Unconstrained and Constrained Optimal Control of Piecewise Deterministic Markov Processes

Oswaldo Costa, François Dufour, Alexey Piunovskiy

Universidade de Sao Paulo

Institut de Mathématiques de Bordeaux
INRIA Bordeaux Sud-Ouest

University of Liverpool

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02)

Outline

Introduction

Davis (80's)

General class of non-diffusion stochastic hybrid models:
deterministic trajectory punctuated by random jumps.

Applications

Engineering systems, biology, operations research, management science, economics, dependability and safety, . . .

Parameters of the model

- the state space: $\mathbf{X}$ open subset of $\mathbb{R}^d$ (boundary $\partial\mathbf{X}$).
- the flow: $\phi(x,t) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ satisfying
  $\phi(x,t+s) = \phi(\phi(x,s),t)$ for all $x \in \mathbb{R}^d$ and $(t,s) \in \mathbb{R}^2$.
  $\to$ active boundary:
  $\Delta = \{z \in \partial\mathbf{X} : z = \phi(x,t) \text{ for some } x \in \mathbf{X} \text{ and } t \in \mathbb{R}_+^*\}$.
  For $x \in \overline{\mathbf{X}} \doteq \mathbf{X} \cup \Delta$,

$$t^*(x) = \inf\{t \in \mathbb{R}_+ : \ \phi(x,t) \in \Delta\}.$$

- $\mathbf{A}$ is the action space, assumed to be a Borel space.
  $\mathbf{A}^g \in \mathcal{B}(\mathbf{A})$ (respectively $\mathbf{A}^i \in \mathcal{B}(\mathbf{A})$) is the set of *gradual* or *continuous* (respectively *impulsive*) actions satisfying
  $\mathbf{A} = \mathbf{A}^i + \mathbf{A}^g$.

Parameters of the model

- The set of *feasible* actions in state $x \in \overline{\mathbf{X}}$ is $\mathbf{A}(x) \subset \mathbf{A}$. Let us introduce the following sets $\mathbf{K} = \mathbf{K}^i \cup \mathbf{K}^g$ with

$$\mathbf{K}^g = \{(x, a) \in \mathbf{X} \times \mathbf{A}^g : a \in \mathbf{A}(x)\}$$

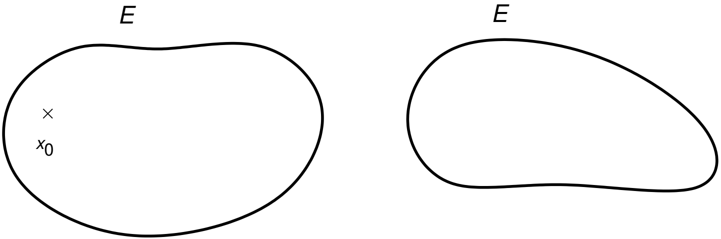
$$\mathbf{K}^i = \{(x, a) \in \Delta \times \mathbf{A}^i : a \in \mathbf{A}(x)\}$$

- The jumps intensity λ which is a $\mathbb{R}_+$-valued measurable function defined on $\mathbf{K}^g$.
- The stochastic kernel Q on $\mathbf{X}$ given $\mathbf{K}$ satisfying $Q(\mathbf{X} \setminus \{x\}|x, a) = 1$ for any $(x, a) \in \mathbf{K}^g$. It describes the state of the process after any jump.

Uncontrolled process

Definition of a PDMP

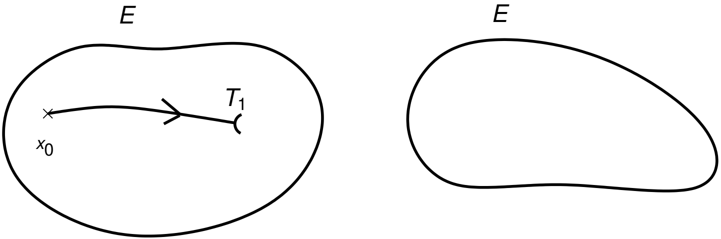
Parameters: flow ϕ, intensity of the jumps λ, transition kernel Q



Uncontrolled process

Definition of a PDMP

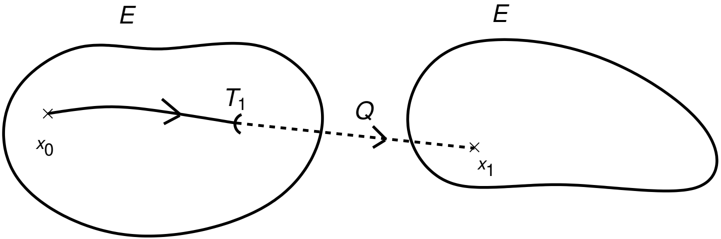
Parameters: flow ϕ, intensity of the jumps λ, transition kernel Q



