outline of the presentation

Whittle criterion

Let us call a criterion Crit in the *Whittle form* whenever it is given by a backward induction of the form:

$$\begin{cases} \mathtt{Crit}\big(t, x(\cdot), u(\cdot), w(\cdot)\big) & = \psi\Big(t, x(t), u(t), w(t), \mathtt{Crit}\big(t+1, x(\cdot), u(\cdot), w(\cdot)\big)\Big) \;, \\ & \quad t = t_0, \ldots, T-1 \;, \\ \mathtt{Crit}\big(T, x(\cdot), u(\cdot), w(\cdot)\big) & = \mathtt{UtilFin}(T, x(T), w(T)) \;. \end{cases}$$

General operator

When $\Omega = \Omega_1 \times \Omega_2$, for an adequate function A, we have

$$\mathbb{G}_{\Omega}[A] = \mathbb{G}_{(w_1,w_2)}\left[A\left(w_1, w_2\right)\right] = \mathbb{G}_{w_1}\left[\mathbb{G}_{w_2}\left[A\left(w_1, w_2\right)\right]\right] .$$

$\mathbb{G}$-linearity

The function $\psi : \{t_0, \dots, T-1\} \times \mathbb{X} \times \mathbb{U} \times \mathbb{W} \times \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is assumed to be $\mathbb{G}$-linear in its last argument in the sense that:

$$\mathbb{G}_{w(t),w(t+1),\dots,w(T)} \left[\psi\left(t, x, u, w(t), A\left(w(t+1), \dots, w(T)\right)\right)\right] =$$

$$\mathbb{G}_{w(t)} \left[\psi\left(t, x, u, w, \mathbb{G}\left[A\left(w(t+1), \dots, w(T)\right)\right]\right)\right] .$$

- When $\mathbb{G}$ is the fear operator $\mathbb{F}$, ψ is assumed to be continuously increasing in its last argument[1].
  - maximin $\psi(t, x, u, w, C) = \min(\texttt{Util}(t, x, u, w), C)$
  - additive $\psi(t, x, u, w, C) = \texttt{Util}(t, x, u, w) + C$
  - multiplicative $\psi(t, x, u, w, C) = \texttt{Util}(t, x, u, w) \times C$
- When $\mathbb{G}$ is the expectation operator $\mathbb{E}$,
  - $\psi(t, x, u, w, C) = g(t, x, u, w) + \beta(t, x, u, w)C$ includes
  - additive $\psi(t, x, u, w, C) = \texttt{Util}(t, x, u, w) + C$
  - multiplicative $\psi(t, x, u, w, C) = \texttt{Util}(t, x, u, w) \times C$

[1] $\psi(t, x, u, w, C^\sharp) \geq \psi(t, x, u, w, C^\flat)$ whenever $-\infty \leq C^\flat \leq C^\sharp \leq +\infty$, and $C_n \to C \Rightarrow \psi(t, x, u, w, C_n) \to \psi(t, x, u, w, C)$.

General dynamic programming equation

$$
\begin{cases}
V(T,x) & := \quad \mathbb{G}_{w \in \mathbb{S}(T)} \left[\texttt{UtilFin}(T,x,w) \right] , \\
V(t,x) & := \quad \max_{u \in \mathbb{B}(t,x)} \mathbb{G}_{w \in \mathbb{S}(t)} \Big[\psi\Big(t,x,u,w, V\big(t+1, \texttt{Dyn}(t,x,u,w)\big) \Big) \Big] .
\end{cases}
$$

Ambiguity?

$$\mathbb{G}_{\Omega}[A] = \min_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}_{w(\cdot)} A(w(\cdot))$$

In general

$$\mathbb{G}_{\Omega}[A] = \mathbb{G}_{(w_1,w_2)}\left[A\left(w_1,w_2\right)\right] \neq \mathbb{G}_{w_1}\left[\mathbb{G}_{w_2}\left[A\left(w_1,w_2\right)\right]\right] .$$

but

$$\begin{cases} \texttt{Crit}\big(t, x(\cdot), u(\cdot), w(\cdot)\big) & = \psi\Big(t, x(t), u(t), \texttt{Crit}\big(t+1, x(\cdot), u(\cdot), w(\cdot)\big)\Big) , \\ & \quad t = t_0, \ldots, T-1 , \\ \texttt{Crit}\big(T, x(\cdot), u(\cdot), w(\cdot)\big) & = \texttt{UtilFin}(T, x(T)) . \end{cases}$$