

[M Gubinelli - Controle des chaines de Markov - MI MD 2009/2010 - 20091218 - poly 3 - v.7]

# Charges de Markov contr

## 1 Chaî

On suppose des évolutions aléatoires sur un espace  $M$  en temps discret et que on peut modifier par la choix, chaque pas de temps, d'une action  $a$  dans un ensemble pré-défini admissibles  $A$ . Donné un état  $x$  au temps initial  $k$  et une politique de choix des actions de contrôle  $u$  on peut considérer la succession aléatoire  $X_k, \dots, X_n$  des états visités par notre système sujet à la politique  $u$ . Le problème que on va se poser est l'optimisation d'un quelque critère moyen par la choix d'une politique de contrôle (tout simplement un contrôle).

Soit donc  $M$  une espace d'états dénombrable,  $A$  l'espace des action admissibles et  $(M)$  l'espace des mesures de probabilité sur  $M$ . On considère une fonction  $P: N \times A \times M \rightarrow (M)$  qui donne au temps  $n \in N$ , une action  $a \in A$  et un état  $x \in M$  détermine la probabilité  $P_{n,a}(x, y) = P_{n,a}(x)(\{y\})$  que l'état  $y$  soit la suite suivante soit  $y \in M$ . La fonction  $P$  spécifie la dynamique aléatoire de notre système. Soit

$$M_k = \{(n, x_k, \dots, x_n) : n \in N, k \leq n, x_k, \dots, x_n \in M\}$$

on appelle une politique de contrôle (ou simplement un contrôle) une fonction  $u: M_k \rightarrow A$  et on appelle  $C_k$  l'espace des controls relatives à l'instant initial  $k = 0$ . La politique de contrôle  $u \in C_k$  est donc une règle qui au temps  $n$ , aient observé la succession d'états  $(x_k, \dots, x_n)$ , détermine l'action  $u_n(x_k, \dots, x_n) \in A$  pour modifier l'évolution future de notre système aléatoire. Un contrôle Markovien et stationnaire est un contrôle  $u \in C_k$  qui dépend seulement de l'état actuel du système, c-à-d tel que  $u_n(x_k, \dots, x_n) = u(x_n)$  pour une quelque fonction  $u: M \rightarrow A$ .

Soit l'espace canonique  $\Omega = M^N$  avec la tribu produit  $F$  (sur chaque composante on considère la tribu discrète des toutes les parties de  $M$ ). Soit  $X_n(\omega) = \omega_n$  la projection sur la  $n$ -ème composante de  $\Omega$ .

Donné un temps initial  $k \in N$ , un état initial  $x \in M$  et un contrôle  $u \in C_k$  on considère la probabilité  $P_{(k,x)}^u$  telle que

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

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

Lemme 1. On a que  $P_{(k,x)}^u$  vérifie (1)ssi,  $n \geq k$  on a

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

Démonstration. Exercice.

Une façon canonique de construire un processus contrôlé est de considérer une fonction

$$G: N \times M \times E \rightarrow M$$

et un espace de probabilité  $(E, F, P)$  muni d'une suite de v.a. iid  $(Z_n)_{n \geq k}$  à 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 appelée une référence aléatoire contrôlée ou un système dynamique aléatoire contrôlé. Il est facile de montrer (exercice) que la suite  $(X_n)_{n \geq k}$  vérifie

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

où

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

Réciproquement, pour toute fonction  $P: \mathbb{N} \times \mathcal{M} \times \mathcal{M} \rightarrow [0, 1]$  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  $Z_n \sim U([0, 1])$  et construire une  $G$  appropriée). La correspondance entre processus contrôlés et systèmes dynamiques contrôlés n'est pas univoque (plusieurs systèmes dynamiques contrôlés de points 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  $PF: \mathbb{N} \times \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$  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

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

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

## 2 Principe d'optimalité

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

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

l'espérance  $E_{(k,x)}^u$  est par rapport à la probabilité  $P_{(k,x)}^u$ . 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  $\sup_{n \geq k} \sup_{x,a} |c(n, x, a)| < +\infty$ . 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 appelée 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)

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

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

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

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 défini par  $\tilde{u}_n(X_{k+1}, \dots, X_n) = u_n(X_k, X_{k+1}, \dots, X_n)$ . On a que  $U_n = \tilde{u}_n(X_{k+1}, \dots, X_n) = \tilde{U}_n$ . Si on pose

