Annales scientifiques de l'Université de Clermont-Ferrand 2
Série Mathématiques

ROBERT FAURE

Méthodes pratiques actuellement utilisables en programmation dynamique

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 8, série *Mathématiques*, n° 2 (1962), p. 63-74

<http://www.numdam.org/item?id=ASCFM_1962__8_2_63_0>

MÉTHODES PRATIQUES ACTUELLEMENT UTILISABLES EN PROGRAMMATION DYNAMIQUE

Robert FAURE
Conseiller scientifique à la Compagnie des machines Bull

(Conférence prononcée au cours de la journée des Constructeurs de Machines à Calculer)

INTRODUCTION -

Empêché, bien contre son gré, de venir à Clermont aujourd'hui, mon collègue et ami A. Kaufmann m'a demandé de le remplacer devant vous.

Lorsque, il y a peu de jours, j'ai connu le sujet qu'il me fallait aborder, j'avoue avoir été quelque peu surpris : pouvait-on faire état, en effet, d'une connexion quelconque entre l'oeuvre de Blaise Pascal et la *programmation dynamique* au sens de Bellman ?

Eh bien !, ce rapport qui m'échappait doit sans doute être recherché dans les liens qui ne cessèrent d'exister entre celui que nous honorons ici et l'un des plus fameux mathématiciens de tous les temps, Pierre de Fermat.

Ainsi, c'est en lisant l'*Historia trochoidis* de Pascal que nous apprenons que, dès 1638, J. de Beaugrand connaissait la méthode *de maximis et minimis* de Fermat(1). Or, il est incontestable que le principe de Fermat, selon lequel "les rayons lumineux sont les extrêmales du chemin optique"(2), donne l'une des premières expressions des théorèmes d'optimalité.

D'autre part, il suffit de lire les lettres de Pascal à Fermat pour se rendre compte de la part qu'a prise ce dernier dans l'invention de la *règle des partis*, dont il avait lui-même une solution. Le 29 juillet 1654, Pascal écrit : "*Je ne doute plus que je ne sois maintenant dans la vérité, après la rencontre admirable où je me trouve avec vous... Je vois bien que la vérité est la même à Toulouse qu'à Paris*". Le 26 août 1654 : "*Je recevrai votre réponse avec respect et avec joie, quand même votre sentiment me serait contraire*". Et, le 27 octobre 1654, ayant reconnu l'identité des résultats fournis par les deux méthodes(3) : "*Votre dernière lettre m'a parfaitement satisfait*". Il n'est pas moins évident que la programmation dynamique est également dominée par la règle des partis, chaque fois qu'elle s'applique à un univers aléatoire.

En quelque sorte, Pascal et Fermat possédaient déjà, à eux deux, les instruments fondamentaux qui ont servi à Bellman, trois cents ans après, à édifier sa théorie(4).

Il y a bien d'autres conjonctions entre les deux grands ancêtres de la programmation dynamique : outre qu'ils ont concouru à résoudre, par des méthodes variées, plusieurs problèmes, tel celui des carrés magiques, il est reconnu qu'ils ont eu, l'un et l'autre, parmi les premiers, la notion de démonstration par récurrence(5) et l'idée de l'élément différentiel(6).

----

(1) Les oeuvres de Fermat, *Varia opera mathematica*, n'ont été publiées qu'en 1679, c'est-à-dire quatorze ans après sa mort, par son fils Samuel. Quant à l'*Historia trochoidis*, dans ses textes latin et français, elle est datée du 10 décembre 1658.

(2) Ce n'est pas le lieu d'examiner ici pourquoi Fermat n'était pas tout à fait parvenu à cette généralité.

(3) "*La méthode de Pascal est d'une admirable simplicité et, en formant une équation aux différences finies, il invente une des deux méthodes analytiques du calcul des probabilités. L'autre méthode, qui repose sur la théorie des combinaisons, fut donnée dans le même temps par Fermat. La correspondance si curieuse entre ces deux grands esprits nous fait assister à la genèse des premiers principes du calcul des probabilités*". Emile Picard, Eloges et discours académiques, Gauthier-Villars, 1931.

(4) Richard Bellman : *Dynamic Programming*. Princeton University Press, 1957.

(5) Voir ce que Pascal écrit lui-même de la proposition XI du *Traité des ordres numériques* : "*Cette même proposition... est tombée dans la pensée de notre célèbre Conseiller de Toulouse, M. de Fermat,...*". Or la démonstration de la proposition suppose l'emploi de la récurrence.

