Stochastic Interventions and Hybrid Control Models

Jose Luis Menaldi ...based on joint works with ...
H. Jasso-Fuentes (Mexico), T. Prieto-Rumeau (Spain), M. Robin (France)

Department of Mathematics
Wayne State University
Detroit, Michigan 48202, USA
(e-mail: menaldi@wayne.edu)

Stochastic Hybrid Systems - ARO July 8-9 and Oct. 29-30, UConn Storrs, CT

Content

Stochastic Interventions and Hybrid Control Models

No quoted references, see papers mentioned for details ...

Comments: MDPs = Markov decision processes ...

• Notation..., Instantaneous actions, un-discounted costs, infinite horizon.

• Meaning of 'impulse MDPs', the impulse = instantaneous change of state. (inventory model)

• Meaning of 'switching MDPs', the switching = instantaneous change dynamic. (production/maintenance model)

• General 'MDPs with interventions', the intervention = instantaneous change of some fundamental part the model itself.

• (0) read state x_t, (1) chose control a_t, (2) apply dynamics x_{t+1}, and then iterate for $t = 0, 1, \ldots$, also referred to as 'sequential decision'

• Everything is more visible in continuous time, **but** we begin with discrete time models.

Discrete Time Model (MDP)

- Markov decision process: $\mathfrak{M} = (X, A, \mathcal{K}, Q, c, \alpha)$, X (state), A (action), $\mathcal{K} = \{(x, a) \in X \times A : x \in X, a \in A(x)\}$ measurable, w/meas. selectors $\mathfrak{F}$; System dynamics = transition kernel $Q : \mathcal{B}(X) \times \mathcal{K} \to [0, 1]$,

$$Qu(x, a) = \int_X u(y)\, Q(dy \mid x, a), \quad \forall (x, a) \in \mathcal{K}.$$

from $u \in \mathcal{M}^+(X)$ into $Qu \in \overline{\mathcal{M}}^+(\mathcal{K})$.

- Sequential Control: Given initial state $x_0 \in X = H_0$, choose an action $a_0 \in A(x_0)$, a policy $\nu = \{\nu_t\} \in \Pi$ (transition probability measures on A given H_t), such that $\nu_t(A(x_t) \mid h_t) = 1$ for all $h_t \in H_t = \mathcal{K}^t \times X$ and $t = 0, 1, \ldots$, with $H_t = \mathcal{K}^t \times X$ (=history up to time t) and $H_\infty = \mathcal{K}^\infty$, with $h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$, $\omega = (x_0, a_0, \ldots, x_t, a_t, \ldots) \in H_\infty$.
- Expected discounted cost: $c \geq 0$ running cost, $0 \leq \alpha \leq 1$ discount

$$J(x, \nu) = \mathbb{E}_x^\nu \Big\{ \sum_{t=0}^{\infty} c(x_t, a_t) \prod_{j=0}^{t-1} \alpha(x_j, a_j) \Big\}, \quad J^*(x) = \inf_\nu J(x, \nu).$$

Comments 1:

- Under these assumptions, a probability P_x^ν on H_∞ is constructed.
- There are other 'equivalent' ways of presenting this model, e.g.,

$$x_{t+1} = F(x_t, a_t, w_t), \quad a_t = \nu_t(w_{-1}, x_0, a_0, w_0, \ldots, x_{t-1}, a_{t-1}, w_{t-1}, x_t),$$

where $\{w_t\}$ is the 'disturbance sequence' (or IID noise), $a_0 = \nu_0(w_{-1}, x_0)$.

- A typical/classic discounted cost looks like

$$J(x, \nu) = \mathbb{E}_x^\nu\Big\{ \sum_{t=0}^{\infty} c(x_t, a_t)\, \alpha^t \Big\}, \quad 0 < \alpha < 1,$$

our point is variable discount factor $\alpha(x, a)$

Assumption MDP: (A1)

(1) Running cost $c \in \mathcal{L}^+(\mathcal{K})$ (lsc), $x \mapsto \inf_{a \in A(x)} c(x,a)$ is bounded.
(2) Discount factor $\alpha \in \mathcal{L}^+(\mathcal{K})$ (by definition, α takes values in $[0,1]$).
(3) Q is weakly continuous, i.e, $u \in C_b(X)$ implies $Qu \in C_b(\mathcal{K})$.
(4) $x \mapsto A(x)$ is compact-valued and upper semicontinuous.
(5) The optimal discounted cost $J^*(x)$ is finite for every $x \in X$. $\square$

Note: (1) $= \exists\, f \in \mathfrak{F}$ such that $\sup_{x \in X} c(x, f(x)) < \infty$, but c may be unbounded. Later, we show examples of (5) satisfied in term of data. Also (2)–(4) are standard. There are only a few references with variable discount, always $0 < \epsilon \leq \alpha(x,a) \leq 1 - \epsilon$. Main interest, why?