Uncontrolled process

Definition of a PDMP

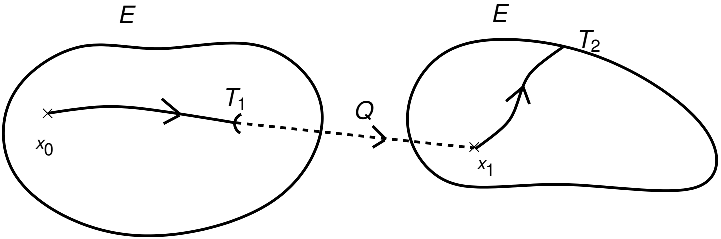
Parameters: flow ϕ, intensity of the jumps λ, transition kernel Q



Uncontrolled process

Definition of a PDMP

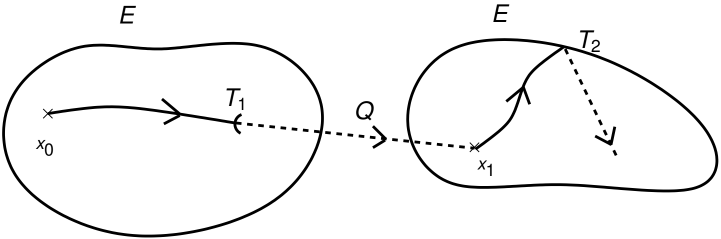
Parameters: flow ϕ, intensity of the jumps λ, transition kernel Q



Uncontrolled process

Definition of a PDMP

Parameters: flow ϕ, intensity of the jumps λ, transition kernel Q



Construction of the controlled process

The canonical space $\Omega = \left(\bigcup_{n=1}^{\infty} \Omega_n\right) \bigcup \left(\mathbf{X} \times (\mathbb{R}_+^* \times \mathbf{X})^\infty\right)$ with $\Omega_n = \mathbf{X} \times (\mathbb{R}_+^* \times \mathbf{X})^n \times (\{\infty\} \times \{x_\infty\})^\infty$.

Introduce the mappings $X_n : \Omega \to \mathbf{X}_\infty = \mathbf{X} \cup \{x_\infty\}$ by $X_n(\omega) = x_n$ and $\Theta_n : \Omega \to \overline{\mathbb{R}}_+^*$ by $\Theta_n(\omega) = \theta_n$; $\Theta_0(\omega) = 0$ where

$$\omega = (x_0, \theta_1, x_1, \theta_2, x_2, \dots) \in \Omega.$$

In addition $T_n(\omega) = \sum_{i=1}^{n} \Theta_i(\omega) = \sum_{i=1}^{n} \theta_i$ with $T_\infty(\omega) = \lim_{n \to \infty} T_n(\omega)$.

$\mathbf{H}_n$ is the set of path up to n.
$H_n = (X_0, \Theta_1, X_1, \dots, \Theta_n, X_n)$ is the history of the process up to n.

Construction of the process

The controlled process $\{\xi_t\}_{t\in\mathbb{R}_+}$:

$$\xi_t(\omega) = \begin{cases} \phi(X_n, t - T_n) & \text{if } T_n \leq t < T_{n+1} \text{ for } n \in \mathbb{N}; \\ x_\infty, & \text{if } T_\infty \leq t. \end{cases}$$

The flow is not controlled.

Admissible strategies and conditional distribution

An admissible control strategy is a sequence $u = (\pi_n, \gamma_n)_{n \in \mathbb{N}}$ such that, for any $n \in \mathbb{N}$,

- π_n is a stochastic kernel on $\mathbf{A}^g$ given $\mathbf{H}_n \times \mathbb{R}_+^*$: $\pi_n(da|h_n, t) = 1$ for $t \in]0, t^*(x_n)[$,
- γ_n is a stochastic kernel on $\mathbf{A}^i$ given $\mathbf{H}_n$: $\gamma_n(da|h_n) = 1$

where $h_n = (x_0, \theta_1, x_1, \ldots \theta_n, x_n) \in \mathbf{H}_n$.

The set of admissible control strategies is denoted by $\mathcal{U}$.