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

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

$$\begin{aligned} E_{n \leq k+1} c(n, X_n, U_n) &= \sum_{y \in M} P(X_{k+1} = y) E_{n \leq k+1} [c(n, X_n, U_n) | X_{k+1} = y] \\ &= \sum_{y \in M} P(X_{k+1} = y) E_{P_{n,a}(x,y)}^{n \leq k+1} c(n, \tilde{X}_n, \tilde{U}_n) \\ &= \sum_{y \in M} P_{n,a}(x, y) V^{\tilde{U}}(k+1, y) = P V^{\tilde{U}}(k, x, a). \end{aligned}$$

Cela nous donne que

$$V(k, x) = \inf_{u \in C_k} V^u(k, x) = \inf_{u \in C_k} [c(k, x, a) + P V^{\tilde{U}}(k, x, a)].$$

Optimiser sur  $u \in C_k$  est équivalent à optimiser la choix initiale  $a = u_k(x)$  et le contrôle  $\tilde{u} \in C_{k+1}$  qui donne la stratégie sur les étapes suivantes donc

$$\begin{aligned} V(k, x) &= \inf_a \inf_{\tilde{u} \in C_k} [c(k, x, a) + P V^{\tilde{U}}(k, x, a)] = \inf_a [c(k, x, a) + P(\inf_{\tilde{u} \in C_k} V^{\tilde{U}})(k, x, a)] \\ &= \inf_a [c(k, x, a) + (P V)(k, x, a)] \end{aligned}$$

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  $V(k, x) = \sup_{u \in C_k} V^u(k, x)$  et l'équation de Bellman prend la forme

$$V(k, x) = \sup_a [c(n, x, a) + P V(n, x, a)].$$

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

Lemme 4. Si la fonction de coût et la fonction  $G$  ne dépendent pas du temps, c'est-à-dire si

$$V^u(k, x) = E_{n \leq k} c(X_n, U_n)$$

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

$$V(x) = \inf_a [c(x, a) + P V(x, a)]. \quad (5)$$

Démonstration. On considère  $u \in C_{k+1}$  et

$$V^u(k+1, x) = E_{n \leq k+1} c(X_n, U_n) = E_{n \leq k} c(X_{n+1}, U_{n+1})$$

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  $\tilde{u} \in C_k$  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 contrôle associé au système dynamique  $(G, (Z_{n+1})_{n \geq 1})$  avec contrôle  $\tilde{u}$  et état initiale  $(k, x)$ , donc

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

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

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

ce qui donne l'éq. (5).

### 3 Contrôle en horizon fini

L'équation de Bellman est un outil puissant pour caractériser (et des fois déterminer) les politiques optimales 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  $r: N \times M \times A \rightarrow \mathbb{R}$  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  $C: M \rightarrow \mathbb{R}$ . Le gain moyen associé au processus contrôlé par  $u$  et démarré en  $(k, x)$  est

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

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  $V(N-1, x)$ ,  $V(N-2, x)$  et ainsi de suite jusqu'à déterminer  $V(k, x)$  au temps initial. L'équation de Bellman a donc une seule solution.

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

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

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

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

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

Alors pour tout  $k \leq n < N-1$  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

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

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

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

**Exemple 6.** (Exercer une option d'achat) On a la possibilité d'acheter un actif à un prix fixe  $p$  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  $Y_{n+1} = Y_n + \sigma_{n+1} \epsilon_n$  où  $(\epsilon_n)_{n \geq 1}$  est une suite iid indépendante. L'objectif est de maximiser le gain moyen relatif à la 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 donné par la suite des valeurs de notre option et on prend comme espace d'états l'ensemble  $M = \mathbb{R} \times \{0, 1\}$  car à un instant déterminé soit on possède encore l'option et sa valeur est  $x \in \mathbb{R}$ , soit on a exercé l'option et alors on décide de faire la conventionnelle de 160 dans l'état 0. L'espace des actions est  $A = \{0, 1\}$ , 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 réviser la dynamique contrôlée comme dynamique aléatoire contrôlée. La fonction de gain est donnée par  $r(n, x, a) = a(x - q)$  et la dynamique aléatoire par

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

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

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

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

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

L'équation de Bellman est alors donnée par

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

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

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

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

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

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

$$V^u(x) = E_{(0,x)}^u \sum_{m=0}^{\infty} r(X_m, U_m)$$

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

$$V_n^u(x) = E_{(0,x)}^u \sum_{m=0}^n r(X_m, U_m), \quad V_n(x) = \sup_{u \in \mathcal{C}_0} V_n^u(x).$$

Par convergence monotone  $V_n^u(x) \rightarrow V^u(x)$  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 fonctions  $V_n(x)$  peuvent être calculées par récurrence.

