[M. Gubinelli - Controle des chaines de Markov - M1 MMD 2009/2010 - 20091218 - poly 3 - v.7]

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 $\mathcal{A}$. Donné un état $x \in \mathcal{M}$ au temps initial $k \in \mathbb{N}$ et une politique de choix des actions de contrôle u on peut considérer la succession aléatoire X_k, ..., X_n, ... des états visité par notre système sujet à la politique u. Le problème que on va se poser est l'optimisation d'un quelque critère moyenne par la choix d'une politique de contrôle (tout simplement un contrôle).

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

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

on appelle une politique de contrôle (ou simplement un contrôle) une fonction $u\colon \mathcal{M}_k^* \to \mathcal{A}$ et on appelle $\mathcal{C}_k$ l'espace des controls relatives à l'instant initiale $k \geqslant 0$. La politique de contrôle $u \in \mathcal{C}_k$ est donc une règle qui au temps n, aient observé la succession d'états $(x_k, ...x_n)$, détermine l'action $u_n(x_k, ..., x_n) \in \mathcal{A}$ pour modifier l'évolution future de notre système aléatoire. Un contrôle Markovien et stationnaire est un contrôle $u \in \mathcal{C}_k$ qui dépend seulement de l'état actuel du système, c-à-d tel que $u_n(x_k, ..., x_n) = \varphi(x_n)$ pour une quelque fonction $\varphi\colon \mathcal{M} \to \mathcal{A}$.

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

Donné un temps initial $k \in \mathbb{N}$, un état initial $x \in \mathcal{M}$ et un contrôle $u \in \mathcal{C}_k$ on considère la probabilité $\mathbb{P}^u_{(k,x)}$ telle que

$$\mathbb{P}^u_{(k,x)}(X_k = x_k, ..., X_{n+1} = x_{n+1}) = \delta_{x,x_k} \prod_{i=k}^{n} P_{k,u_i(x_k,...,x_i)}(x_i, x_{i+1}) \qquad \forall n \geqslant k. \tag{1}$$

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

Lemme 1. *On a que $\mathbb{P}^u_{(k,x)}$ vérifie (1) ssi, $\forall n \geqslant k$ on a*

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

Démonstration. Exercice. □

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

$$G\colon \mathbb{N} \times \mathcal{M} \times \mathcal{A} \times E \to \mathcal{M}$$

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

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

où $U_n = u_n(X_k, ..., X_n)$. 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 $(X_n)_{n \geqslant k}$ vérifie

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

où

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

Réciproquement, pour tout fonction $P\colon \mathbb{N}\times\mathcal{A}\times\mathcal{M}\to\Pi(\mathcal{M})$ il est possible de trouver un espace auxiliaire E, une suite iid $(Z_n)_{n\geqslant k}$ 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 contrôlées différents peuvent avoir la même loi et donc correspondre au même processus contrôlé).

Donnée une fonction $F\colon \mathbb{N}\times\mathcal{M}\to\mathbb{R}$ on définit la fonction $PF\colon \mathbb{N}\times\mathcal{M}\times\mathcal{A}\to\mathbb{R}$ 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 $c\colon \mathbb{N}\times\mathcal{M}\times\mathcal{A}\to\mathbb{R}$ 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. Donné 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 $(X_n)_{n\geqslant k}$ est déterminé par

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

où l'espérance $\mathbb{E}^u_{(k,x)}$ est par rapport à la probabilité $\mathbb{P}^u_{(k,x)}$. Pour que cet expression ait un sens on admet que une des condition suivantes est satisfaite: $c(n,\,x,\,a)\geqslant 0$, $c(n,\,x,\,a)\leqslant 0$ ou $\sum_{n\geqslant k}\sup_{x,a}|c(n,\,x,\,a)|<+\infty$. 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 $V\colon \mathbb{N}\times\mathcal{M}\to\mathbb{R}$.

