

# Chaînes de Markov contrôlées

## 1 Chaînes contrôlées

On s'intéresse à des évolutions aléatoires sur un espace  $\mathcal{M}$  en temps discret et que on peut modifier par la choix, à chaque pas de temps, d'une action  $a$  dans un ensemble préfixé d'action admissibles  $\mathcal{A}$ . Donn  un  tat  $x \in \mathcal{M}$  au temps initial  $k \in \mathbb{N}$  et une politique de choix des actions de contr le  $u$  on peut consid rer la succession al atoire  $X_k, \dots, X_n, \dots$  des  tats visit  par notre syst me sujet   la politique  $u$ . Le probl me que on va se poser est l'optimisation d'un quelque crit re moyenne par la choix d'une politique de contr le (tout simplement un contr le).

Soit donc  $\mathcal{M}$  une espace d' tats d nombrable,  $\mathcal{A}$  l'espace des action admissibles et  $\Pi(\mathcal{M})$  l'espace des mesures de probabilit  sur  $\mathcal{M}$ . On consid re une fonction  $P: \mathbb{N} \times \mathcal{A} \times \mathcal{M} \rightarrow \Pi(\mathcal{M})$  qui donn  un temps  $n \in \mathbb{N}$ , une action  $a \in \mathcal{A}$  et un  tat  $x \in \mathcal{M}$  d termine la probabilit   $P_{n,a}(x, y) = P_{n,a}(x)(\{y\})$  que l' tat   l' tape suivante soit  $y \in \mathcal{M}$ . La fonction  $P$  sp cifie la dynamique al atoire de notre syst me. Soit

$$\mathcal{M}_k^* = \{(n, x_k, \dots, x_n) : n \in \mathbb{N}, k \leq n, x_k, \dots, x_n \in \mathcal{M}\}$$

on appelle une politique de contr le (ou simplement un contr le) une fonction  $u: \mathcal{M}_k^* \rightarrow \mathcal{A}$  et on appelle  $\mathcal{C}_k$  l'espace des controls relatives   l'instant initiale  $k \geq 0$ . La politique de contr le  $u \in \mathcal{C}_k$  est donc une r gle qui au temps  $n$ , aient observ  la succession d' tats  $(x_k, \dots, x_n)$ , d termine l'action  $u_n(x_k, \dots, x_n) \in \mathcal{A}$  pour modifier l' volution future de notre syst me al atoire. Un contr le Markovien et stationnaire est un contr le  $u \in \mathcal{C}_k$  qui d pend seulement de l' tat actuel du syst me, c- -d tel que  $u_n(x_k, \dots, x_n) = \varphi(x_n)$  pour une quelque fonction  $\varphi: \mathcal{M} \rightarrow \mathcal{A}$ .

Soit  $\Omega$  l'espace canonique  $\Omega = \mathcal{M}^{\mathbb{N}}$  avec la tribu produit  $\mathcal{F}$  (sur chaque composante on consid re la tribu discr te des toutes les parties de  $\mathcal{M}$ ). Soit  $X_n(\omega) = \omega_n$  la projection sur la  $n$ -eme composante de  $\omega \in \Omega$ .

Donn  un temps initial  $k \in \mathbb{N}$ , un  tat initial  $x \in \mathcal{M}$  et un contr le  $u \in \mathcal{C}_k$  on consid re la probabilit   $\mathbb{P}_{(k,x)}^u$  telle que

$$\mathbb{P}_{(k,x)}^u(X_k = x_k, \dots, X_{n+1} = x_{n+1}) = \delta_{x,x_k} \prod_{i=k}^n P_{k,u_i(x_k, \dots, x_i)}(x_i, x_{i+1}) \quad \forall n \geq k. \quad (1)$$

On appelle le processus  $(X_n)_{n \geq k}$  un processus contr l .