Lemme 7. On a l'équation

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

Démonstration. (Exercice, utiliser l'homogénéité)

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

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

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

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

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

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  $c: M \rightarrow \mathbb{R}$  (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

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

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 coûts partiels

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

On remarque que

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

si  $n \rightarrow \infty$ , c'est à dire

$$V_n^u(x) \xrightarrow[n \rightarrow \infty]{} V^u(x) \quad V_n(x) \xrightarrow[n \rightarrow \infty]{} V(x)$$

pour tout  $n \geq 0$ . En optimisant sur  $u$  on obtient de même

$$V_n(x) \xrightarrow[n \rightarrow \infty]{} V(x)$$

ce qui nous donne aussi

$$|V_n(x) - V(x)| \xrightarrow[n \rightarrow \infty]{} 0.$$

**Lemme 9.** On a l'équation

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

**Démonstration.** On considère le problème non-homogène d'optimisation associé à la fonction

$$W_n^u(k, x) = \mathbb{E}_{(k,x)}^u \sum_{m=k}^{\infty} \beta^m d(m, X_m, U_m)$$

avec  $d(n, x, a) = \beta^n c(x, a)$  et  $u \in \mathcal{C}_k$ . On remarque que  $W_n^u(0, x) = V_n^u(x)$ . L'équation de Bellman associée à  $W_{n+1}(k, x) = \inf_{u \in \mathcal{C}_k} W_{n+1}^u(k, x)$  est

$$W_{n+1}(k, x) = \inf_{a \in A} [d(k, x, a) + \beta W_{n+1}(k+1, x, a)]$$

car le processus est homogène et donc le probabilité de transition ne dépend pas du temps. Or pour  $k=0$  on a que

$$W_{n+1}^u(1, x) = \mathbb{E}_{(1,x)}^u \sum_{m=1}^{\infty} \beta^m d(m, X_m, U_m)$$

pour  $u \in \mathcal{C}_1$ . Par le même argument utilisé dans la preuve du lemme 4 sur l'homogénéité on a que cette quantité est équivalente à

$$W_{n+1}^u(1, x) = \mathbb{E}_{(0,x)}^u \sum_{m=0}^{\infty} \beta^m d(m+1, X_m, U_m) = \mathbb{E}_{(0,x)}^u \sum_{m=0}^{\infty} \beta^m d(m, X_m, U_m) = W_n^u(0, x)$$

où  $\tilde{C}_0$  est défini par  $\tilde{U}_k(x_0, \dots, x_k) = U_{k+1}(x_0, \dots, x_k)$  pour tout  $k \geq 0$  et  $u \in C_1$ . Donc

$$W_{n+1}(1, x) = \inf_{u \in C_1} W_{n+1}^u(1, x) = \inf_{\tilde{u} \in C_0} W_n^{\tilde{u}}(0, x) = W_n(0, x) = V_n(x)$$

ce qui donne l'équation

$$V_{n+1}(x) = W_{n+1}(0, x) = \inf_{a \in A} [d(0, x, a) + \beta V_n(x, a)].$$

Remarque 10. La même preuve peut être utilisée pour montrer que  $V$  est solution de

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

Il suffit de considérer le cas  $n = \infty$  dans l'argument.

**Théorème 11.** *Le coût total moyen actualisé minimal est lié à la unique solution bornée de l'équation d'optimalité*

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

De plus, toute application  $V : M \rightarrow \mathbb{R}$  telle que

$$V(x) = [c + \beta V](x, u(x)), \quad x \in M$$

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

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

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

Soit  $F$  une solution bornée de (7) et soit  $u \in C_0$  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

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

et

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

qui donne que

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

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

$$F(x) = V^u(x)$$

et par arbitraire de  $u$  que  $F = V$ .

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

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

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

$$F(x) = [c + \beta F](x, \tilde{u}(x)) \quad \forall n \geq 0, x \in M$$

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

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

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

$$F(x) = E_{(0,x)}^{\tilde{u}} \left[ \sum_{m=0}^{\infty} \tilde{c}(X_m, \tilde{u}(X_m)) \right] = V^{\tilde{u}}(x) \leq V(x)$$

et par l'arbitraire de  $\epsilon > 0$  on conclut que  $F(x) = V(x)$  et donc que  $F(x) = V(x)$ .

