HACHIRO AKAMA

Un aspect de la programmation dynamique

Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche, tome 2 (1957), p. 27-36

<http://www.numdam.org/item?id=BURO_1957__2__27_0>

UN ASPECT DE LA PROGRAMMATION DYNAMIQUE

(problème des mines d'or)

par

Hachiro AKAMA

de l'Institut technique de recherche de la Défense Nationale de Tokyo

I - INTRODUCTION

La théorie de la programmation dynamique, qui se caractérise par l'application du principe de la stratégie optimum à des programmes récurrents, est devenue récemment un des sujets les plus importants de recherche opérationnelle, et les travaux sur cette question se multiplient: citons ceux du groupe de M: Richard Bellman aux Etats-Unis et de l'équipe de M. G. Th. Guilbaud en France.

Nous nous proposons de faire ici un exposé limité du problème dit des "mines d'or", en nous bornant à en souligner quelques aspects, fondamentaux au point de vue mathématique.

Il y a lieu de signaler que la formulation de tous les problèmes de programmation dynamique s'appuie sur les bases suivantes :

1. Le programme (ou processus) dont il s'agit est constitué par une série de décisions consécutives.

2. L'état du système se décrit complètement au moyen d'un ensemble fini de paramètres, i.e., par un vecteur dans un espace abstrait.

3. Chaque décision produit une transformation de ce vecteur.

4. A chaque stade, la décision est prise conformément au principe suivant : quelle que soit l'histoire passée du système, cette décision doit constituer avec les suivantes la stratégie optimum à l'égard de l'état actuel du système.

II - MISE EN ÉQUATION DU PROBLÈME

Supposons qu'il y ait deux mines d'or, A et B dont les richesses exploitables sont respectivement x et y unités de quantité d'or. Si la seule machine à notre disposition est employée dans la mine A pendant une période déterminée, il y a la probabilité p_1 pour que nous obtenions une quantité d'or $r_1 x$, la machine étant encore en bon état, tandis qu'il y a la probabilité $(1-p_1)$ pour que la machine se détériore sans exploiter aucun or. Il en est de même de la mine B, à laquelle sont liées la quantité d'or $r_2 y$ et la probabilité p_2.

Le problème consiste à chercher une stratégie optimum pour maximiser la quantité d'or exploitée au cours de plusieurs périodes.

Soit $f_n(x,y)$ l'espérance mathématique de la quantité d'or exploitée durant n périodes par la stratégie optimum. On obtient d'après le principe d'optimalité, la formule

$$(2.1) \qquad f_n(x,y) = \text{Max} \begin{cases} A : p_1 r_1 x + p_1 f_{n-1}[(1-r_1)x, y] \\ B : p_2 r_2 y + p_2 f_{n-1}[(x, (1-r_2) y] \end{cases}$$

qui indique dans quelle mine la machine doit travailler à la première période : si elle est employée dans A, il y a une probabilité p_1 pour qu'on obtienne pendant cette période une quantité d'or $r_1 x$ et pour qu'il reste encore dans les mines A et B des richesses respectives $(1-r_1)x$ et y, exploitables durant les (n-1) périodes suivantes.

Il est tout naturel d'examiner la limite de (2.1).

$$(2.2) \qquad f(x,y) = \text{Max} \begin{cases} A : p_1 r_1 x + p_1 f[(1-r_1)x, y] \\ B : p_2 r_2 y + p_2 f[x, (1-r_2)y] \end{cases}$$

quand n tend vers l'infini.

Le passage à la limite, quoiqu'il soulève la question de l'existence et de l'unicité de la solution $f(x,y)$, a pour effet de simplifier la structure du problème, puisqu'il reste toujours une infinité de stades après un nombre fini de périodes. C'est ainsi que l'on arrivera à choisir une décision sans tenir compte ni du passé ni de l'avenir très lointain, mais en tenant compte seulement du présent et du futur assez proche. On considérera maintenant n comme la mesure du temps, et non plus comme le nombre des périodes.