**Lemme 1.** *On a que  $\mathbb{P}_{(k,x)}^u$  v rifie (1) ssi,  $\forall n \geq k$  on a*

$$\mathbb{P}_{(k,x)}^u(X_{n+1} = x_{n+1} | X_k = x_k, \dots, X_n = x_n) = P_{n,u_n(x_k, \dots, x_n)}(x_n, x_{n+1}).$$

**D monstration.** Exercice. □

Une fa on canonique de construire un processus contr l  est de consid rer une fonction

$$G: \mathbb{N} \times \mathcal{M} \times \mathcal{A} \times E \rightarrow \mathcal{M}$$

et un espace de probabilit   $(\Omega, \mathcal{F}, \mathbb{P})$  muni d'une suite de v.a. iid  $(Z_n)_{n \geq k}$    valeurs dans l'espace auxiliaire  $E$ . On pose alors

$$X_k = x, \quad X_{n+1} = G(n, X_n, U_n, Z_{n+1}), \quad n \geq k \quad (2)$$

où  $U_n = u_n(X_k, \dots, X_n)$ . Une suite aléatoire construite de cette façon est appelé une récurrence aléatoire contrôlée ou un système dynamique aléatoire contrôlé. Il est facile de montrer (exercice) que la suite  $(X_n)_{n \geq k}$  vérifie

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_k = x_k, \dots, X_n = x_n) = P_{n, u_n(x_k, \dots, x_n)}(x_n, x_{n+1}) \quad (3)$$

où

$$P_{n, a}(x, y) = \mathbb{P}(G(n, x, a, Z_1) = y). \quad (4)$$

Réciproquement, pour tout fonction  $P: \mathbb{N} \times \mathcal{A} \times \mathcal{M} \rightarrow \Pi(\mathcal{M})$  il est possible de trouver un espace auxiliaire  $E$ , une suite iid  $(Z_n)_{n \geq k}$  et une fonction  $G$  tels que les équations (3) et (4) soient satisfaites (exercice. sugg. prendre  $E = [0, 1]$  et les  $Z_n \sim \mathcal{U}([0, 1])$  et construire une  $G$  approprié). La correspondance entre processus contrôlés et systèmes dynamiques contrôlés n'est pas univoque (plusieurs systèmes dynamiques contrôlés différents peuvent avoir la même loi et donc correspondre au même processus contrôlé).

Donnée une fonction  $F: \mathbb{N} \times \mathcal{M} \rightarrow \mathbb{R}$  on définit la fonction  $PF: \mathbb{N} \times \mathcal{M} \times \mathcal{A} \rightarrow \mathbb{R}$  par

$$(PF)(n, x, a) = \sum_{y \in \mathcal{M}} P_{n, a}(x, y) F(n+1, y)$$

Dans le cas d'un système dynamique contrôlé on a que

$$PF(n, x, a) = \mathbb{E}[F(n+1, G(n, x, a, Z_1))].$$

On remarque que dans cette égalité le membre de droite est bien défini même dans le cas d'espace d'états non discret.

## 2 Principe d'optimalité

Soit  $c: \mathbb{N} \times \mathcal{M} \times \mathcal{A} \rightarrow \mathbb{R}$  une fonction que représente un coût pour unité de temps et qui peut dépendre du temps, de l'état actuel du système et de l'action choisi pour continuer. Donnée un état initiale  $(k, x) \in \mathbb{N} \times \mathcal{M}$  et un contrôle  $u \in \mathcal{C}_k$ , le coût totale moyen pour le processus contrôlé  $(X_n)_{n \geq k}$  est déterminé par

$$V^u(k, x) = \mathbb{E}_{(k, x)}^u \sum_{n \geq k} c(n, X_n, U_n)$$

où l'espérance  $\mathbb{E}_{(k, x)}^u$  est par rapport à la probabilité  $\mathbb{P}_{(k, x)}^u$ . Pour que cet expression ait un sens on admet que une des condition suivantes est satisfaite:  $c(n, x, a) \geq 0$ ,  $c(n, x, a) \leq 0$  ou  $\sum_{n \geq k} \sup_{x, a} |c(n, x, a)| < +\infty$ . On veut trouver un contrôle  $u^*$  qui minimise ce coût moyen parmi tout les controls admissibles:

$$V^{u^*}(k, x) = V(k, x) = \inf_{u \in \mathcal{C}_k} V^u(k, x).$$

On appelle la valeur optimale  $V(k, x)$  du coût moyen obtenu à partir de l'état  $(k, x)$  est appelé la fonction valeur  $V: \mathbb{N} \times \mathcal{M} \rightarrow \mathbb{R}$ .