Dynamic Programming and Comments 2:

Given $u \in \mathcal{M}^+(X)$ define $Tu(x) = \inf_{a \in A(x)} \{c(x,a) + \alpha(x,a)Qu(x,a)\}$, $\forall x \in X$, could be $+\infty$. If $Tu(x) = c(x, f(x)) + \alpha(x, f(x))Qu(x, f(x))$, $\forall x \in X$ with $f \in \mathfrak{F}$, then T admits a measurable selector at u.

The *dynamic programming equation* (DPE)

$$u(x) = Tu(x) \quad \text{for each } x \in X, \tag{1}$$

and look for solution $u \in \mathcal{M}^+(X)$.

- First, need to relate the DPE (1) with the optimal cost $J^*(x)$.
- Need 'general Assumptions' to accommodate instantaneous controls ... Hybrid models.
- As seen later, variable discount factor is key element!

Value Iteration Procedure (VIP)

Under Assumption MDP

• If $u \in \mathcal{L}^+(X)$ then $Tu \in \overline{\mathcal{L}}^+(X)$ and T has a measurable selector at u, and if $u \in \mathcal{L}_b^+(X)$ then $Tu \in \mathcal{L}_b^+(X)$.
• VIP: Define $v_{k+1} = Tv_k$ for $k \geq 0$ beginning with $v_0 = 0$. Then, $\{v_k\}$ converges pointwise and monotonically to some $v^* \in \mathcal{L}^+(X)$ with $v^* \leq J^*$.
• The limiting function v^* is a solution of the DPE (1). $\square$

Theorem

Suppose Assumption MDP and let $v^* \in \mathcal{L}^+(X)$ be the limiting function as above. Then the optimal discounted cost J^* equals v^*, and it is the minimal solution in $\mathcal{L}^+(X)$ of the DPE (1). Moreover, any measurable selector of T at J^* is an optimal deterministic stationary policy ν^* (i.e., $\exists\, f^* \in \mathfrak{F} \mid \nu_t^*(B \mid h_t) = \delta_{f^*(x_t)}(B)$, $\forall B \in (\mathcal{B}(X)$, $h_t \in H_t$, $t = 0, 1, \ldots)$.

Comments 3:

- Theorem gives satisfaction from the DPE view point.

- Need to translate the 'abstract' condition (5) (i.e., $J^*(x) < \infty\ \forall x \in X$) in Assumption (A1) to 'something' on the data of the model ...

- Assumptions on the variable discount $\alpha(x, a)$ ...

- Optimal cost $J^* = v^* \in \mathcal{L}^+(X)$, but we would like $\in \mathcal{L}_b^+(X)$

Definition: $f \in \mathfrak{F}$, $C \in \mathcal{B}(X)$ is *small* if $\exists\, t \in \mathbb{N}$ and a nontrivial measure μ on $(X, \mathcal{B}(X))$ with $P_x^f\{x_t \in B\} \geq \mu(B)$ for all $B \in \mathcal{B}(X)$ and $x \in C$. We will also say that C is μ-small for $f \in \mathfrak{F}$ at stage (or time) t. $\square$

- For 'small sets' see Meyn and Tweedie's book (CUP 2009), this is 'similar/related' to some 'ergodic' conditions ...

