Reduction of continuous-time control to discrete-time control

A. Jean-Marie

OCOQS Meeting, 24 January 2012

Outline

Progress

Problem statement

Consider some continuous-time, discrete-event, infinite-horizon control problem.

The standard way to analyze such problems is to reduce them to a discrete-time problem using some embedding of a discrete-time process into the continuous-time one.

The optimal policy is deduced from the solution of the discrete-time problem.

Problem statement (ctd)

There are various ways to place the observation points:

- jump instants,
- controllable event instants,
- uniformization instants.

They may result in different value functions.

Question

Is there a way to “play” with the embedding process in order to obtain structural properties of the optimal policy?

Progress

A basic continuous-time control model

As a starting point, consider:

- a continuous-time, piecewise-constant process $\{X(t); t \geq 0\}$ over some discrete state space $\mathcal{X}$;
- a sequence of decision instants $\{T_n; n \in N\}$, endogenous
- a finite set of actions $\mathcal{A}$;
- at a decision point t, given the current state $x = X(t)$, there is a feasible set of actions $\mathcal{A}_x \subset \mathcal{A}$.
  Assuming that action $a \in \mathcal{A}_s$ is applied,
  - a reward $r(x, a, y)$ is obtained;
  - the state jumps to a random $T_a(x)$ with distribution $P_{xay} = \mathbb{P}(T_a(x) = y)$;
  - given y, the next decision point is at $t + \tau$, where τ has an exponential distribution with parameter λ_y.
- between decision points, a reward is accumulated at $\ell(x(t))$, piecewise constant by assumption.

Basic model (ctd.)

Reward criterion: expected total discounted reward. Given $X(0) = x$,

$$J(x) = \mathbb{E}\Big\{ \int_0^\infty e^{-\alpha t} \ell(X(t)) \mathrm{d}t + \sum_{n=1}^{\infty} e^{-\alpha T_n} r(X(T_n^-), A(T_n), X(T_n^+)) \Big\} .$$

The goal is to find the optimal feedback control $d : \mathcal{X} \to \mathcal{A}$ (with the constraint that $d(x) \in \mathcal{A}_x$ for all x) to maximize J.

Basic embedding

Features of this model:

- control is instantaneous and localized in time
- evolution is strictly Markovian
- immediate generalization to semi-Markov decision/transition instants.

Two possibilities for the observation of the process:

- just before a transition/control: $\rightarrow V^{-}(x)$
- just after a transition/control: $\rightarrow V^{+}(x)$

Question:

What is their relation with $J(x)$?

Direct Bellman equations

Conditioning on T_1, the first decision point, we get:

$$\begin{aligned} V^+(x) &= \frac{1}{\alpha + \lambda_x}\left[\ell(x) \;+\; \lambda_x V^-(x)\right] \\ V^-(x) &= \max_{a \in \mathcal{A}_x} \left\{ \sum_y P_{xay}\left(r(x,a,y) + V^+(y)\right) \right\} \\ &= \max_{a \in \mathcal{A}_x} \left\{ \mathbb{E}\left(r(x,a,T_a(x)) + V^+(T_a(x))\right) \right\} \;. \end{aligned}$$

Basic functional equations

Eliminating V^+ or V^- leads to two forms of Bellman's equation:

Bellman Equations

$$\begin{aligned} V^+(x) &= \frac{1}{\alpha + \lambda_x}\Bigg[\ell(x) \\ &\qquad + \lambda_x \max_{a \in \mathcal{A}_x} \sum_y P_{xay}\left[r(x,a,y) + V^+(y)\right]\Bigg] \\ V^-(x) &= \max_{a \in \mathcal{A}_x} \sum_y P_{xay}\Bigg[r(x,a,y) \\ &\qquad + \frac{1}{\alpha + \lambda_y}\left[\ell(y) + \lambda_y V^-(y)\right]\Bigg]. \end{aligned}$$

Progress

Uniformization *à la carte*

For each state x, define $\nu_x \geq \lambda_x$ and introduce a new, uncontrollable transition point after $\tau \sim \text{Exp}(\nu_x)$.
Extend the state space to $\mathcal{X} \times \{r, u\}$,
r = regular event, u = uniformization event.
Table of rewards and transition probabilities:

x'	a	y'	$r(x', a, y')$	$P_{x'ay'}$
(x, r)	a	(y, r)	$r(x, a, y)$	$\frac{\lambda_y}{\nu_y} P_{xay}$
(x, r)	a	(y, u)	$r(x, a, y)$	$\frac{\nu_y - \lambda_y}{\nu_y} P_{xay}$
(x, u)	$*$	(x, r)	0	$\frac{\lambda_y}{\nu_y}$
(x, u)	$*$	(y, u)	0	$\frac{\nu_y - \lambda_y}{\nu_y}$

Running reward: $\ell(x, e) = \ell(x)$; transition rate: $\lambda(x, e) = \nu_x$.