Admissible strategies and conditional distribution

For an admissible control strategy $u = (\pi_n, \gamma_n)_{n\in\mathbb{N}}$, we can equivalently consider the random processes with values in $\mathcal{P}(\mathbf{A}^g)$ and $\mathcal{P}(\mathbf{A}^i)$ respectively as

$$\pi(da|t) = \sum_{n\in\mathbb{N}} I_{\{T_n<t\leq T_{n+1}\}}\pi_n(da|H_n, t - T_n)$$

and

$$\gamma(da|t) = \sum_{n\in\mathbb{N}} I_{\{T_n<t\leq T_{n+1}\}}\gamma_n(da|H_n),$$

for $t \in \mathbb{R}_+^*$.

Admissible strategies and conditional distribution

Interaction of $u = (\pi_n, \gamma_n)_{n\in\mathbb{N}}$ and the parameters of the model:

- the intensity of jumps

$$\lambda_n^u(h_n, t) = \int_{\mathbf{A}^g} \lambda(\phi(x_n, t), a)\pi_n(da|h_n, t),$$

and the corresponding rate of jumps

$$\Lambda_n^u(h_n, t) = \int_{]0,t]} \lambda_n^u(h_n, s)ds,$$

- the distribution of the state after a (*stochastic*) jump

$$Q_n^{g,u}(dx|h_n, t) = \frac{1}{\lambda_n^u(h_n, t)} \int_{\mathbf{A}^g} Q(dx|\phi(x_n, t), a)\lambda(\phi(x_n, t), a)\pi_n(da|h_n, t)$$

- the distribution of the state after a (*boundary*) jump

$$Q_n^{i,u}(dx|h_n) = \int_{\mathbf{A}^i} Q(dx|\phi(x_n, t^*(x_n)), a)\gamma_n(da|h_n).$$

Admissible strategies and conditional distribution

We want the joint distribution of the next sojourn time and state be given by G_n

$$
\begin{aligned}
& G_n(\Gamma_1 \times \Gamma_2 | h_n) \\
& \quad = \Big[I_{\{x_n = x_\infty\}} + e^{-\Lambda_n^u(h_n,+\infty)} I_{\{x_n \in \mathbf{X}\}} I_{\{t^*(x_n)=\infty\}} \Big] \delta_{(+\infty, x_\infty)}(\Gamma_1 \times \Gamma_2) \\
& \quad + I_{\{x_n \in \mathbf{X}\}} \Big[\delta_{t^*(x_n)}(\Gamma_1) Q_n^{i,u}(\Gamma_2 | h_n) e^{-\Lambda_n^u(h_n, t^*(x_n))} I_{\{t^*(x_n) < \infty\}} \\
& \quad + \int_{]0, t^*(x_n)[\cap \Gamma_1} Q_n^{g,u}(\Gamma_2 | h_n, t) \lambda_n^u(h_n, t) e^{-\Lambda_n^u(h_n,t)} dt \Big],
\end{aligned}
$$

where $\Gamma_1 \in \mathcal{B}(\overline{\mathbb{R}}_+^*)$, $\Gamma_2 \in \mathcal{B}(\mathbf{X}_\infty)$ and
$h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \mathbf{H}_n$.

Admissible strategies and conditional distribution

Consider an admissible strategy $u \in \mathcal{U}$ and an initial state $x_0 \in \mathbf{X}$

$$\mathbb{P}^u_{x_0}\Big((\Theta_{n+1}, X_{n+1}) \in \Gamma_1 \times \Gamma_2 \big| \mathcal{F}_{T_n}\Big) \overset{?}{=} G_n\big(\Gamma_1 \times \Gamma_2 \big| H_n\big)$$

$\Longrightarrow$ the conditional distribution of (Θ_{n+1}, X_{n+1}) given $\mathcal{F}_{T_n}$ under $\mathbb{P}^u_{x_0}$ is $G_n(\cdot|H_n)$ ($\{\mathcal{F}_t\}$ is the natural filtration of the process).

Admissible strategies and conditional distribution

Consider an admissible strategy $u \in \mathcal{U}$ and an initial state $x_0 \in \mathbf{X}$. There exists a probability $\mathbb{P}^u_{x_0}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}^u_{x_0}(\{X_0 = x_0\}) \quad = \quad 1$$

and the positive random measure ν defined on $\mathbb{R}^*_+ \times \mathbf{X}$ by

$$\nu(dt, dx) = \sum_{n \in \mathbb{N}} \frac{G_n(dt - T_n, dx | H_n)}{G_n([t - T_n, +\infty] \times \mathbf{X}_\infty | H_n)} I_{\{T_n < t \leq T_{n+1}\}}$$

is the compensator of

$$\mu(dt, dx) = \sum_{n \geq 1} I_{\{T_n(\omega) < \infty\}} \delta_{(T_n(\omega), X_n(\omega))}(dt, dx).$$

with respect to $\mathbb{P}^u_{x_0}$ (Jacod, *Multivariate point processes*, 1975).