Variable Discount Assumption: (A2) [$|$ = such that]

For $f \in \mathfrak{F}$ and $0 \le \beta \le 1$ set $L(U)_{f,\beta} = \{x \in X : \alpha(x, f(x)) \le (>) \beta\}$.
(a) $\exists\ f \in \mathfrak{F} \mid \sup_{x \in X} c(x, f(x)) = \mathfrak{c} < \infty$, and $\mid \exists\ 0 \le \beta < 1$ and μ measure on $(X, \mathcal{B}(X)) \mid \mu(L_{f,\beta}) > 0$, so that $U_{f,\beta}$ is μ-small for f at t, i.e., $\exists\, t \mid\ P_x^f\{x_t \in B\} \ge \mu(B),\ \forall B \in \mathcal{B}(X),\ x \in U_{f,\beta}$.
(b) $\exists\ 0 < \delta < 1 \mid \forall (x, a) \in \mathcal{K} : \alpha(x, a) \ge 1 - \delta \Rightarrow c(x, a) \ge \delta$, i.e., roughly speaking, discount factors close to one yield a running positive cost. $\square$

Proposition

(i) *If (A2)(a) holds then* $\sup_{x \in X} J(x, f) < \infty$, *i.e., (A1)(5) is satisfied.*
(ii) *Under (A2)(b), if* $x \in X$ *and* $\nu \in \Pi$ *are such that* $J(x, \nu) < \infty$ *then* $\prod_{i=0}^{k} \alpha(x_i, a_i)$ *converges to 0, with* P_x^ν*-probability one as* $k \to \infty$.

Theorem

If (A1) and (A2) hold then $J^* \in \mathcal{L}_b^+(X)$. *Furthermore it is the unique solution in* $\mathcal{L}_b^+(X)$ *of the DPE (1).*

Comments 4 - This Model includes:

- Considering 'instantaneous actions'.
- Such as impulses, switchings, hybrid models.
- Stopping the dynamics at τ.
- Conditional stopping actions, like $\mathfrak{a}$.
- Variable discount factor (its role!).
- Absorption actions a_∂ and absorption states ∂.
- Markov Decision Processes/DPE with Stopping or VI.
- Markov Decision Processes/DPE with Absorption.

Switching MDP with Stopping

Switching is an instantaneous change of regime or mode or configuration of the dynamics (better known in continuous time).
In our context, there are $N = \{i = 1, \ldots, N\}$ controlled Markov chains, all taking values in Y and with common action space V; both Y and V are Borel spaces. Transition kernels $Q_i(B \mid y, a)$, $\forall B \in \mathcal{B}(Y)$, $\forall (y, a) \in Y \times A$.

Controlled switching model with stopping: $X = (Y \times N) \cup \{\partial\}$ is the State space, where ∂ is an isolated absorption state. In the pair $(y, k) \in X$, $y \in Y$ is the 'fast' variable (main state of the system) and $k \in N$ is the 'slow' variable indicating the current regime/mode/configuration.
Action space $= V \cup \{a_\partial\} \cup N$, the latter being $N + 1$ isolated points. The set of available actions are

$$A(y, k) = \tilde{A}(y) \cup \{1, \ldots, k-1, k+1, \ldots, N\}, \quad \tilde{A}(y) \subset V \cup \{a_\partial\}$$

for a state $(y, k) \in Y \times N$ and $A(\partial) = \{a_\partial\}$, i.e., a_∂ is an absorption(-stopping) action.

Dynamics of Switching MDP with Stopping

Starting from a state $(y, k) \in X$ we have

$$Q(dz \times dm \mid (y,k), a) = \begin{cases} Q_k(dz \times dm \mid y, a) & \text{if } a \in V, \\ \delta_{(y,a)}(dz \times dm) & \text{if } 1 \le a \le N, a \neq k, \\ \delta_\partial(dz) & \text{if } a = a_\partial, \end{cases}$$

note that $Q(\{\partial\}|(y,k),a) = 0$ for any $(y,k) \in Y \times N$ and $a \in A(y,k)$, while $Q(\{\partial\}|\partial, a_\partial) = 1$.
A typical or common situation is

$$Q(dz \times dm \mid (y,k), a) = Q_k(dz \mid y, a)\, \delta_{\{k\}}(dm) \quad \text{if } a \in V,$$

i.e., when modes remain constant except when a switching is applied.

Switching Costs

Running costs $\ell_i : Y \times V \to \mathbb{R}^+$ under each of the N controlled Markov chains; switching cost $l : N \times N \mapsto \mathbb{R}^+$ and terminal cost $\ell_0 : Y \to \mathbb{R}^+$. Thus, $c(\partial, a_\partial) = 0$ and

$$c(y,k,a) = \ell_k(y,a)1_V(a) + \ell_0(y)1_{\{a_\partial\}}(a) + l(k,a)1_{\{1,\dots,N\}}(a),$$

represents cost for state $(y,k) \in Y \times N$ and action $a \in V \cup \{a_\partial\} \cup N$.