**Théorème 2.** *La fonction valeur satisfait l'équation (dit de Bellman ou de la programmation dynamique)*

$$V(k, x) = \inf_{a \in \mathcal{A}} [c(k, x, a) + (PV)(k, x, a)]$$

**Démonstration.** Sans perte de généralité on peut supposer que le processus contrôlé  $(X_n)_{n \geq k}$  est un système dynamique contrôlée associé à la fonction  $G$  et à la suite aléatoire  $(Z_n)_{n \geq k}$ . Donc

$$V^u(k, x) = \mathbb{E} \sum_{n \geq k} c(n, X_n, U_n) = c(k, x, u_k(x)) + \mathbb{E} \sum_{n \geq k+1} c(n, X_n, U_n)$$

où  $X_k = x$ ,  $X_{n+1} = G(n, X_n, U_n, Z_{n+1})$  et  $U_n = u_n(X_k, \dots, X_n)$ . Soit  $a = u_k(x)$  et  $\tilde{u} \in \mathcal{C}_{k+1}$  le contrôle définit par  $\tilde{u}_n(x_{k+1}, \dots, x_n) = u_n(x, x_{k+1}, \dots, x_n)$ . On a que  $U_n = \tilde{u}_n(X_{k+1}, \dots, X_n) = \tilde{U}_n$ . Si on pose

$$\tilde{X}_{k+1} = y, \quad \tilde{X}_{n+1} = G(n, \tilde{X}_n, \tilde{U}_n, Z_{n+1}) \quad \forall n > k+1$$

on aura que  $\mathbb{P}(\tilde{X}_n = X_n, \forall n \geq k+1 | X_{k+1} = y) = 1$  et donc

$$\begin{aligned} \mathbb{E} \sum_{n \geq k+1} c(n, X_n, U_n) &= \sum_{y \in \mathcal{M}} \mathbb{P}(X_{k+1} = y) \mathbb{E} \left[ \sum_{n \geq k+1} c(n, X_n, U_n) | X_{k+1} = y \right] \\ &= \sum_{y \in \mathcal{M}} \underbrace{\mathbb{P}(X_{k+1} = y)}_{P_{n,a}(x,y)} \mathbb{E} \underbrace{\sum_{n \geq k+1} c(n, \tilde{X}_n, \tilde{U}_n)}_{V^{\tilde{u}}(k+1,y)} \\ &= \sum_{y \in \mathcal{M}} P_{n,a}(x, y) V^{\tilde{u}}(k+1, y) = PV(k, x, a). \end{aligned}$$

□

**Remarque 3.** On peut vouloir résoudre un problème de maximisation au lieu d'un problème de minimisation. Dans ce cas la fonction valeur est définie par  $V(k, x) = \sup_{u \in \mathcal{C}_k} V^u(k, x)$  et l'équation de Bellman prends la forme

$$V(k, x) = \sup_{a \in \mathcal{A}} [c(n, x, a) + PV(n, x, a)].$$

On dit que un processus contrôlé est homogène si la fonction  $P$  ne dépend pas du temps, i.e. si  $P: \mathcal{A} \times \mathcal{M} \rightarrow \Pi(\mathcal{M})$ . Similairement on dit que un système dynamique contrôlé est homogène si la fonction  $G$  ne dépend pas du temps:  $G: \mathcal{M} \times \mathcal{A} \times E \rightarrow \mathcal{M}$ . Un processus contrôlé est homogène ssi il est équivalent à un système dynamique homogène.

**Lemme 4.** Si la fonction de coût et la fonction  $G$  ne dépendants pas du temps, c-à-d si

$$V^u(k, x) = \mathbb{E} \sum_{n \geq k} c(X_n, U_n)$$

et

$$X_k = x, \quad X_{n+1} = G(X_n, U_n, Z_{n+1}), \quad n \geq k$$

alors la fonction valeur  $V(k, x)$  ne dépend pas du temps et l'équation de Bellman devient

$$V(x) = \inf_{a \in \mathcal{A}} [c(x, a) + PV(x, a)]. \quad (5)$$

**Démonstration.** On considère  $u \in \mathcal{C}_{k+1}$  et

$$V^u(k+1, x) = \mathbb{E} \sum_{n \geq k+1} c(X_n, U_n) = \mathbb{E} \sum_{n \geq k} c(X_{n+1}, U_{n+1})$$