Relationships

Lemma

Let $V(\cdot)$ be the direct value function and $V_u(\cdot,\cdot)$ be the uniformized value function. Then:

$$
\begin{aligned}
V_u^-(x,r) &= V^-(x) \\
V_u^-(x,u) &= V^+(x) \\
V_u^+(x,r) &= \frac{1}{\alpha+\nu_x}\left(\ell(x)+\nu_x V^-(x)\right) \\
V_u^+(x,u) &= \frac{1}{\alpha+\nu_x}\left(\ell(x)+\nu_x V^+(x)\right) .
\end{aligned}
$$

Interpretations

No uniformization ($\lambda_x = \mu_x$):

$$V_u^+(x,r) \;=\; \frac{1}{\alpha+\lambda_x}(\ell(x)+\lambda_x V^-(x)) \;=\; V^+(x)$$

$$V_u^+(x,u) \;=\; \mathbb{E}\left\{\int_0^{T_1} e^{-\alpha u}\ell(x)\mathrm{d}u \;+\; e^{-\alpha T_1}V^+(x)\right\} .$$

Hyper-frequent uniformization ($\nu_x \to \infty$):

$$\lim_{\nu_x\to\infty} V_u^+(x,r) \;=\; V^-(x) \;=\; V_u^-(x,u)$$

$$\lim_{\nu_x\to\infty} V_u^+(x,u) \;=\; V^+(x) \;=\; V_u^-(x,r) .$$

No discounting ($\alpha \to 0$):

$$V_u^+(x,r) \;\sim\; \frac{\ell(x)}{\nu_x} + V^-(x)$$

$$V_u^+(x,u) \;\sim\; \frac{\ell(x)}{\nu_x} + V^+(x) .$$

Bellman equations for the uniformized process

Lemma

The basic value functions V^+ and V^- satisfy:

$$V^+(x) = \frac{1}{\alpha + \nu_x}\Big[\ell(x) + (\nu_x - \lambda_x)V^+(x) + \lambda_x \max_{a \in \mathcal{A}_x} \sum_y P_{xay}\left[r(x,a,y) + V^+(y)\right]\Big]$$

$$V^-(x) = \frac{1}{\alpha + \nu_x}\Big[(\nu_x - \lambda_x)V^-(x) + (\alpha + \lambda_x) \max_{a \in \mathcal{A}_x} \sum_y P_{xay}\Big[r(x,a,y) + \frac{1}{\alpha + \lambda_y}(\ell(y) + \lambda_y V^-(y))\Big]\Big]$$

Progress

The event model

If transitions have several "types", the strictly markovian model requires to extend the state space: $x = (s, e)$ with s the actual system state, and e the event type. We get:

$$
\begin{aligned}
V^+(s,e) &= \frac{1}{\alpha + \lambda_{s,e}} \Big[\ell(s,e) \\
&\quad + \lambda_{s,e} \max_{a \in \mathcal{A}_{s,e}} \sum_{s'} \sum_{e'} P((s,e);a;(s',e')) \\
&\qquad \left\{ r((s,e),a,(s',e')) + V^+(s',e') \right\} \Big] \\
V^-(s,e) &= \max_{a \in \mathcal{A}_{s,e}} \sum_{s'} \sum_{e'} P((s,e);a;(s',e')) \\
&\quad \left[r((s,e),a,(s',e')) + \frac{\ell(s',e') + \lambda_{s',e'} V^-(s',e')}{\alpha + \lambda_{s',e'}} \right]
\end{aligned}
$$

The event model

Question

Under which conditions is it possible to “get rid” of the event part in the state representation.

Is it possible that:

$$V^{+}(s,e) \;=\; V^{+}(s) \qquad \forall e?$$

Progress

Application: arrival control in the $M/M/1$

Let λ and μ denote the arrival and service rates. Reward R for each accepted customer, and (negative) running reward $\ell(s)$ for keeping s customers in queue.

Markovian state: $x \in \mathbb{N} \times \{a, d\}$ (numbered 1/0 in Puterman). The equations for the value function, after uniformization at uniform rate $\lambda + \mu$, are:

$$V_P(s,d) = \frac{1}{\alpha + \lambda + \mu} \left[\ell(s) + \mu V_P((s-1)^+, d) + \lambda V_P(s,a) \right]$$

$$V_P(s,a) = \max \left\{ R + \frac{1}{\alpha + \lambda + \mu} \left[\ell(s+1) + \mu V_P(s,d) + \lambda V_P(s+1,a) \right. \right.$$

$$\left. \frac{1}{\alpha + \lambda + \mu} \left[\ell(s) + \mu V_P(s-1,d) + \lambda V_P(s,a) \right] \right\} .$$

Where is the observation?

But Puterman p. 568 says:

> *The system is in state $< s, 0 >$ if there are s jobs in the system and no arrivals. We observe this state when a transition corresponds to a departure. [...] The state $< s, 1 >$ occurs when there are s jobs in the system and a new job arrives.*

In our notation, this would correspond to setting:

$$\begin{aligned} V_P(s,d) &= V_u^+((s+1,d),r) \\ V_P(s,a) &= V_u^-((s,a),r) \ . \end{aligned}$$

Work in progress....

Lunch time!