Discount factor is $\alpha(\cdot, a_\partial) = 0$ and

$$\alpha(y,k,a) = \varrho_k(y,a)1_V(a) + 1_{\{1,\dots,N\}}(a), \quad \forall (y,k) \in Y \times N.$$

where $\varrho_k : Y \times V \to (0,1)$, for $1 \le k \le N$, are the discount functions of each of the N controlled Markov chains.

All, under the assumptions (A1) and (A2).

DPE with Switching and Stopping

$u : Y \times N \to \mathbb{R}^+$ solves the dynamic programming equation (DPE) if for $(y,k) \in Y \times N$ we have

$$u(y,k) = \min\left\{\mathcal{M}u(y,k), \mathcal{H}u(y,k), \ell_0(y)\right\},$$

where the operators $\mathcal{H}$ and $\mathcal{M}$ are defined as

$$\begin{aligned}\mathcal{H}u(y,k) &:= \inf_{a \in A(y,k)\cap V}\left\{\ell_k(y,a) + \varrho_k(y)\int_X u(z,k)Q_k(dz|y,a)\right\},\\ \mathcal{M}u(y,k) &:= \min_{a \in A(y,k)\cap(N\setminus\{k\})}\left\{l(k,a) + u(y,a)\right\},\end{aligned}$$

we put $u(\partial) = 0$.

Typical example: pollution accumulation problem.

• Joint work with H. Jasso-Fuentes and T. Prieto-Rumeau, Discrete-time Control with Non-constant Discount Factor Mathematical Methods of Operations Research, 2020, 92, 377–399. (and several others 2017 – 2020)

Impulse/Switching/Hybrid Control - Comments 6:

- Contrast with continuous time models.
- Impulse versus Switching control.
- Hybrid control: Example 'wait for a signal' constraint.
- Back to discrete time models ...
- Specifying the model in between two consecutive 'ticks of a clock'.
- Back to the 'black box' ...

Continuous Time Models

Impulse Control $\nu = \{\theta_i, \xi_i : i = 1, 2, \ldots\}$ produces (x_t^ν, y_t^ν).
y_t = 'signal clock', i.e., y_t = 'time since last signal' or 'waiting time'
$\{(x_t^0, y_t^0) : t \geq 0\}$ is the uncontrolled Markov evolution (of the state) and $\{(x_t^i, y_t^i) : t \geq \theta_i\}$ denotes the Markov evolution after the i-impulse, i.e., only the first i impulses are applied and the Markov process restart anew at time $\theta_i < \infty$ with initial condition $(x_{\theta_i}^i, y_{\theta_i}^i) = (\xi_i, 0)$, since $y_{\theta_i}^{i-1} = 0$.
Also the sequence $\{\tau_k^i : k \geq 1\}$ of signals after θ_i is given by the functional $\tau_{k+1}^i = \inf\{t > \tau_k^i : y_t^i = 0\}$, beginning with $\tau_0^i = \theta_i < \infty$.
Control is allowed only when the signal arrives
Costs: running, impulse, and optimal $u(x, y) = \inf_\nu\{J_{x,y}\}$, with

$$J_{x,y}(\nu) = \mathbb{E}_{x,y}^\nu\Big\{ \int_0^\infty \mathrm{e}^{-\alpha t} f(x_t, y_t)\mathrm{d}t + \sum_{i=0}^\infty \mathrm{e}^{-\alpha\theta_i} c(x_{\theta_i}^{i-1}, \xi_i)\Big\},$$

$x_{\theta_i}^{i-1}$ is the value of the process just before the impulse.

Assumption (ACT)

E (locally) compact Polish space, $(\Omega, \mathbb{F}, x_t, y_t, P_{xy})$ Markov(-Feller) process on $\Omega = D(\mathbb{R}^+, E \times \mathbb{R}^+)$, filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$, (x_t, y_t) canonical process with values in $E \times \mathbb{R}^+$, infinitesimal generator $A_{xy} = A_x + A_y$.

(a) x_t is a Markov process by itself (*reduced state*), with a C_0-semigroup $\Phi_x(t)$ (i.e., $\Phi_x(t) C_0(E) \subset C_0(E)$, $\forall t \geq 0$), and infinitesimal generator A_x with domain $\mathcal{D}(A_x) \subset C_0(E)$.