La formule (2.2) se généralise comme suit :

$$(2.3) \qquad \sum (\omega; s^*) \,\|\, f(\omega) = \underset{\delta_1}{\text{Max}}\left[g(\omega;\delta_1) + P(\delta_1)\, f(T_{\delta_1}\omega)\right]$$

ω est un vecteur dans l'espace à k dimensions, et représente l'état du système,

$\delta_1 = \{\delta_1^1, \delta_1^2, \ldots, \delta_1^k\}$ est l'ensemble fini des décisions possibles à l'instant $n = 1$;

T_{δ_1} est un opérateur (éventuellement aléatoire) associé à δ_1, et transformant le vecteur en un autre vecteur $T_{\delta_1}\omega$, qui vérifie la condition

$$|\omega|^2 \geqslant |T_{\delta_1}\omega|^2$$

$P(\delta_1)$ est un coefficient (de probabilité) associé à δ_1, et soumis à la condition

$$0 \leqslant P(\delta_1) \leqslant \alpha < 1$$

$g(\omega\,;\delta_1)$ désigne l'espérance mathématique du gain immédiat résultant de la décision δ_1, et $f(\omega)$ est l'espérance mathématique du gain total

$\sum(\omega;s^*)$ représente le processus au cours duquel on obtient le gain optimum $f(\omega)$ par la stratégie optimum s^*.

Par définition δ_1 fait partie de s^*.

Des considérations pratiques imposent

$$0 \leqslant f(\omega) \leqslant K$$

K étant un nombre fini positif dépendant de $\sum$, et

$$f(\omega) \geqslant f(T_{\delta_1}\omega)$$

Soient

$$\omega = (x_1, \dots, x_k)$$

$$\delta_1 = (1, 2, \dots, k)$$

$T_{\delta_1^i}\,\omega = (x_1, \dots, c'_{ij}\,x_i, \dots, x_k)$ avec la probabilité p_{ij} $(j = 1, 2, \dots \ell)$

$$g(\omega\,;\delta_1^i) = \sum_{j=1}^{l} p_{ij}\,c_{ij}\,x_i$$

$$P(\delta_j^i) = \sum_{j=1}^{l} p_{ij} < 1$$

$$c_{ij} + c'_{ij} = 1$$

(2.3) nous donne alors

(2.4)
$$f(x_1, \dots, x_k)$$
$$= \underset{1 \leqslant i \leqslant k}{\text{Max}} \left\{ \sum_{j=1}^{l} p_{ij} \left[c_{ij}\,x_i + f(x_1, \dots, c'_{ij}\,x_i, \dots x_k \right] \right\}$$

dont l'équation (2.2) est un cas particulier.

On déduit de (2.3)

$$\sum(\omega\,;s^*) \,\|\, f(\omega) = \underset{\delta_1}{\text{Max}} \left[g(\omega\,;\delta_1) + P(\delta_1)\,f(T_{\delta_1}\omega) \right]$$

la relation

(2.5)
$$f(\omega) = \underset{D_n}{\text{Max}} \left[g(\omega\,;D_n) + P(D_n)\,f(T_{D_n}\omega) \right]$$

où
$$D_n = \delta_1 . \delta_2 \dots . \delta_n$$

et
$$g(\omega\,;D_n) = g(\omega\,;\delta_1) + P(\delta_1)\,g(T_{\delta_1}\omega\,;\delta_2) + \dots$$
$$+ P(\delta_1) \dots P(\delta_{n-1})\,g(T_{D_{n-1}}\omega\,;\delta_n)$$
$$T_{D_n} = T_{\delta_n} \dots T_{\delta_1}$$

Cette relation exprimant l'invariance de $\sum(\omega , s^*)$ par rapport à l'échelle du temps, reflète un aspect intéressant du problème.

III - L'EXISTENCE ET L'UNICITÉ DE LA SOLUTION

Comme on l'a déjà vu, l'optimum doit être défini par

$$f(\omega) = \lim_{n \to \infty} f_n(\omega)$$

(3.1) $$f(\omega) = \lim_{n \to \infty} \operatorname{Max}_{\delta_1}\left[g(\omega ; \delta_1) + P(\delta_1) f_{n-1}(T_{\delta_1}\omega)\right]$$

ce qui équivaut à

(3.2) $$\sum (\omega ; S^*) = \lim_{n \to \infty} \sum\nolimits_n (\omega \ ; S_n^*)$$

On voit immédiatement que l'existence de l'optimum $f(\omega)$ équivaut à celle de la stratégie optimum S^*.

Il convient maintenant d'introduire la notion importante de troncature du processus ; on le fera au moyen de l'arbre représentatif de ce processus (Fig. 1). Chaque embranchement correspond à un des états possibles du système, le nombre des branches au nombre de décisions attachées à cet état.