Outline

1. Piecewise deterministic Markov processes
   - Introduction
   - Parameters of the model
   - Construction of the controlled process
   - Admissible strategies
2. Optimization problems
   - Unconstrained and constrained problems
   - Assumptions
3. Non explosion
4. The unconstrained problem and the dynamic programming approach
5. The constrained problem and the linear programming approach

Unconstrained and constrained problems

Cost functions

- $(C_j^g)_{j\in\{0,1,\ldots,p\}}$ associated with a continuous action. Real-valued mapping defined on $\mathbf{K}^g$.
- $(C_j^i)_{j\in\{0,1,\ldots,p\}}$ associated with an impulsive action on the boundary. Real-valued mapping defined on $\mathbf{K}^i$.

The associated infinite-horizon discounted criteria corresponding to an admissible control strategy $u \in \mathcal{U}$ are given by

$$\mathcal{V}_j(u, x_0) = \mathbb{E}^u_{x_0}\bigg[\int_{]0,+\infty[} e^{-\alpha s}\int_{\mathbf{A}(\xi_s)} C_j^g(\xi_s, a)\pi(da|s)ds\bigg] + \mathbb{E}^u_{x_0}\bigg[\int_{]0,+\infty[} e^{-\alpha s} I_{\{\xi_{s-}\in\Delta\}}\int_{\mathbf{A}(\xi_{s-})} C_j^i(\xi_{s-}, a)\gamma(da|s)\mu(ds, \mathbf{X})\bigg]$$

for any $j \in \{0, 1, \ldots, p\}$.

Unconstrained and constrained problems

- The optimization problem without constraint consists in minimizing the performance criterion

$$\inf_{u \in \mathcal{U}} \mathcal{V}_0(u, x_0).$$

- The optimization problem with p constraints consists in minimizing the performance criterion

$$\inf_{u \in \mathcal{U}} \mathcal{V}_0(u, x_0)$$

  such that the constraint criteria

$$\mathcal{V}_j(u, x_0) \leq B_j$$

  are satisfied for any $j \in \mathbb{N}_p^*$, where $(B_j)_{j \in \mathbb{N}_p^*}$ are real numbers representing the constraint bounds.

Different classes of strategies

- *feasible*, if $u \in \mathcal{U}$ and $\mathcal{V}_j(u, x_0) \leq B_j,$ for $j \geq 1$.
- *stationary*, if for some $(\pi, \gamma) \in \mathcal{P}^g \times \mathcal{P}^i$ the control strategy $u = (\pi_n, \gamma_n)_{n \in \mathbb{N}}$ is given by $\pi_n(da|h_n, t) = \pi(da|\phi(x_n, t))$ and $\gamma_n(db|h_n) = \gamma(db|\phi(x_n, t^*(x_n)))$.
- *non-randomized stationary*, if $\pi_n(\cdot|h_n, t) = \delta_{\varphi^s(\phi(x_n,t))}(\cdot)$ and $\gamma_n(\cdot|h_n) = \delta_{\varphi^s(\phi(x_n,t))}(\cdot)$, where $\varphi^s : \overline{\mathbf{X}} \to \mathbf{A}$ is a measurable mapping satisfying $\varphi^s(y) \in \mathbf{A}(y)$ for any $y \in \overline{\mathbf{X}}$.

Hypotheses

Assumption A. There are constants $K \geq 0, \varepsilon_1 > 0$ and $\varepsilon_2 \in [0,1[$ such that

(A1) *For any* $(x,a) \in \mathbf{K}^g$, $\lambda(x,a) \leq K$

(A2) $\inf_{(z,b)\in \mathbf{K}^i} Q(A_{\varepsilon_1}|z,b) \geq 1-\varepsilon_2$, *with* $A_{\varepsilon_1} = \{x \in \mathbf{X} : t^*(x) > \varepsilon_1\}$.

Assumption B.

(B1) *The set* $\mathbf{A}(y)$ *is compact for every* $y \in \overline{\mathbf{X}}$.