(b) $A_y \varphi(x,y) = \partial_y \varphi(x,y) + \lambda(x,y)[\varphi(x,0) - \varphi(x,y)]$ is the infinitesimal generator of the *signal* process y_t, $\lambda \geq 0$ and $\lambda \in C_b(E \times \mathbb{R}^+)$, i.e., it has jumps to zero at times $\tau_1, \ldots, \tau_n \to \infty$ and $y_t = t - \tau_n$ for $\tau_n \leq t < \tau_{n+1}$.

(c) Beside $P_{x0}\{\tau_1 < \infty\} = 1$, for some constant $K > 0$,

$$\mathbb{E}_{x0}\{\tau_1\} := \mathbb{E}_x\Big\{ \int_0^\infty t\lambda(x_t, t) \exp\Big(- \int_0^t \lambda(x_s, s) \mathrm{d}s \Big) \mathrm{d}t \Big\} =$$

$$= \mathbb{E}_x\Big\{ \int_0^\infty \exp\Big(- \int_0^t \lambda(x_s, s) \mathrm{d}s \Big) \mathrm{d}t \Big\} \leq K, \quad \forall x. \quad \square$$

Comments 7:

• In a continuous time models, keeping the system as a controlled Markov-Feller process is really very general. (!?) [certainly including switching diffusion processes with jumps]

• It can be argued that provided a very large state-space, any process is Markov. (!?)

• It can be argued that if the ‘continuity’ (of the Feller property) is dropped, then the model may have an incomplete description, this means that the discontinuity should be understood and resolved. (!?)

Dynamic Programming Principle

Usual conditions on costs f, c, and on $\Gamma(x)$ (impulse at x) and on $Mv(x) = \inf_{\xi \in \Gamma(x)} \{c(x,\xi) + v(\xi)\}$, the jumps operator. □

- Dynamic Programming Principle, with $u_0(x) = u_0(x,0)$, shows

$$u(x,y) = \mathbb{E}_{xy}\Big\{ \int_0^\tau \mathrm{e}^{-\alpha t} f(x_t,y_t)\mathrm{d}t + \mathrm{e}^{-\alpha\tau} \min\{Mu,u\}(x_\tau,y_\tau)\Big\}, \tag{2}$$

$$\begin{aligned} u(x,y) &= \mathbb{E}_{xy}\Big\{ \int_0^\tau \mathrm{e}^{-\alpha t} f(x_t,y_t)\mathrm{d}t + \mathrm{e}^{-\alpha\tau} u(x_\tau,y_\tau)\Big\}, \quad y>0, \\ u_0(x) &= \min\Big\{\mathbb{E}_{x0}\Big\{ \int_0^\tau \mathrm{e}^{-\alpha t} f(x_t,y_t)\mathrm{d}t + \mathrm{e}^{-\alpha\tau} u_0(x_\tau)\Big\}, Mu_0(x)\Big\}, \end{aligned} \tag{3}$$

$$u(x,y) = \mathbb{E}_{xy}\Big\{ \int_0^\tau \mathrm{e}^{-\alpha t} f(x_t,y_t)\mathrm{d}t + \mathrm{e}^{-\alpha\tau} u_0(x_\tau)\Big\}, \tag{4}$$

and so, if $u_0(x) = u(x,0)$ is known then the last equality yields $u(x,y)$.

- However, getting u_0 is like solving a discrete time problem.

Switching Control

(1) reduced state $x_t = (x'_t, n_t)$ belongs to $E = E' \times N$;
(2) x'_t is a Markov process with a C_0-semigroup $\Phi^i(t)$ when $n = i$;
(3) n_t is a Markov chain with generator $Q = (q_{ij})$, i.e., $A_x = A_{x'} + Q$.
(could have n_t a semi-Markov process and $N = \{1, 2, \ldots\}$)

A switching control is a particular case of impulse control (conversely) but!

• However the specific assumptions make same differences, specifically, the switching cost. Moreover, there are more differences in a ergodic cost setting.

• Impulse and Switching Models (with and without constraint) are particular cases of Hybrid control Models.

• Joint work with M Robin, Hybrid Models and Switching Control with Constraints Communications on Stochastic Analysis, 2019, 13, 1–29. (and several others, 2016 – 2020)

• ACTUALLY, THERE ARE MANY OTHER REFERENCES QUOTED IN THE ABOVE WORKS !!!