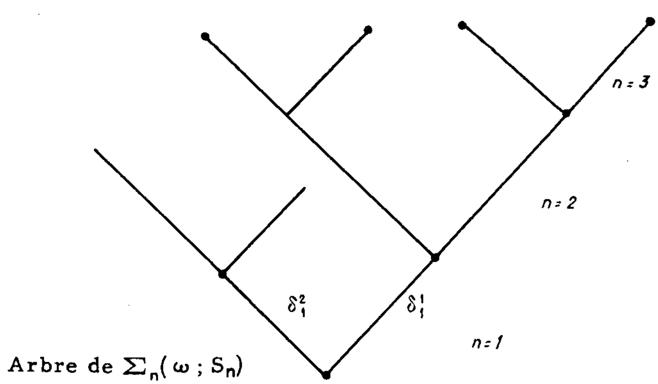

Fig. 1

Le processus de la Fig. 1 comporte deux décisions à l'instant n = 1, si bien que le système admet deux états possibles à l'instant n = 2. Si le système se trouve dans un état g à l'instant n, on peut retracer le processus en remontant dans le temps, c'est-à-dire en descendant par les branches consécutives jusqu'en bas. La suite $\{\delta_1 , \delta_2 , \dots \delta_{n-1}\}$ alors retrouvée constitue la stratégie par laquelle le système a été conduit à l'état g où il se trouve à l'instant n.

La troncature, dont on se servira dans la suite, signifie l'énumération de sous-processus à partir d'un certain instant n. Si l'on tronque le processus $\sum_n(\omega\,;S_n)$ de la Fig. 1 à l'instant n = 1, on obtient deux troncs dégradés :

$$\left\{\sum_{n-1}(T_{\delta_1^1}\omega\ ;S_{n-1}),\sum_{n-1}(T_{\delta_1^2}\omega\ ;S_{n-1})\right\}$$

THÉORÈME I

Si le nombre de décisions à chaque instant est fini le processus $\sum(\omega\,;S^*)$ admet une solution optimum unique $f(\omega)$ pour laquelle il existe une stratégie optimum S^*.

DÉMONSTRATION

1) Commençons par démontrer l'existence de $f(\omega)$, ce qui revient à démontrer, sous une autre forme, l'existence de S^*. A cet effet, nous recourons à un raisonnement par induction portant sur la durée n du processus.

Pour n = 0, le théorème est évident : il y a une stratégie optimum S_0^* qui indique de ne rien faire.

Pour n = 1

$$f_1(\omega) = \operatorname*{Max}_{\delta_1}\left\{g(\omega\,;\delta_1)\right\}$$

δ_1 étant un ensemble fini, il y a une stratégie δ_1^* pour laquelle

$$f_1(\omega) = g(\omega\,;\delta_1^*)$$

c'est-à-dire

$$\delta_1^* \in S_1^*$$

Supposons donc que le théorème soit vrai pour tous les processus de durée inférieure à m. Par la troncature à n = 1, on obtient k troncs

$$\sum_{m-1}(T_{\delta_1^i}\omega\ ;\ s_{m-1}^{i\,*})$$

$$i = 1,\ldots,k$$

chacun admet, par hypothèse, une stratégie optimum s_{m-1}^{i*} qui conduit à la solution optimum $f_{m-1}(\omega)$.

Posons :

$$F_m(\omega\,;\delta_1^i) = g(\omega\,;\delta_1^i) + f_{m-1}(T_{\delta_1^i}\omega\ ;\ s_{m-1}^{i\,*})$$

On constitue ainsi un ensemble fini

$$F_m(\omega\,;\delta_1^1)\ ,\quad F_m(\omega\,;\delta_1^2)\ ,\ldots,\quad F_m(\omega\,;\delta_1^k)$$

tel que, par définition,

$$F_m(\omega\,;\delta_1^*) = \operatorname*{Max}_{\delta_1}\ F_m(\omega\,;\delta_1)$$

On définit alors une stratégie S_m^* de la façon suivante : si l'embranchement g_1 correspond à l'état du système à l'instant n = 1,

$$S_m^*(g_1) = \delta_1^*$$

si g est un des embranchements de $\sum_{m-1}(T_{\delta_1^i}\omega \; ; \; s_{m-1}^{i*})$

$$S_m^*(g) = s_{m-1}^{i*}(g)$$

On obtient ainsi à partir des définitions ci-dessus,

$$\sum\nolimits_m (\omega \; ; S_m^*) \; ;$$

(3.3)
$$F_m(\omega \; ; \delta_1^*) = \underset{\delta_1}{\mathrm{Max}} \left[g(\omega \; ; \delta_1) + f_{m-1}(T_{\delta_1}\omega) \right]$$

$$= f_m(\omega)$$

ce qui établit l'existence d'une stratégie optimum S^* conduisant à une solution optimum $f(\omega)$ par un processus de durée quelconque.