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 $(X_n)_{n\geqslant k}$ est un système dynamique contrôlée associé à la fonction G et à la suite aléatoire $(Z_n)_{n\geqslant k}$. Donc

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

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

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

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

$$\mathbb{E}\sum_{n\geqslant k+1} c(n, X_n, U_n) = \sum_{y\in\mathcal{M}} \mathbb{P}(X_{k+1}=y)\mathbb{E}[\sum_{n\geqslant k+1} c(n, X_n, U_n)|X_{k+1}=y]$$

$$= \sum_{y\in\mathcal{M}} \underbrace{\mathbb{P}(X_{k+1}=y)}_{P_{n,a}(x,y)} \underbrace{\mathbb{E}\sum_{n\geqslant k+1} c(n, \tilde{X}_n, \tilde{U}_n)}_{V^{\tilde{u}}(k+1,y)}$$

$$= \sum_{y\in\mathcal{M}} P_{n,a}(x,y) V^{\tilde{u}}(k+1, y) = PV^{\tilde{u}}(k, x, a).$$

Cela nous donne que

$$V(k,x) = \inf_{u\in\mathcal{C}_k} V^u(k,x) = \inf_{u\in\mathcal{C}_k} [c(k,x,a) + PV^{\tilde{u}}(k,x,a)].$$

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

$$V(k,x) = \inf_{a\in\mathcal{A}} \inf_{\tilde{u}\in\mathcal{C}_k} [c(k,x,a) + PV^{\tilde{u}}(k,x,a)] = \inf_{a\in\mathcal{A}} [c(k,x,a) + P(\inf_{\tilde{u}\in\mathcal{C}_k} V^{\tilde{u}})(k,x,a)]$$

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

□

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 dépends pas du temps, i.e. si $P\colon \mathcal{A}\times\mathcal{M} \to \Pi(\mathcal{M})$. Similairement on dit que un système dynamique contrôlé est homogène si la fonction G ne dépends pas du temps: $G\colon \mathcal{M}\times\mathcal{A}\times E \to \mathcal{M}$. 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*

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

et

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

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

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

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

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

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

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

et donc $V(k,x)=V(k+1,x)$ pour tout $k\geqslant 0$. 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 $N\geqslant 0$ un temps et soit $r\colon \mathbb{N}\times\mathcal{M}\times\mathcal{A}\to\mathbb{R}$ 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 $C\colon\mathcal{M}\to\mathbb{R}$. Le gain moyen associé au processus contrôlé par u et démarrant en (k,x) est

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

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

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

pour tout $k\leqslant n<N$ 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 jusque 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*

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

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 $(X_n)_{n\geqslant k}$ le processus contrôlé associé. Soit

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

Alors pour tout $k\leqslant n\leqslant N-1$ 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

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

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

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

□

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 $n=0,\ldots,N-1$. Le prix de marché de l'actif est modélisé par une suite $(Y_n)_{n\geqslant 0}$ donnée par $Y_{n+1}=Y_n+\varepsilon_{n+1}$ où $(\varepsilon_n)_{n\geqslant 1}$ 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 Δ. 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 $(\varepsilon_n)_{n \geqslant 0}$. En particulier la fonction de transition P est de la forme

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

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

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

L'équation de Bellman est alors donnée par

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

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 $V(k, x) \geqslant V(k+1, x)$ pour tout $0 \leqslant k \leqslant N$ et tout $x \in \mathbb{R}$.

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

4 Contrôle en horizon infini: cas des gains positifs

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

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

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

$$V^u_n(x) = \mathbb{E}^u_{(0,x)} \sum_{0 \leqslant m \leqslant n-1} r(X_m, U_m), \qquad V_n(x) = \sup_{u \in \mathcal{C}_0} V^u_n(x).$$

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

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

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*

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

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 $F\colon \mathcal{M} \to \mathbb{R}_+$ un autre solution non-négative de (6). Alors $F(x) \geqslant 0 = V_0(x)$. Supposons par induction que $F \geqslant V_n$, alors

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

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

5 Contrôle en horizon infini: cas des coûts actualisés

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

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

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

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

On remarque que

$$|V^u_n(x) - V^u(x)| \leqslant \sum_{m\geqslant n} \beta^m C = C\frac{\beta^n}{1-\beta} = \varepsilon_n \to 0$$

si $n \to \infty$, c'est à dire

$$V^u_n(x) - \varepsilon_n \leqslant V^u(x) \leqslant V^u_n(x) + \varepsilon_n$$

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

$$V_n(x) - \varepsilon_n \leqslant V(x) \leqslant V_n(x) + \varepsilon_n$$

ce qui nous donne aussi

$$|V_n(x) - V(x)| \leqslant \varepsilon_n \to 0.$$

Lemme 9. *On a l'équation*

$$V_{n+1}(x) = \inf_{a\in\mathcal{A}} [c(x,a) + \beta PV_n(x,a)] \qquad n \geqslant 0, x \in \mathcal{M}$$