(B2) *The kernel* Q *is weakly continuous.*

(B3) *The function* λ *is continuous on* $\mathbf{K}^g$.

(B4) *The flow* ϕ *is continuous on* $\mathbb{R}_+ \times \mathbb{R}^p$.

(B5) *The function* t^* *is continuous on* $\overline{\mathbf{X}}$.

Assumption C.

(C1) The multifunction Ψ^g *from* $\mathbf{X}$ *to* $\mathbf{A}$ *defined by* $\Psi(x) = \mathbf{A}(x)$ *is upper semicontinous. The multifunction* Ψ *from* Δ *to* $\mathbf{A}$ *defined by* $\Psi^i(z) = \mathbf{A}(z)$ *is upper semicontinous.*

(C2) The cost function C_0^g *(respectively,* C_0^i*) is bounded and lower semicontinuous on* $\mathbf{K}^g$ *(respectively,* $\mathbf{K}^i$*).*

Outline

1. Controlled piecewise deterministic Markov processes
   - Introduction
   - Parameters of the model
   - Construction of the process
   - Admissible strategies
2. Optimization problems
   - Unconstrained and constrained problems
   - Different classes of strategies
   - Hypotheses
3. Non explosion
4. The unconstrained problem and the dynamic programming approach
5. The constrained problem and the linear programming approach

Lemma

Suppose Assumption A is satisfied. Then there exists $M < \infty$ *such that, for any control strategy* $u \in \mathcal{U}$ *and for any* $x_0 \in \mathbf{X}$

$$\mathbb{E}^u_{x_0}\Big[\sum_{n\in\mathbb{N}^*} e^{-\alpha T_n}\Big] \leq M \text{ and } \mathbb{P}^u_{x_0}(T_\infty < +\infty) = 0.$$

Elements of proof:

- For any control strategy u, $x_0 \in \mathbf{X}$ we have for any $j \in \mathbb{N}$
$$\mathbb{P}^u_{x_0}(\Theta_{j+2} + \Theta_{j+1} > \varepsilon_1 | H_j) \geq e^{-2K\varepsilon_1}(1 - \varepsilon_2).$$
- Now,
$$\begin{aligned}
&\mathbb{E}^u_{x_0}\Big[e^{-\alpha(\Theta_{j+1}+\Theta_{j+2})}|H_j\Big] \\
&\quad \leq \mathbb{P}^u_{x_0}(\Theta_{j+1} + \Theta_{j+2} \leq \varepsilon_1 | H_j) \\
&\qquad + e^{-\alpha\varepsilon_1}\mathbb{P}^u_{x_0}(\Theta_{j+1} + \Theta_{j+2} > \varepsilon_1 | H_j) \\
&\quad = 1 + [e^{-\alpha\varepsilon_1} - 1]\mathbb{P}^u_{x_0}(\Theta_{j+1} + \Theta_{j+2} > \varepsilon_1 | H_j) \\
&\quad \leq 1 + [e^{-\alpha\varepsilon_1} - 1][1 - \varepsilon_2]e^{-2K\varepsilon_1} = \kappa < 1.
\end{aligned}$$

Elements of proof:

- For any $j \in \mathbb{N}^*$,

$$\mathbb{E}^u_{x_0}\left[e^{-\alpha T_{2j+1}}\right] = \mathbb{E}^u_{x_0}\left[e^{-\alpha T_{2j-1}} \mathbb{E}^u_{x_0}\left[e^{-\alpha(\Theta_{2j}+\Theta_{2j+1})} \middle| H_{2j-1}\right]\right]$$
$$\leq \kappa \mathbb{E}^u_{x_0}\left[e^{-\alpha T_{2j-1}}\right],$$

  and so

$$\mathbb{E}^u_{x_0}\left[e^{-\alpha T_{2j+1}}\right] \leq \kappa^j \mathbb{E}^u_{x_0}\left[e^{-\alpha T_1}\right] \leq \kappa^j.$$

  Similarly,

$$\mathbb{E}^u_{x_0}\left[e^{-\alpha T_{2j+2}}\right] \leq \kappa^j \mathbb{E}^u_{x_0}\left[e^{-\alpha T_2}\right] \leq \kappa^j.$$

  for any $j \in \mathbb{N}$.
- Therefore,

$$\mathbb{E}^u_{x_0}\Big[\sum_{n\in\mathbb{N}^*} e^{-\alpha T_n}\Big] \leq \frac{2}{1-\kappa}.$$

Outline

There are two approaches to deal with such problems:

- the associated discrete-stage Markov decision model:
  - A. Almudevar. A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes, 2001.
  - N. Bauerle and U. Rieder. Optimal control of piecewise deterministic Markov processes with finite time horizon, 2010.
  - O.L.V Costa and F. Dufour. Continuous average control of piecewise deterministic Markov processes, 2013.
  - M.H.A. Davis. Control of piecewise-deterministic processes via discrete-time dynamic programming, 1986.
  - L. Forwick, M. Schal, and M. Schmitz. Piecewise deterministic Markov control processes with feedback controls and unbounded costs, 2004.
  - M. Schal. On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance, 1998.
  - A.A. Yushkevich. On reducing a jump controllable Markov model to a model with discrete time, 1980.

There are two approaches to deal with such problems:

- the the infinitesimal approach (HJB equation):
  - M.H.A. Davis. Markov models and optimization, volume 49 of Monographs on Statistics and Applied Probability, 1993.
  - M.A.H. Dempster and J.J. Ye. Necessary and sufficient optimality conditions for control of piecewise deterministic processes, 1992.
  - M.A.H. Dempster and J.J. Ye. Generalized Bellman-Hamilton-Jacob optimality conditions for a control problem with boundary conditions, 1996.
  - A.A. Yushkevich. Bellman inequalities in Markov decision deterministic drift processes. Stochastics, 1987

Notation and preliminary results:

- $\mathbb{A}(\overline{\mathbf{X}})$ is the set of functions $g \in \mathbb{B}(\overline{\mathbf{X}})$ such that for any $x \in \overline{\mathbf{X}}$, the function $g(\phi(x,\cdot))$ is absolutely continuous on $[0, t^*(x)] \cap \mathbb{R}_+$.
- Let $g \in \mathbb{A}(\overline{\mathbf{X}})$, there exists a real-valued measurable function $\mathcal{X}g$ defined on $\mathbf{X}$ satisfying for any $t \in [0, t^*(x)[$

$$g(\phi(x,t)) = g(x) + \int_{[0,t]} \mathcal{X}g(\phi(x,s))ds.$$

- Let $R \in \mathcal{P}(X|Y)$. Then $Rf(y) \doteq \int_X f(x)R(dx|y)$ for any $y \in Y$ and measurable function f. For any measure η on $(Y, \mathcal{B}(Y))$, $\eta R(\cdot) \doteq \int_Y R(\cdot|y)\eta(dy)$.
- $q(dy|x,a) \doteq \lambda(x,a)\big[Q(dy|x,a) - \delta_x(dy)\big]$

Sufficient conditions for the existence of a solution for the HJB equation associated with the optimization problem.

Theorem

Suppose assumptions A, B and C hold. Then there exist $W \in \mathbb{A}(\overline{\mathbf{X}})$ *and* $\mathcal{X}W \in \mathbb{B}(\mathbf{X})$ *satisfying*

$$-\alpha W(x) + \mathcal{X}W(x) + \inf_{a \in A^g(x)} \Big\{ C_0^g(x,a) + qW(x,a) \Big\} = 0,$$

for any $x \in \mathbf{X}$*, and*

$$W(z) = \inf_{b \in A^i(z)} \Big\{ C_0^i(z,b) + QW(z,b) \Big\},$$

for any $z \in \Delta$*. Moreover, for any* $x \in \mathbf{X}$

$$W(x) = \inf_{u \in \mathcal{U}} \mathcal{V}_0(u,x).$$

Sufficient conditions for the existence of an optimal strategy.

Theorem

Suppose assumptions A, B and C hold. There exists a measurable mapping $\widehat{\varphi} : \overline{\mathbf{X}} \to \mathbf{A}$ *such that* $\widehat{\varphi}(y) \in \mathbf{A}(y)$ *for any* $y \in \overline{\mathbf{X}}$ *and satisfying*

$$C_0^g(x, \widehat{\varphi}(x)) + qW(x, \widehat{\varphi}(x)) = \inf_{a \in \mathbf{A}(x)} \Big\{ C_0^g(x, a) + qW(x, a) \Big\}$$

for any $x \in \mathbf{X}$*, and*

$$C_0^i(z, \widehat{\varphi}(z)) + QW(z, \widehat{\varphi}(z)) = \inf_{b \in \mathbf{A}(z)} \Big\{ C_0^i(z, b) + QW(z, b) \Big\}.$$

for any $z \in \Delta$*. Moreover, the stationary non-randomized strategy* $\widehat{\varphi}$ *is optimal.*