où  $X_{k+1} = x$ ,  $X_{n+1} = G(X_n, U_n, Z_{n+1})$ ,  $n \geq k+1$ ,  $U_n = u_n(X_{k+1}, \dots, X_n)$ . Soit maintenant  $\tilde{X}_n = X_{n+1}$ . On a que  $\tilde{X}_k = x$  et pour  $n \geq k$   $\tilde{X}_{n+1} = G(\tilde{X}_n, u_{n+1}(\tilde{X}_k, \dots, \tilde{X}_n), Z_{n+2})$ . Soit alors  $\tilde{u} \in \mathcal{C}_k$  tel que  $\tilde{u}_n(x_k, \dots, x_n) = u_{n+1}(x_k, \dots, x_n)$  et  $\tilde{U}_n = \tilde{u}_n(\tilde{X}_k, \dots, \tilde{X}_n) = U_{n+1}$ . Le processus  $(\tilde{X}_n)_{n \geq k}$  est le processus contrôlé associé au système dynamique  $(G, (Z_{n+1})_{n \geq 1})$  avec contrôle  $\tilde{u}$  et état initiale  $(k, x)$ , donc

$$\mathbb{E}_{(k+1,x)}^u \sum_{n \geq k} c(X_{n+1}, U_{n+1}) = \mathbb{E} \sum_{n \geq k} c(\tilde{X}_n, \tilde{U}_n) = \mathbb{E}_{(k,x)}^{\tilde{u}} \sum_{n \geq k} c(X_n, U_n) = V^{\tilde{u}}(k, x)$$

et donc  $V(k, x) = V(k+1, x)$  pour tout  $k \geq 0$ . Soit  $V(x) = V(0, x)$ , l'équation de Bellman est

$$V(x) = V(0, x) = \inf_{a \in \mathcal{A}} \{c(x, a) + \mathbb{E}[V(1, G(x, a, Z_1))]\} = \inf_{a \in \mathcal{A}} \{c(x, a) + \mathbb{E}[V(G(x, a, Z_1))]\}$$

ce qui donne l'eq. (5). □

### 3 Contrôle en horizon fini

L'équation de Bellman est un outil puissant pour caractériser (et des fois déterminer) les politiques optimaux dans les problèmes de contrôle. Le cas plus simple est l'optimisation en horizon fini qui on va analyser dans cette section. Soit  $N \geq 0$  un temps et soit  $r: \mathbb{N} \times \mathcal{M} \times \mathcal{A} \rightarrow \mathbb{R}$  une fonction de gain pour laquelle on fait l'hypothèse que  $r(n, x, a) = 0$  pour tout  $n > N$  et que  $r(N, x, a) = C(x)$  pour une fonction  $C: \mathcal{M} \rightarrow \mathbb{R}$ . Le gain moyen associé au processus contrôlé par  $u$  et démarré en  $(k, x)$  est

$$V^u(k, x) = \mathbb{E}_{(k, x)}^u \left[ \sum_{n=k}^{N-1} r(n, X_n, U_n) + C(X_N) \right].$$

On veut maximiser cette quantité en fonction de la politique  $u$ . La fonction valeur  $V(k, x) = \sup_u V^u(k, x)$  satisfait l'équation de Bellman

$$V(n, x) = \sup_a \left[ r(n, x, a) + \sum_y P_{n, a}(x, y) V(n+1, y) \right]$$

pour tout  $k \leq n < N$  et en plus on a la condition au bord  $V(N, x) = C(x)$ . Par récurrence rétrograde on peut alors trouver  $V(N-1, \cdot)$ ,  $V(N-2, \cdot)$  et ainsi de suite jusqu'à déterminer  $V(k, x)$  au temps initiale. L'équation de Bellman a donc une seule solution.

**Théorème 5.** *Supposons que  $u$  est un contrôle Markovien tel que*

$$V(k, x) = (c + PV)(k, x, u_k(x)) \quad 0 \leq k \leq N-1, x \in \mathcal{M}$$

*alors  $u$  est optimale pour tout  $(k, x) \in \mathbb{N} \times \mathcal{M}$ , i.e.  $V(k, x) = V^u(k, x)$ .*