## 6 Contrôle en horizon infini: cas des coûts positifs

Dans cette section on fait l'hypothèse d'avoir des coûts  $c: M \times A \rightarrow \mathbb{R}_+$  positifs et homogènes dans le problème de minimisation et on définit

$$V^u(x) = E_{(0,x)}^u \left[ \sum_{m=0}^{\infty} c(X_m, U_m) \right], \quad V(x) = \inf_{u \in \mathcal{C}_0} V^u(x).$$

Comme dans le cas des gains positifs on a la convergence monotone des  $V_n^u(x)$  vers  $V^u(x)$ :

$$V_n^u(x) = E_{(0,x)}^u \left[ \sum_{m=0}^{n-1} c(X_m, U_m) \right] \leq V^u(x)$$

pour  $n \geq 1$ .

**Théorème 12.** Soit  $A$  fini. Alors la fonction valeur  $V$  est la solution positive minimale de l'équation d'optimalité

$$V(x) = \min_{a \in A} (c(x, a) + P V)(x, a), \quad x \in M.$$

De plus, toute application  $u: M \rightarrow A$  telle que

$$V(x) = (c + P V)(x, u(x)), \quad x \in M$$

définit un contrôle markovien homogène qui est optimal pour tout coût initial.

**Démonstration.** Par le principe de programmation dynamique la fonction  $V$  est solution de l'équation d'optimalité. Soit  $F$  une autre solution tel que  $F(x) \geq 0$ , par la récurrence de  $A$  il existe une application  $\tilde{u}: M \rightarrow A$  telle que

$$F(x) = (c + P F)(x, \tilde{u}(x)), \quad x \in M.$$

On a que

$$F(x) = E_{(0,x)}^{\tilde{u}}[M_0] = E_{(0,x)}^{\tilde{u}}[M_n] = V_n^{\tilde{u}}(x) + E_{(0,x)}^{\tilde{u}}[F(X_n)] \geq V_n^{\tilde{u}}(x)$$

et la limite pour  $n \rightarrow \infty$  on obtient  $F(x) \geq V^{\tilde{u}}(x) = V(x)$ . Si  $F = V$  on peut prendre  $u = \tilde{u}$  et vérifier que  $V = V^u$  et donc que  $u$  donne un contrôle optimal.

**Corollaire 13.** (Itération de la fonction valeur) Soit  $V_n(x) = \inf_{u \in \mathcal{C}_0} V_n^u(x)$ . On a que  $V_n \leq V$ .

**Démonstration.** Par convergence monotone on a que  $V(x) = \lim_n V_n(x)$  est bien définie et positive. Par l'équation d'optimalité en horizon fini

$$V_{n+1}(x) = \inf_{a \in A} [c(x, a) + P V_n(x, a)]$$

et en prenant la limite pour  $n \rightarrow \infty$  on a que

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

mais alors  $V \leq V$ . D'autre part  $V_n \leq V_n^u \leq V^u$  et donc  $V \leq V^u$  et  $V = V^u$  ce qui donne que  $V = V$ .



En pratique on peut donc trouver la fonction  $V$  par approximation avec des problèmes en horizon fini  $V_n$ . On peut aussi chercher à trouver des politiques de contrôle non optimales. En effet si  $u: M \rightarrow A$  est un contrôle markovien alors on sait que

$$V^u(x) = (c + PV^u)(x, u(x)), \quad x \in M$$

et si  $V^u$  ne satisfait pas l'équation d'optimalité on peut trouver un contrôle  $\tilde{u}: M \rightarrow A$  meilleur dans le sens que

$$V^u(x) > (c + PV^u)(x, \tilde{u}(x)), \quad x \in M$$

avec une inégalité stricte pour un quelque  $x_0 \in M$ . Alors, évidemment,  $V^u - V_0^{\tilde{u}} = 0$  et si on suppose que  $V^u - V_n^{\tilde{u}}$  on a

$$V^u(x) - (c + PV^u)(x, \tilde{u}(x)) = (c + PV_n^{\tilde{u}})(x, \tilde{u}(x)) - V_{n+1}^{\tilde{u}}(x)$$

ce qui donne que  $V^u - V_n^{\tilde{u}}$  pour tout  $n \geq 0$  et donc que  $V^u - V^{\tilde{u}}$  avec une inégalité stricte pour  $x_0 \in M$ .