Elements of proof:

- Define recursively $\{W_i\}_{i\in\mathbb{N}}$ as

$$W_{i+1}(y) = \mathfrak{B} W_i(y),$$

with $W_0(y) = -K_A I_{A_{\varepsilon_1}}(y) - (K_A + K_B) I_{A^c_{\varepsilon_1}}(y)$ and

$$\mathfrak{B} V(y) = \int_{[0,t^*(y)[} e^{-(K+\alpha)t} \mathfrak{R} V(\phi(y,t))dt + e^{-(K+\alpha)t^*(y)} \mathfrak{T} V(\phi(y,t^*(y))),$$

where

$$\mathfrak{R} V(x) = \inf_{a\in\mathbf{A}(x)} \Big\{ C_0^g(x,a) + qV(x,a) + KV(x) \Big\},$$

and

$$\mathfrak{T} V(z) = \inf_{b\in\mathbf{A}(z)} \Big\{ C_0^i(z,b) + QV(z,b) \Big\}.$$

- W_i is lower semicontinuous and

$$\left|W_i(y)\right| \leq K_A I_{A_{\varepsilon_1}}(y) + (K_A + K_B) I_{A^c_{\varepsilon_1}}(y).$$

- $\mathfrak{B}$ is monotone $\left(V_1 \leq V_2 \Rightarrow \mathfrak{B}V_1 \leq \mathfrak{B}V_2\right)$, $\left\{W_i\right\}_{i\in\mathbb{N}}$ is increasing and $W_i \to W$ and W is bounded and lower semicontinuous.
- $\lim_{i\to\infty} \mathfrak{R}W_i(x) = \mathfrak{R}W(x)$, for any $x \in \mathbf{X}$
  $\lim_{i\to\infty} \mathfrak{T}W_i(z) = \mathfrak{T}W(z)$ for any $z \in \Delta$.

- By using the bounded convergence Theorem,

$$\begin{aligned} W(y) &= \mathfrak{B}W(y) \\ &= \int_{[0,t^*(y)[} e^{-(K+\alpha)t} \mathfrak{R}W(\phi(y,t))dt \\ &\quad + e^{-(K+\alpha)t^*(y)} \mathfrak{T}W(\phi(y,t^*(y))), \end{aligned}$$

where $y \in \overline{\mathbf{X}}$.

- Then $W \in \mathbb{A}(\overline{\mathbf{X}})$ and there exists $\mathcal{X}W \in \mathbb{B}(\mathbf{X})$

$$-\alpha W(x) + \mathcal{X}W(x) + \inf_{a \in A^g(x)} \Big\{ C_0^g(x,a) + qW(x,a) \Big\} = 0,$$

for any $x \in \mathbf{X}$, and

$$W(z) = \inf_{b \in A^i(z)} \Big\{ C_0^i(z,b) + QW(z,b) \Big\},$$

for any $z \in \Delta$.

Outline

The linear programming approach

The method has been extensively studied in the literature

- Continuous and discrete time MDP:
  - Eitan Altman. Constrained Markov decision processes, 1999.
  - Vivek S. Borkar. A Convex Analytic Approach to Markov Decision Processes, 1988.
  - Vivek S. Borkar. Convex analytic methods in Markov decision processes, 2002.
  - Alexey B. Piunovskiy. Optimal control of random sequences in problems with constraints, 1997.
- Controlled martingale problems:
  - Abhay G. Bhatt and Vivek S. Borkar. Occupation measures for controlled Markov processes: characterization and optimality, 1996.
  - K. Helmes and R. H. Stockbridge. Linear programming approach to the optimal stopping of singular stochastic processes, 2007.
  - Richard H. Stockbridge. Time-average control of martingale problems: a linear programming formulation, 1990.

Occupation measure

For any admissible control strategy $u \in \mathcal{U}$, the occupation measure $\eta_u \in \mathcal{M}(\mathbf{K})$ associated with u is defined as follows

$$\begin{aligned}\eta_u(\Gamma) =& \mathbb{E}^u_{x_0}\Big[\int_{\Gamma \cap \mathbf{K}^g} \int_{]0,\infty[} e^{-\alpha s} \delta_{\xi_s}(dx)\pi(da|s)ds\Big] \\ &+ \mathbb{E}^u_{x_0}\Big[\int_{\Gamma \cap \mathbf{K}^i} \sum_{n \in \mathbb{N}^*} e^{-\alpha T_n} \delta_{\xi_{T_n-}}(dz)\gamma(db|T_n-)\Big].\end{aligned}$$

for any $\Gamma \in \mathcal{B}(\mathbf{K})$.