**Démonstration.** Considérons un tel contrôle et soit  $(X_n)_{n \geq k}$  le processus contrôlé associé. Soit

$$M_n = \sum_{j=k}^{n-1} r(j, X_j, U_j) + V(n, X_n) \quad k \leq n \leq N$$

Alors pour tout  $k \leq n \leq N-1$  on a

$$M_{n+1} - M_n = V(n+1, X_{n+1}) - V(n, X_n) - r(n, X_n, U_n)$$

et donc

$$\begin{aligned} \mathbb{E}_{(k, x)}^u [M_{n+1} - M_n | X_n = y] &= \mathbb{E}_{(k, x)}^u [V(n+1, X_{n+1}) - V(n, X_n) - r(n, X_n, U_n) | X_n = y] \\ &= (r + PV)(n, y, u_n(y)) - V(n, y) = 0 \end{aligned}$$

ce qui donne que  $\mathbb{E}_{(k, x)}^u [M_n] = \mathbb{E}_{(k, x)}^u [M_{n+1}]$  pour tout  $k \leq n < N$ . Par conséquent

$$V(k, x) = \mathbb{E}_{(k, x)}^u [M_k] = \mathbb{E}_{(k, x)}^u [M_N] = \mathbb{E}_{(k, x)}^u \left[ \sum_{j=k}^{N-1} r(j, X_j, U_j) + C(X_N) \right] = V^u(k, x).$$

□

**Exemple 6.** (EXERCER UNE OPTION D'ACHAT) On a la possibilité d'acheter un actif à un prix fixé d'avance  $p$  et à un instant quelconque  $n = 0, \dots, N-1$ . Le prix de marché de l'actif est modélisé par une suite  $(Y_n)_{n \geq 0}$  donnée par  $Y_{n+1} = Y_n + \varepsilon_{n+1}$  où  $(\varepsilon_n)_{n \geq 1}$  est une suite iid intégrable. L'objectif est de maximiser le gain moyen relatif à l'utilisation de l'option d'achat: si on décide de l'utiliser au temps  $n$  avec un prix de marché  $Y_n$  alors notre gain serait de  $Y_n - p$ .

Le processus contrôlé est donnée par la suite des valeurs de notre option et on prend comme espace d'états l'ensemble  $\mathcal{M} = \mathbb{R} \cup \{\Delta\}$  car à un instant déterminé soit on possède encore l'option et sa valeur est  $x \in \mathbb{R}$ , soit on a déjà exercé l'option et alors on décide de façon conventionnelle de être dans l'état fictif  $\Delta$ . L'espace des actions est  $\mathcal{A} = \{0, 1\}$ , 0 si on exerce pas et 1 si on décide d'exercer l'option. On n'est pas dans le cas d'espace d'états discret mais on peut réaliser la dynamique contrôlée comme dynamique aléatoire contrôlée. La fonction de gain est donnée par  $r(n, x, a) = a(x - p)$  et la dynamique aléatoire par

$$G(x, a, \varepsilon) = \begin{cases} x + \varepsilon & \text{si } x \in \mathbb{R}, a = 0 \\ \Delta & \text{si } x \in \mathbb{R}, a = 1 \\ \Delta & \text{si } x = \Delta \end{cases}$$

avec espace auxiliaire  $\mathbb{R}$  et suite iid  $(\varepsilon_n)_{n \geq 0}$ . En particulier la fonction de transition  $P$  est de la forme

$$P_{n,0}(x, A) = \mathbb{P}(x + \varepsilon_1 \in A), \quad P_{n,1}(x, \mathbb{R}) = 0, \quad P_{n,1}(x, \{\Delta\}) = 1$$

(sur  $\mathcal{M}$  on considère la tribu  $\sigma(\mathcal{B}(\mathbb{R}), \{\Delta\})$ ) et on a

$$PF(n, x, a) = \begin{cases} \mathbb{E}[F(n+1, x + \varepsilon)] & \text{si } a = 1 \\ F(n+1, \Delta) & \text{si } a = 0 \end{cases}$$

L'équation de Bellman est alors donnée par

$$V(k, x) = \max \{x - p, \mathbb{E}[V(k+1, x + \varepsilon)]\}, \quad 0 \leq k \leq N-1, x \in \mathbb{R}$$

et  $V(N, x) = 0$  (car à  $N$  on ne peut pas exercer l'option). On note que  $V(N-1, x) = (x - p)_+$ .

Montrez que  $V(k, x)$  est une fonction convexe de  $x$  et que  $V(k, x) \geq V(k+1, x)$  pour tout  $0 \leq k \leq N$  et tout  $x \in \mathbb{R}$ .