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

$$W^u_n(k,x) = \mathbb{E}^u_{(k,x)} \sum_{m=k}^{n-1} d(m, X_m, U_m)$$

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

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

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

$$W^u_{n+1}(1,x) = \mathbb{E}^u_{(1,x)} \sum_{m=1}^{n} d(m, X_m, U_m)$$

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 équivalent à

$$W^u_{n+1}(1,x) = \mathbb{E}^{\tilde{u}}_{(0,x)} \sum_{m=0}^{n} d(m+1, X_m, U_m) = \beta\mathbb{E}^{\tilde{u}}_{(0,x)} \sum_{m=0}^{n-1} d(m, X_m, U_m) = \beta W^{\tilde{u}}_n(0,x)$$

où $\tilde{u} \in \mathcal{C}_0$ est défini par $\tilde{u}_k(x_0, ..., x_k) = u_{k+1}(x_0, ..., x_k)$ pour tout $k \geqslant 0$ et $u \in \mathcal{C}_1$. Donc

$$W_{n+1}(1,x) = \inf_{u \in \mathcal{C}_1} W^u_{n+1}(1,x) = \inf_{\tilde{u} \in \mathcal{C}_0} \beta W^{\tilde{u}}_n(0,x) = \beta W_n(0,x) = \beta V_n(x)$$

ce qui donne l'équation

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

□

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

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

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

Théorème 11. *Le coût total moyen actualisé minimale V est l'unique solution bornée de l'équation d'optimalité*

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

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

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

définit un contrôle markovien homogène $u \in \mathcal{C}_0$ (par $u_k(x_0, ..., x_k) = \varphi(x_k)$) 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 $C/(1-\beta)$:

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

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

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

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

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

qui donne que

$$F(x) = \mathbb{E}^u_{(0,x)}[M_0] \leqslant \mathbb{E}^u_{(0,x)}[M_n] = V^u_n(x) + \beta^n \mathbb{E}^u_{(0,x)}[F(X_n)].$$

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

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

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

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

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

et à la limite on obtient $F(x) = V^u(x)$. Alors $F(x) \geqslant V(x)$ 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

$$F(x) \geqslant [c + \beta P F](x, \tilde{u}(x)) - \varepsilon \qquad n \geqslant 0, x \in \mathcal{M}$$

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

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

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

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

et par l'arbitrarité de $\varepsilon > 0$ on conclut que $F(x) \geqslant V(x)$ 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 $c\colon \mathcal{M} \times \mathcal{A} \to \mathbb{R}_+$ positifs et homogènes dans le problème de minimisation et on définit

$$V^u(x) = \mathbb{E}^u_{(0,x)} \sum_{m \geqslant 0} c(X_m, U_m), \qquad V(x) = \inf_{u \in \mathcal{C}_0} V^u(x).$$

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

$$V^u_n(x) = \mathbb{E}^u_{(0,x)} \sum_{m=0}^{n-1} c(X_m, U_m) \nearrow V^u(x)$$

pour $n \to \infty$.

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

$$V(x) = \min_{a \in \mathcal{A}} (c + PV)(x,a), \qquad x \in \mathcal{M}.$$

De plus, toute application $u\colon \mathcal{M} \to \mathcal{A}$ telle que

$$V(x) = (c + PV)(x, u(x)), \qquad x \in \mathcal{M}$$

définit un contrôle markovien homogène qui est optimal pour tout état initiale $x \in \mathcal{M}$.

Démonstration. Par le principe de programmation dynamique la fonction V est solution de l'équation d'optimalité. Soit F un autre solution tel que $F(x) \geqslant 0$, par la finitude de $\mathcal{A}$ il existe une application $\tilde{u}\colon \mathcal{M} \to \mathcal{A}$ telle que

$$F(x) = (c + PF)(x, \tilde{u}(x)), \qquad x \in \mathcal{M}.$$

On a que

$$F(x) = \mathbb{E}^{\tilde{u}}_{(0,x)}[M_0] = \mathbb{E}^{\tilde{u}}_{(0,x)}[M_n] = V^{\tilde{u}}_n(x) + \mathbb{E}^{\tilde{u}}_{(0,x)}[F(X_n)] \geqslant V^{\tilde{u}}_n(x)$$

et à la limite où $n \to \infty$ on obtient $F(x) \geqslant V^{\tilde{u}}(x) \geqslant V(x)$. Si $F = V$ on peut prendre $u = \tilde{u}$ et vérifier que $V \geqslant V^u$ et donc que u donne un contrôle optimale. □