(6) Dans sa lettre à Fermat, du 29. 7. 1654, Pascal dit du chevalier de Méré qu'"*il ne comprend pas qu'une ligne soit divisible à l'infini*" et ajoute : "*jamais je n'ai pu l'en tirer. Si vous pouviez le faire, on le rendrait parfait*".

Il n'est pas jusqu'à leurs visages qui n'aient une indéniable ressemblance(1) : ces mêmes yeux profonds, où transparaît l'étonnante sagesse de ceux qui parviennent aux sources de la science, la sagesse de l'homme étrange que Pascal tenait non seulement pour le plus grand géomètre de l'Europe, mais pour "*le premier homme du monde*"(2), la sagesse aussi de celui qui notait, en 1658 : "*Je ne fais pas de doute que ces règles, étant les véritables, ne doivent être simples, naïves, naturelles comme elles le sont. Ce ne sont pas barbara et baralipton qui forment le raisonnement. Il ne faut pas guinder l'esprit; les manières tendues et pénibles le remplissent d'une sotte présomption par une élévation étrangère et par une enflure vaine et ridicule au lieu d'une nourriture saine et vigoureuse*"(3).

LE PRINCIPE D'OPTIMALITE -

Du coup, nous voilà au coeur de notre sujet. Car ce sont de grandes, mais simples idées qui constituent le noyau des méthodes de programmation dynamique : *le principe d'optimalité.*

S'agissant de recherche opérationnelle, donc d'une politique économique optimale, Bellman exprime, en bref, ceci : "Toute politique optimale ne peut être formée que de sous-politiques optimales".

N'est-il pas pour ainsi dire évident que tout chemin optimal est formé de portions de chemins elles-mêmes optimales ? S'il n'en était pas ainsi pour une portion quelconque, celle-ci pourrait être remplacée par une autre meilleure, et, par conséquent, le chemin ne serait pas optimal, contrairement à l'hypothèse.

On peut même dire que le principe de Bellman prend ici l'allure d'un théorème, tout au moins lorsque le mot "chemin" possède un sens dépourvu d'ambiguité, relativement au phénomène envisagé.

R. Fortet ayant, au cours de ce même colloque, déjà traité des idées de Bellman d'une manière théorique, nous nous bornerons, au cours de cette intervention, à examiner quelques exemples pratiques et à indiquer, le cas échéant, quel secours les machines mathématiques peuvent apporter au traitement de tels problèmes.

On est généralement convenu de considérer que la programmation dynamique concerne l'évolution dans le temps d'un système économique, celle-ci étant, à chaque phase, partiellement aléatoire (intervention du hasard) et partiellement contrôlée (intervention de l'homme).

R. Fortet distingue les évolutions de première et seconde espèces, selon que, à chaque phase, l'intervention du hasard précède la décision humaine ou, au contraire, la suit.

Evidemment, les cas limites sont ceux, opposés :

- où le hasard disparaît (cas déterministe) ;
- où la décision humaine ne se manifeste pas (processus stochastique).

Mais certains caractères de la programmation dynamique, comme la division séquentielle du problème, sont conservés, même dans certains cas déterministes où l'évolution temporelle n'intervient pas.

LE CAS DETERMINISTE -

Imaginons que nous ayons à construire une autoroute entre les villes A et K, ainsi que l'indique la figure 1, les points intermédiaires de passage pouvant être classés en groupes (ou "phases") et les arcs du graphe ayant été valués par l'indication des coûts totaux (coûts de réalisation : chaussées, ouvrages d'art, etc. + coûts sociaux, de regret, etc.).

Dans ces conditions, la meilleure des politiques $F(x_0, \dots, x_4)$ est celle qui correspond au coût minimal, avec :

$$x_0 = \{A\}, \quad x_1 \in \{B, C, D\}, \quad x_2 \in \{E, F, G, H\}, \quad x_3 \in \{I, J\}, \quad x_4 = \{K\}.$$

---

(1) Comparer le portrait de Fermat par Poilly à celui de Pascal par Philippe de Champaigne.

(2) Lettre de Pascal à Fermat du 10 août 1668.

(3) *De l'Esprit géométrique*, Section II.