2°) Démontrons maintenant l'unicité de la solution optimum. D'après (2.5)

$$f(\omega) = \underset{D_n}{\mathrm{Max}} \left[g(\omega \; ; D_n) + P(D_n)\, f(T_{D_n}\omega) \right]$$

$$= g(\omega \; ; D_n^*) + P(D_n^*)\, f(T_{D_n^*}\omega)$$

$$\geqslant g(\omega \; ; D'_n) + P(D'_n)\, f(T_{D'_n}\omega)$$

S'il y a une autre solution $K(\omega)$ de $\sum(\omega \; ; s^*)$ telle que

$$K(\omega) = \underset{D_n}{\mathrm{Max}} \left[g(\omega \; ; D_n) + P(D_n)\, K(T_{D_n}\omega) \right]$$

$$= g(\omega \; ; D'_n) + P(D'_n)\, K(T_{D'_n}\omega)$$

$$\geqslant g(\omega \; ; D_n^*) + P(D_n^*)\, K(T_{D_n}\omega)$$

on a :

$$P(D'_n)\, f(T_{D'_n}\omega) - P(D'_n)\, K(T_{D'_n}\omega) \leqslant f(\omega) - K(\omega)$$

$$\leqslant P(D_n^*)\, f(T_{D_n^*}\omega) - P(D_n^*)\, K(T_{D_n^*}\omega)$$

Soient :

$$f(T_{D_n^*}\omega) = a_n^* \quad , \qquad K(T_{D_n^*}\omega) = b_n^*$$

$$f(T_{D'_n}\omega) = a'_n \quad , \qquad K(T_{D'_n}\omega) = b'_n$$

les suites $\{a_n^*\}$ $\{b_n^*\}$ $\{a'_n\}$ $\{b'_n\}$

sont non croissantes par rapport à n. D'autre part :

$$P(D_n) \leqslant \alpha^n \qquad\qquad 0 < \alpha < 1$$

$$P(D'_n) \leqslant \alpha^n$$

Les inégalités précédentes prennent alors la forme

$$- \alpha^n(|a'_n|+|b'_n|) \leqslant f(\omega) - K(\omega) \leqslant \alpha^n(|a^*_n|+|b^*_n|)$$

Lorsque n tend vers l'infini, on obtient ainsi :

$$f(\omega) = K(\omega)$$

ce qui établit l'unicité de la solution optimum.

Nous allons établir ci-dessous quelques propriétés appartenant à notre processus.

THÉORÈME II

La suite $f_n(\omega)$ est monotone, non décroissante, et converge uniformément vers la solution optimum $f(\omega)$.

D'après (2.3), (2.5) et (3.3),

$$f_n(\omega) = \underset{\delta_1}{\mathrm{Max}} \left\{ g(\omega;\delta_1) + P(\delta_1) f_{n-1}(T_{\delta_1}\omega) \right\}$$

$$= \underset{D_n}{\mathrm{Max}} \left\{ g(\omega;D_n) = g(\omega;D_n^*) \right\};$$

$$f_{n+1}(\omega) = \underset{\delta_1}{\mathrm{Max}} \left\{ g(\omega;\delta_1) + P(\delta_1) f_n(T_{\delta_1}\omega) \right\};$$

$$f_{n+1}(\omega) = \underset{D_n}{\mathrm{Max}} \left\{ g(\omega;D_n) + P(D_n) f\ (T_{D_n}\omega) \right\}$$

$$\geqslant g(\omega;D_n^*) + P(D_n^*) f\ (T^*_{D_n}\omega)$$

$$\geqslant f_n(\omega)$$

La suite $f_n(\omega)$ est donc monotone et non décroissante. D'autre part $f_n(\omega)$ est une fonction continue de ω ; et elle est bornée par K ; la convergence uniforme de $f_n(\omega)$ vers $f(\omega)$ en résulte.

THÉORÈME III

Si la fonction $g(\omega;\delta_1)$ est convexe, la solution optimum $f(\omega)$ l'est aussi.

Par induction sur la durée n, on peut établir cette propriété pour toute fonction $f_n(\omega)$, et par suite pour $f(\omega)$.

THÉORÈME IV

Si k désigne une constante

$$k\,f(\omega) = f(k\omega)$$

En particulier,

$$f(0) = 0$$

L'égalité entre $kf(\omega)$ et $f(k\omega)$ vient du fait que les deux fonctions satisfont à la même équation.

IV - FONCTION DE DÉCISION

En raison du caractère présenté par le problème, il importe de déterminer la stratégie optimum S^* formée par la suite des décisions optimum à chaque instant, mais non pas de calculer la valeur de la solution optimum de $f(\omega)$.