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

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

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

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

$$V_\infty(x) = \inf_{a \in \mathcal{A}} [c(x,a) + PV_\infty(x,a)]$$

mais alors $V_\infty \geqslant V$. D'autre part $V_n \leqslant V^u_n \leqslant V^u$ et donc $V_\infty \leqslant V^u$ et $V_\infty \leqslant V$ ce qui donne que $V_\infty = V$. □

En pratique on peut donc trouver la fonction V par approximation avec des problèmes en horizon fini V_n. On peut aussi chercher d'améliorer des politique de contrôle non optimales. En effet si $u\colon \mathcal{M} \to \mathcal{A}$ est un contrôle markovien alors on sait que

$$V^u(x) = (c + PV^u)(x, u(x)), \qquad x \in \mathcal{M}$$

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

$$V^u(x) \geqslant (c + PV^u)(x, \tilde{u}(x)), \qquad x \in \mathcal{M}$$

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

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

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

7 Optimisation de la moyenne sur des long temps

Ici on considère une fonction coût $c\colon \mathcal{M} \times \mathcal{A} \to \mathbb{R}$ bornée et on définit

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

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

$$\lambda = \lim_{n\to\infty} \frac{V_n^u(x)}{n}$$

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

$$\lambda \leqslant \liminf_{n\to\infty} \frac{V_n^{\tilde{u}}(x)}{n}.$$

La valeur λ est alors appelé le *coût minimale par unité de temps* en démarrant de x.

Théorème 14. *Si il existe une constante λ et une fonction bornée $\theta\colon \mathcal{M} \to \mathbb{R}$ tels que*

$$\lambda + \theta(x) \leqslant (c + P\theta)(x, a), \qquad x \in \mathcal{M}, a \in \mathcal{A}.$$

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

$$\liminf_{n\to\infty} \frac{V^u(x)}{n} \geqslant \lambda.$$

Démonstration. Soit

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

Alors

$$M_{n+1} - M_n = \theta(X_{n+1}) - \theta(X_n) + c(X_n, U_n) - \lambda$$

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

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

Donc

$$\theta(x) = \mathbb{E}^u_{(0,x)}[M_0] = \mathbb{E}^u_{(0,x)}[M_n] = \mathbb{E}^u[\theta(X_n)] - n\,\lambda + V_n^u(x)$$

et

$$\frac{V_n^u(x)}{n} \geqslant \lambda + \frac{\theta(x)}{n} - \frac{\mathbb{E}^u[\theta(X_n)]}{n} \to \lambda$$

car θ est bornée. □

Un argument similaire donne

Théorème 15. *Si il existe une constante λ, une fonction bornée θ et un contrôle u tels que*

$$\lambda + \theta(x) \geqslant (c + P\theta)(x, u(x)), \qquad x \in \mathcal{M}$$

alors pour tout $x \in \mathcal{M}$,

$$\limsup_{n\to\infty} \frac{V^u(x)}{n} \leqslant \lambda.$$

Donc si λ, θ satisfont

$$\lambda + \theta(x) = \inf_{a\in\mathcal{A}} (c + P\theta)(x, a)$$

et si l'infimum est atteint à $u(x)$ pour tout $x \in \mathcal{M}$ alors u est un contrôle optimal pour tout $x \in \mathcal{M}$.

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

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

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

$$\liminf_{n\to\infty} \frac{V^u(x)}{n} \geqslant \lambda_k^-.$$

De pus, si il existe $u\colon \mathcal{M} \to \mathcal{A}$ tel que

$$V_{k+1}(x) = (c + PV_k)(x, u(x)), \qquad x \in \mathcal{M},$$

alors

$$\limsup_{n\to\infty} \frac{V^u(x)}{n} \leqslant \lambda_k^+.$$

Démonstration. On remarque que

$$\lambda_k^- + V_k(x) \leqslant V_{k+1}(x) \leqslant (c + PV_k)(x, a), \qquad x \in \mathcal{M}, a \in \mathcal{A}$$

et

$$\lambda_k^+ + V_k(x) \geqslant V_{k+1}(x) = (c + PV_k)(x, u(x)), \qquad x \in \mathcal{M},$$

et donc on peut appliquer les théorèmes précédentes avec $\theta = V_k$. □