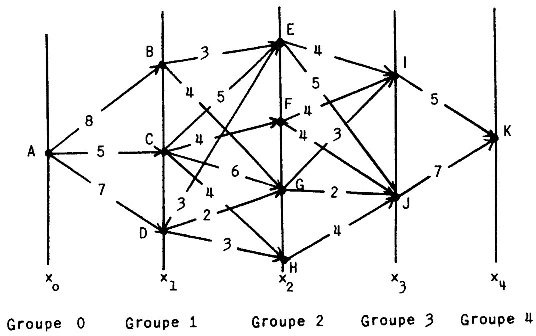

Fig. 1

On peut, par exemple, déterminer, pour la variable attachée à chaque phase de la progression sur le terrain, le chemin optimal depuis l'origine :

x_2	chemin optimal	coût
E	ACE, ADE	10
F	ACF	9
G	ADG	9
H	ACH	9

x_3	chemins possibles	optimal	coût
I	ACEI, ADEI, ACFI, ADGI,	ADGI	12
J	ACEJ, ADEJ, ACFJ, ADGJ, ACHJ	ADGJ	11

x_4	possibles	optimal	coût
K	ADGIK ADGJK	ADGIK	17

Dans ce problème, on aurait aussi bien pu commencer la résolution par la fin, ou encore regrouper des phases successives(1).

Comme il s'agit, en fait, de la recherche du chemin minimal dans un graphe, l'*algorithme de Ford* est applicable ; mieux, il est simplifiable, les points étant groupés par phases successives.

Considérons un programme dynamique déterministe, portant sur un nombre fini N de phases et tel que l'ensemble discret des valeurs possibles de la variable d'état, à chaque phase, soit également fini. Numérotons de O à M les sommets du graphe, dans l'ordre des phases, les sommets de la phase 1 portant les numéros 1 à α_1, ceux de la phase 2, $\alpha_1 + 1$ à α_2, et ainsi de suite jusqu'à $\alpha_N = M$. On sait (2) qu'on affecte d'abord le sommet x_o de l'indice $\lambda_o = 0$ et tous les autres sommets $x_i (i > 0)$ de l'indice $\lambda_i = +\infty$. Ensuite, partant du sommet x_o, on examine successive-

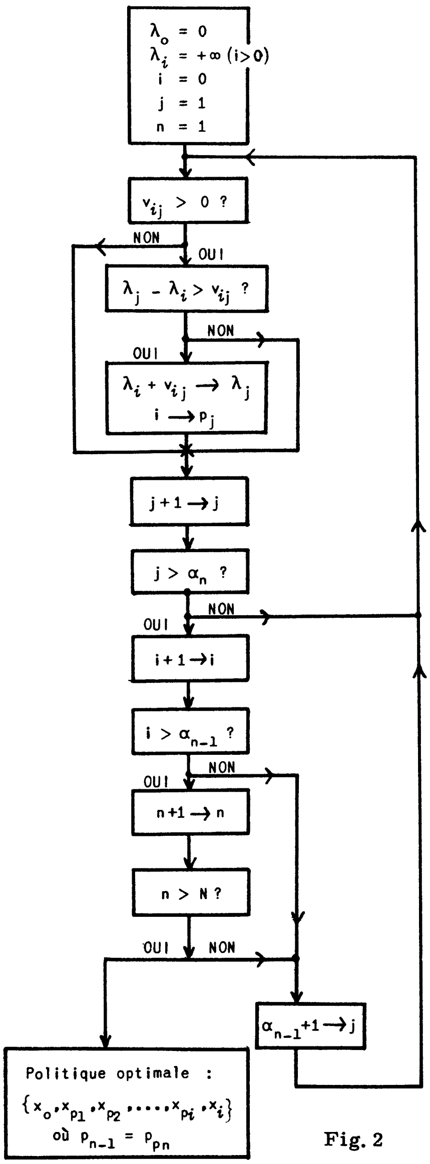

Fig. 2

----

(1) Le problème est dit non ordonné.

(2) Voir, par exemple : C. Berge, *Théorie des graphes et applications*, Dunod, 1958.

5

ment *tous les arcs* (x_i, x_j) tels que $\lambda_j - \lambda_i > v(x_i, x_j)$ et, à chaque fois, on remplace λ_j par $\lambda_i + v(x_i, x_j)$. Mais ici, pour $\alpha_{n-1} + 1 \leqslant i \leqslant \alpha_n$ (sommet de la phase n), les seuls sommets d'extrémité initiale x_i ont leur extrémité terminale en un sommet x_j de la phase n+1, telle que : $\alpha_n + 1 \leqslant j \leqslant \alpha_{n+1}$.

Dans ces conditions, l'organigramme et le processus de traitement(1) se simplifient (figure 2).