Examinons d'abord le cas le plus simple, celui de la fonction $f(x,y)$, pour étudier la règle de décision, c'est-à-dire pour définir la fonction de décision à chaque instant.

Comme la fonction $f(x,y)$ est homogène, la fonction de décision ne dépendra que du rapport x/y, et sera représentée par une demi-droite séparatrice L passant par l'origine dans le premier quadrant x-y.

Considérons l'ensemble des points M définis de la façon suivante par leurs coordonnées m_1 et m_2 dans un espace à deux dimensions :

$$M = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} a_1 x + p_1 f(c_1 x, y) \\ a_2 y + p_2 f(x, c_2 y) \end{pmatrix}$$

avec $x+y = 1$.

La continuité, la convexité et la monotonie de $f(x,y)$ par rapport à ses variables x et y impliquent que l'ensemble des points M est constitué par une courbe continue qui est la borne inférieure de la contenue convexe M^* (convex hull) engendrée par M (fig. 2).

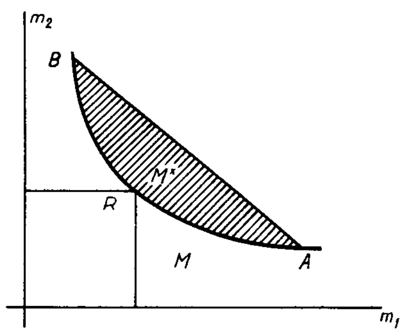

Fig. 2

Les deux extrémités A, b sont

$$A = \begin{pmatrix} a_1 + p_1 f(c_1, 0) \\ p_2 f(1, 0) \end{pmatrix}$$

$$B = \begin{pmatrix} p_1 f(0, 1) \\ a_2 + p_2 f(0, c_2) \end{pmatrix}$$

Il est bien clair (sinon le problème serait trivial) qu'il existe sur M un point R, dont les deux coordonnées sont égales, qui comporte deux stratégies optimum S_A^* , S_B^* , la première commence par l'opération A, l'autre par B. On en conclut que l'état du système correspondant au point R, admet deux stratégies optimum. A cause de l'intervention du coefficient c dans l'argument de f, B succède à A dans S_A^*. Il en est ainsi de S_B^* où B entraîne A au deuxième stade. Ainsi

$$(S_A^*) \qquad f_{AB}(x,y) = a_1 x + p_1 a_2 y + p_1 p_2 f(c_1 x, c_2 y)$$

$$(S_B^*) \qquad f_{BA}(x,y) = a_2 y + p_2 a_1 x + p_2 p_1 f(c_1 x, c_2 y)$$

égalisant les deux expressions, on arrive à l'équation L

$$\frac{a_1 x}{1-p_1} = \frac{a_2 y}{1-p_2}$$

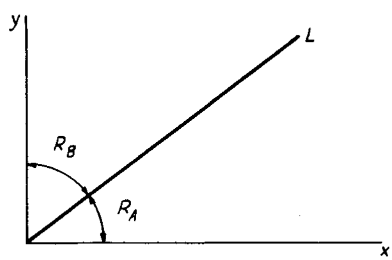

Fig. 3

Il suit de là que, dans la Fig. 3, les points situés au-dessous de L, définis par

R_A

$$\frac{a_1 x}{1-p_1} > \frac{a_2 y}{1-p_2}$$

exigent A pour la première décision optimum, et les points dans la région R_B exigent B.

Dans le cas général de la fonction de $f(x_1, \ldots x_k)$ soumise à l'égalité (2.4), le nombre de décisions est supérieur à deux. Il suffit alors de recourir à la troncature, et d'appliquer la méthode ci-dessus à tous les couples de décisions pour choisir entre les troncs obtenus. On doit ainsi chercher

$$\operatorname*{Max}_{1 \leq i \leq k} \frac{\sum_{j=1}^{l} p_{ij} \; c_{ij} \; x_i}{1 - \sum_{j=1}^{l} p_{ij}} = D(\omega \, ; i^*)$$

et prendre la décision (i^*) qui correspond à ce maximum.

Si p_{ij} est une répartition de probabilité, comme dans la première partie de cet article, le numérateur de l'expression précédente est l'espérance mathématique du gain immédiat, tandis que le dénominateur est la probabilité pour que le processus se termine par l'opération envisagée. Nous en déduisons la conclusion que la stratégie optimum pour un avenir lointain consiste à faire de son mieux à chaque instant.

BIBLIOGRAPHIE :

1. Richard BELLMAN : Dynamic programming. Princeton, 1957.
2. Samuel KARLIN : The structure of dynamic programming models. Naval Research Logistics Quarterly. 2 (1955), pp. 285-294.