Linear programming approach

The infinite-horizon discounted criteria can be rewritten as

$$\mathcal{V}_j(u, x_0) = \mathbb{E}^u_{x_0}\left[\int_{]0,+\infty[} e^{-\alpha s} \int_{\mathbf{A}(\xi_s)} C^g_j(\xi_s, a)\pi(da|s)ds\right]$$

$$+ \mathbb{E}^u_{x_0}\left[\int_{]0,+\infty[} e^{-\alpha s} I_{\{\xi_{s-}\in\Delta\}} \int_{\mathbf{A}(\xi_{s-})} C^i_j(\xi_{s-}, a)\gamma(da|s)\mu(ds, \mathbf{X})\right]$$

$$= \eta^g_u(C^g_j) + \eta^i_u(C^i_j)$$

where η^g_u (resp. η^i_u) denotes the restriction of η_u to $\mathbf{K}^g$ (resp. $\mathbf{K}^i$).

Admissible measure

A finite measure $\eta \in \mathcal{M}(\mathbf{K})$ is called admissible if, for any $(W, \mathcal{X}W) \in \mathbb{A}(\overline{\mathbf{X}}) \times \mathbb{B}(\mathbf{X})$, the following equality holds

$$\int_{\mathbf{X}} [\alpha W(x) - \mathcal{X}W(x)] \widehat{\eta}^g(dx) + \int_{\Delta} W(z) \widehat{\eta}^i(dz)$$
$$= W(x_0) + \int_{\mathbf{K}^g} qW(x,a) \eta^g(dx,da) + \int_{\mathbf{K}^i} QW(z,b) \eta^i(dz,db).$$

with $\widehat{\eta}^g$ (resp. $\widehat{\eta}^i$) denotes the marginal of η^g (resp. η^i) w.r.t. to $\mathbf{X}$.

Occupation and admissible measures

The next important result shows the link between the set of admissible measures and the set of occupation measures.

Theorem

Suppose Assumption A is satisfied. Then the following assertions hold.

i) *For any control strategy $u \in \mathcal{U}$, the occupation measure η_u is admissible.*

ii) *Suppose that the measure η is admissible. Then there exist stochastic kernels $\pi \in \mathcal{P}^g$ and $\gamma \in \mathcal{P}^i$ for which the stationary control strategy $u = (\pi, \gamma) \in \mathcal{U}_s$ satisfies $\eta = \eta_u$.*

Linear programming approach

The constrained linear program, labeled $\mathbb{LP}$, is defined as

$$\inf_{(\eta^g,\eta^i)\in\mathbb{M}} \eta^g(C_0^g) + \eta^i(C_0^i)$$

where $\mathbb{M}$ is the set of measures (η^g, η^i) in $\mathcal{M}(\mathbf{K}^i) \times \mathcal{M}(\mathbf{K}^g)$ such that $\eta^g + \eta^i$ is admissible and satisfies

$$\eta^g(C_j^g) + \eta^i(C_j^i) \leq B_j.$$

Linear programming approach

Theorem

Suppose Assumption A holds and the cost functions C_j^g and C_j^i are bounded from below for any $j \in \mathbb{N}_p$. Then the values of the constrained control problem and the linear program $\mathbb{LP}$ are equivalent:

$$\inf_{(\eta^g,\eta^i)\in\mathbb{M}} \eta^g(C_0^g) + \eta^i(C_0^i) = \inf_{u\in\mathcal{U}^f} \mathcal{V}_0(u, x_0).$$

Linear programming approach

Theorem

Suppose Assumptions A, B and (C1) are satisfied. Assume the cost functions C_j^g (resp. C_j^i) are bounded from below and lower semicontinuous on $\mathbf{K}^g$ (resp. $\mathbf{K}^i$) for any $j \in \mathbb{N}_p$.

If the set of feasible strategies is non empty then the $\mathbb{LP}$ is solvable and there exists a stationary feasible strategy u^* satisfying

$$\begin{aligned}\eta^g_{u^*}(C_0^g) + \eta^i_{u^*}(C_0^i) &= \inf_{(\eta^g,\eta^i)\in\mathbb{M}} \eta^g(C_0^g) + \eta^i(C_0^i) \\ &= \inf_{u\in\mathcal{U}^f} \mathcal{V}_0(u, x_0) = \mathcal{V}_0(u^*, x_0).\end{aligned}$$