Soit  $p_k = \inf \{x \geq 0 : V(k, x) = x - p\}$ . Montrez que  $p_k$  est décroissant en  $k$  et que la politique optimale est d'exercer l'option de que  $Y_k \geq p_k$ .

## 4 Contrôle en horizon infini: cas des gains positifs

On se donne un processus contrôlé homogène et une fonction gain homogène et positive  $r: \mathcal{M} \times \mathcal{A} \rightarrow \mathbb{R}_+$ . Si  $u \in \mathcal{C}_0$  on définit le gain total moyen

$$V^u(x) = \mathbb{E}_{(0,x)}^u \sum_{m \geq 0} r(X_m, U_m)$$

et la fonction valeur du problème de maximisation de ce gain  $V(x) = \sup_{u \in \mathcal{C}_0} V^u(x)$ . Pour tout  $n \geq 0$  soit

$$V_n^u(x) = \mathbb{E}_{(0,x)}^u \sum_{0 \leq m \leq n-1} r(X_m, U_m), \quad V_n(x) = \sup_{u \in \mathcal{C}_0} V_n^u(x).$$

Par convergence monotone  $V_n^u(x) \nearrow V^u(x)$  et donc

$$\sup_n V_n(x) = \sup_n \sup_{u \in \mathcal{C}_0} V_n^u(x) = \sup_{u \in \mathcal{C}_0} \sup_n V_n^u(x) = \sup_{u \in \mathcal{C}_0} V^u(x) = V(x).$$

Les fonction  $V_n(x)$  peuvent être calculé par récurrence.

**Lemme 7.** *On a l'équation*

$$V_{n+1}(x) = \sup_{a \in \mathcal{A}} [r(x, a) + PV_n(x, a)].$$

**Démonstration.** (Exercice, utiliser l'homogénéité). □

**Théorème 8.** *La fonction valeur en horizon infini  $V$  est la plus petite solution non-négative de l'équation*

$$V(x) = \sup_{a \in \mathcal{A}} [r(x, a) + PV(x, a)], \quad x \in \mathcal{M}. \quad (6)$$

Tout contrôle  $u \in \mathcal{C}_0$  tel que  $V^u$  satisfait cette équation est optimal, pour tout état initial  $x \in \mathcal{M}$ .

**Démonstration.** Par le principe d'optimalité on sait que  $V$  satisfait l'équation. Soit maintenant  $F: \mathcal{M} \rightarrow \mathbb{R}_+$  un autre solution non-négative de (6). Alors  $F(x) \geq 0 = V_0(x)$ . Supposons par induction que  $F \geq V_n$ , alors

$$F(x) = \sup_{a \in \mathcal{A}} [r(x, a) + PF(x, a)] \geq \sup_{a \in \mathcal{A}} [r(x, a) + PV_n(x, a)] = V_{n+1}(x)$$

et donc  $F \geq V_n$  pour tout  $n \geq 0$  ce qui implique que  $V = \sup_n V_n \leq F$ . □

## 5 Contrôle en horizon infini: cas des coûts actualisés

Ici on considère un processus contrôlé homogène, une fonction coût homogène  $c: \mathcal{M} \times \mathcal{A} \rightarrow \mathbb{R}$  (non nécessairement positive) et bornée  $|c(x, a)| \leq C < \infty$  et une constante  $\beta \in ]0, 1]$ . Si  $u \in \mathcal{C}_0$  on définit le coût total moyen actualisé

$$V^u(x) = \mathbb{E}_{(0,x)}^u \sum_{m \geq 0} \beta^m c(X_m, U_m)$$

et le coût total moyen actualisé minimale  $V(x) = \inf_{u \in \mathcal{C}_0} V^u(x)$ . Pour tout  $n \geq 0$  on définit aussi les coût partiels

$$V_n^u(x) = \mathbb{E}_{(0,x)}^u \sum_{m=0}^{n-1} \beta^m c(X_m, U_m) \quad V_n(x) = \inf_{u \in \mathcal{C}_0} V_n^u(x).$$