Il faut attirer l'attention sur le fait que l'application, ci-dessus, de l'algorithme de Ford ne concerne que les problèmes où les coûts ne sont pas actualisables. S'ils l'étaient, la solution pourrait être modifiée. Ainsi, imaginons que v(I,K) = 5,95, dans un problème qui se traduirait, pour tous les autres points, par le graphe de la figure 1, mais dans lequel les coûts seraient actualisables et choisissons un taux de 10 %.

On aurait les tableaux :

x_2	chemin optimal	coût
E	ACE	9,545
F	ACF	8,636
G	ADG	8,818
H	ACH	8,636

x_3	chemin optimal	coût
I	ADGI	11,296
J	ADGJ	10,470

x_4	chemin optimal	coût
K	ADGJK	15,727

Alors que le problème non actualisé comporte la solution ADGIK, avec un coût de 17,95 (contre 18 à ADGJK), le problème actualisé a comme solution ADGJK, avec un coût actualisé de 15,727 (contre 15,764 à ADGIK).

Il serait, bien entendu, aisé de modifier encore l'algorithme en remplaçant, à toute époque x, λ_j par $\lambda_i + \dfrac{v(x_i, x_j)}{(1+\alpha)^{n-1}}$ et non par $\lambda_i + v(x_i, x_j)$, lorsque $\lambda_j - \lambda_i > v(x_i, x_j)$.

En général, l'étude d'un programme dynamique déterministe s'étendant sur un nombre N de phases fini et non actualisable par nature (non temporel ou pouvant être considéré comme tel) conduit à des *formules d'optimisation séquentielle*, telles que :

$$f_{o,n}(x_o, x_n) = \underset{x_{n-1} \in X_{n-1}(x_o,\, x_n)}{\text{OPT}} \left[f_{o,n-1}(x_o, x_{n-1}) + v_n(x_{n-1}, x_n)\right], \quad n = 1, 2, \ldots, N,$$

avec :

$$f_{o,1}(x_o, x_1) = v_1(x_o, x_1).$$

Dans les problèmes non ordonnés, la décomposition en phases et l'optimisation peuvent se faire dans un sens quelconque ; dans les problèmes faiblement ordonnés, la décomposition et l'optimisation peuvent se faire selon un certain nombre déterminé d'ordres. Dans les problèmes fortement ordonnés, une seule décomposition et un seul ordre d'optimisation sont possibles.

Remarques.

1/ Les formules se modifient aisément lorsque les valeurs aux limites ne sont pas imposées, mais la difficulté des calculs peut être différente selon l'ordre d'optimisation choisie (s'il en existe plusieurs).

2/ Les valeurs pouvant être prises par les variables d'état ne sont pas nécessairement choisies dans des ensembles discrets. La formulation s'étend à des choix dans des ensembles continus mais bornés.

3/ La méthode a été présentée pour des fonctions économiques additives de phase en phase, mais elle est valable pour des fonctions multiplicatives :

$$F(x_o, x_1, \ldots, x_N) = v_1(x_o, x_1) \,.\, v_2(x_1, x_2) \,.\, \ldots \,.\, v_N(x_{N-1}, x_N)$$

ou des produits de composition, au sens de Borel :

$$F(x_o, x_1, \ldots, x_N) = v_1(x_o, x_1) * v_2(x_1, x_2) * \ldots * v_N(x_{N-1}, x_N).$$

4/ A l'heure actuelle, on ne peut que remarquer l'absence d'algorithme général pour les cas si divers qu'on peut envisager.

---

(1) A. Kaufmann et R. Cruon. *La méthode de programmation dynamique*, Dunod, à paraître.

INTRODUCTION D'UN HORIZON ILLIMITE -

Même en avenir déterminé, les problèmes de caractère économique se posent différemment lorsque l'horizon est limité ou, au contraire, illimité. On peut considérer, outre le cas d'un horizon fermé dans les deux sens, celui d'un horizon ouvert à gauche et fermé à droite, ou fermé à gauche et ouvert à droite, ou encore ouvert dans les deux sens.

Le cas le plus important, économiquement parlant, est évidemment celui d'un horizon fermé à gauche et illimité à droite, ce qui revient à faire tendre le nombre N de phases vers l'infini ($0 \leqslant N \leqslant +\infty$).

Bien entendu, il est rare que la fonction $F(x_o, x_1, \ldots, x_N)$ converge, pour $N \longrightarrow +\infty$, vers une limite finie ; en fait, ceci a lieu pour les programmes où le domaine de décision se rétrécit de phase en phase.