## 7 Optimisation de la moyenne sur des long temps

Ici on considère une fonction coût  $c: M \times A \rightarrow \mathbb{R}$  bornée et on définit

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

On dit que un contrôle  $u$  est optimale en départ de  $x$  si la limite

$$= \lim_n \frac{V_n^u(x)}{n}$$

existe et si pour tout autre contrôle  $\tilde{u}$  on a que

$$\liminf_n \frac{V_n^{\tilde{u}}(x)}{n} \geq \lim_n \frac{V_n^u(x)}{n}.$$

La valeur est alors appelée le coût minimal par unité de temps en départ de  $x$ .

**Théorème 14.** Si il existe une constante  $\beta$  et une fonction bornée  $\gamma: M \rightarrow \mathbb{R}$  tels que

$$\gamma(x) = (c + P\gamma)(x, a), \quad x \in M, a \in A.$$

Alors pour tout contrôle  $u \in \mathcal{C}_0$  et tout  $x \in M$ ,

$$\liminf_n \frac{V_n^u(x)}{n} \geq \gamma(x).$$

**Démonstration.** Soit

$$M_n = \gamma(X_n) + \sum_{k=0}^{n-1} c(X_k, U_k) - \beta \gamma(X_0).$$

Alors

$$M_{n+1} - M_n = \gamma(X_{n+1}) - \gamma(X_n) + c(X_n, U_n) - \beta \gamma(X_0)$$

et pour tout  $y \in M, a \in A$ :

$$\mathbb{E}[M_{n+1} - M_n | X_n = y, U_n = a] = P\gamma(y, a) - \gamma(y) + c(y, a) - \beta \gamma(X_0) = 0$$

Donc

$$\gamma(x) = \mathbb{E}_{(0,x)}^u[M_0] = \mathbb{E}_{(0,x)}^u[M_n] = \mathbb{E}^u[\gamma(X_n)] - \beta \gamma(X_0) + V_n^u(x)$$

et

$$\frac{V_n^u(x)}{n} = \frac{\gamma(x)}{n} + \frac{\mathbb{E}^u[\gamma(X_n)]}{n} - \frac{\beta \gamma(X_0)}{n}$$

car  $\beta_\infty$  est borné.

Un argument similaire donne

**Théorème 15.** Si il existe une constante  $\beta_\infty$ , une fonction bornée et un contrôle tels que

$$\beta_\infty + (x) = (c + P)(x, u(x)), \quad x \in M$$

alors pour tout  $x \in M$ ,

$$\limsup_n \frac{V^u(x)}{n} \leq \beta_\infty.$$

Donc si  $\beta_\infty$  satisfait

$$\beta_\infty + (x) = \inf_{a \in A} (c + P)(x, a)$$

et si le minimum est atteint ( $u(x)$  pour tout  $x \in M$ ) alors  $u$  est un contrôle optimal pour tout  $x \in M$ .

Soit  $V_n$  la fonction valeur de horizon  $n$  donnée par  $V_0(x) = 0$  et  $V_{n+1}(x) = \inf_a (c + P V_n)(x, a)$ . Soit

$$\beta_k^- = \inf_x \{V_{k+1}(x) - V_k(x)\} \quad \beta_k^+ = \sup_x \{V_{k+1}(x) - V_k(x)\}$$

**Théorème 16.** Pour tout  $k \geq 0$  et tout contrôle on a que

$$\liminf_n \frac{V^u(x)}{n} \geq \beta_k^-.$$

De plus, si il existe  $u: M \rightarrow A$  tel que

$$V_{k+1}(x) = (c + P V_k)(x, u(x)), \quad x \in M,$$

alors

$$\limsup_n \frac{V^u(x)}{n} \leq \beta_k^+.$$

**Démonstration.** On remarque que

$$\beta_k^- + V_k(x) \leq V_{k+1}(x) = (c + P V_k)(x, a), \quad x \in M, a \in A$$

et

$$\beta_k^+ + V_k(x) \geq V_{k+1}(x) = (c + P V_k)(x, u(x)), \quad x \in M,$$

et donc on peut appliquer les théorèmes précédents avec  $\beta_k = \beta_k^-$ .