On remarque que

$$|V_n^u(x) - V(x)| \leq \sum_{m \geq n} \beta^m C = C \frac{\beta^n}{1 - \beta} \rightarrow 0$$

si  $n \rightarrow \infty$ . De même

$$|V_n(x) - V(x)| \leq C \frac{\beta^n}{1 - \beta} \rightarrow 0.$$

**Lemme 9.** *On a l'équation*

$$V_{n+1}(x) = \inf_{a \in \mathcal{A}} [c(x, a) + \beta P V_n(x, a)] \quad n \geq 0, x \in \mathcal{M}$$

**Démonstration.** (Exercice, utiliser l'homogénéité). □

**Théorème 10.** *Le coût total moyen actualisé minimale  $V$  est l'unique solution bornée de l'équation d'optimalité*

$$V(x) = \inf_{a \in \mathcal{A}} [c(x, a) + \beta P V(x, a)] \quad x \in \mathcal{M}. \quad (7)$$

De plus, tout application  $\varphi: \mathcal{M} \rightarrow \mathcal{A}$  tel que

$$V(x) = [c + P V](x, \varphi(x)), \quad x \in \mathcal{M}$$

définit un contrôle markovien homogène  $u \in \mathcal{C}_0$  (par  $u_k(x_0, \dots, x_k) = \varphi(x_k)$ ) qui est optimal pour tout état initial  $x \in \mathcal{M}$ .

**Démonstration.** Est facile de voir que  $V$  est solution de (7) et que  $V$  est bornée par  $C/(1 - \beta)$ :

$$|V(x)| \leq C \sum_{m \geq 0} \beta^m = C/(1 - \beta).$$

Soit  $F$  une solution bornée de (7) et soit  $u \in \mathcal{C}_0$  un contrôle quelconque. Considérons le processus

$$M_n = \sum_{k=0}^{n-1} \beta^k c(X_k, U_k) + \beta^n F(X_n), \quad n \geq 0.$$

Alors

$$M_{n+1} - M_n = \beta^n c(X_n, U_n) + \beta^{n+1} F(X_{n+1}) - \beta^n F(X_n)$$

et

$$\mathbb{E}[M_{n+1} - M_n | X_n = y, U_n = a'] = \beta^n c(y, a') + \beta^{n+1} P F(y, a') - \beta^n F(y) \geq 0$$

qui donne que

$$F(x) = \mathbb{E}_{(0,x)}^u [M_0] \leq \mathbb{E}_{(0,x)}^u [M_n] = V_n^u(x) + \beta^n \mathbb{E}_{(0,x)}^u [F(X_n)].$$

En prenant la limite pour  $n \rightarrow \infty$  et utilisant l'hypothèse de bornitude sur  $F$  on obtient que

$$F(x) \leq V^u(x)$$

et par l'arbitrarité de  $u$  que  $F \leq V$ .

Si il existe un contrôle  $u$  markovien et homogène tel que  $F(x) = [c + \beta PF](x, u(x))$  pour tout  $n \geq 0$  et  $x \in \mathcal{M}$  alors on a que

$$\mathbb{E}[M_{n+1} - M_n | X_n = y, U_n = a'] = 0$$

et à la limite on obtient  $F(x) = V^u(x)$ . Alors  $F(x) \geq V(x)$  et  $F(x) = V(x) = V^u(x)$  ce qu'implique que le contrôle  $u$  est optimal. Si un tel contrôle n'existe pas on peut toujours raisonner de façon approché et considérer un contrôle  $\tilde{u}$  markovien et homogène tel que

$$F(x) \geq [c + \beta PF](x, \tilde{u}(x)) - \varepsilon \quad n \geq 0, x \in \mathcal{M}$$

pour  $\varepsilon > 0$ . Cette inégalité est équivalente à demander que

$$F(x) = [\tilde{c} + \beta PF](x, \tilde{u}(x))$$

pour une certaine fonction  $\tilde{c}(x, a) \geq c(x, a) - \varepsilon$ . Alors par l'argument précédent on obtient que

$$F(x) = \mathbb{E}_{(0,x)}^{\tilde{u}} \sum_{m \geq 0} \beta^m \tilde{c}(X_m, \tilde{u}(X_m)) \geq V^{\tilde{u}}(x) - \frac{\varepsilon}{1-\beta} \geq V(x) - \frac{\varepsilon}{1-\beta}$$

et par l'arbitrarité de  $\varepsilon > 0$  on conclut que  $F(x) \geq V(x)$  et donc que  $F(x) = V(x)$ . □