Dans ces conditions, le plus souvent, on a recours à une *actualisation* pour rendre la fonction économique convergente ; on calcule, par exemple :

$$\lim_{N \to \infty} F(x_o, \ldots, x_N) = v_1(x_o, x_1) + \frac{v_2(x_1, x_2)}{1+\alpha} + \ldots + \frac{v_N(x_{N-1}, x_N)}{(1+\alpha)^{N-1}}$$

et si la limite existe et est unique pour chaque politique, on recouvre un moyen de comparer différentes politiques.

Il existe aussi des problèmes où l'actualisation ne se justifie pas. On peut alors tenter de calculer la :

$$\lim_{N \to \infty} \frac{F(x_o, x_1, \ldots, x_N)}{N},$$

ce qui prend particulièrement un sens lorsqu'on a affaire à des sous-politiques cycliques, comme dans le cas d'un renouvellement d'équipements.

Enfin, la combinaison de l'actualisation et de la valeur moyenne par période peut devenir parfois significative.

Remarque. Un programme déterministe défini par les applications Γ_n :

$$\begin{array}{lll}
x_1 \in \Gamma_1 x_o & X_1(x_o, o) = \Gamma_1 x_o & \\
x_2 \in \Gamma_2 x_1 & X_2(x_o, o) = \Gamma_2 X_1 = \bigcup\limits_{x_1 \in X_1} \Gamma_2 x_1 & \\
\hline
x_n \in \Gamma_n x_{n-1} & X_n(x_o, o) = \Gamma_n X_{n-1} = \bigcup\limits_{x_{n-1} \in X_{n-1}} \Gamma_n x_{n-1} &
\end{array}$$

est *stationnaire* si :

$$\forall_n \quad \Gamma_n x_n = \Gamma x_n$$

et :

$$v_n(x_{n-1}, x_n) = v(x_{n-1}, x_n) ;$$

il est *semi-stationnaire* si seule la première de ces deux conditions est vérifiée.

Les théorèmes de convergence et d'unicité sont beaucoup plus simples dans le cas des programmes stationnaires ou même semi-stationnaires.

LE CAS ALEATOIRE -

Revenons maintenant au cas le plus courant en programmation dynamique, dans lequel le processus n'est que partiellement contrôlé par l'homme.

Il est bien clair que c'est ici que nous deviendra utile la notion pascalienne d'*espérance mathématique*.

Prenons, comme exemple, très simple, un jeu dans lequel chaque partie peut se décrire ainsi : un joueur, sachant au départ du jeu la case où il débute, reçoit en phase I l'autorisation

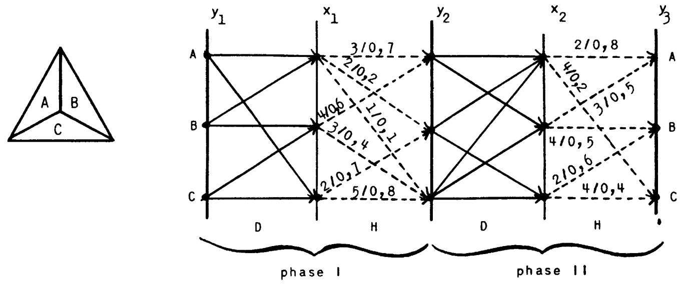

Fig. 3

de choisir la case où il désire aller, le hasard intervenant alors pour lui imposer une nouvelle case, en assortissant ce tirage d'un gain déterminé. En phase II, nouvelle décision du joueur, puis nouveau tirage au hasard et nouveau gain aléatoire.

Recherchons quelles seront les politiques optimales du joueur dans un jeu comprenant de nombreuses parties.

La figure 3 donne tous les renseignements sur la règle de ce jeu de type DH (décision-hasard). A chaque phase de décision, les autorisations de déplacement volontaire sont indiquées en traits pleins ; à chaque tirage au sort, les probabilités de transition et les gains correspondants sont notés sur des traits ponctués.

Nous avons des variables de *décision* : x_1 et x_2, et des variables de *position* : y_1, y_2 et y_3.

Appelons $\bar{Z}_2(y_2, x_2)$ l'espérance mathématique du joueur qui, se trouvant en y_2, décide d'aller en x_2 :

$$\bar{Z}_2(A,A) = 2 \times 0,8 + 4 \times 0,2 = 2,4$$

$$\bar{Z}_2(A,B) = 3 \times 0,5 + 4 \times 0,5 = \boxed{3,5}$$

$$\bar{Z}_2(B,A) = 2,4$$

$$\bar{Z}_2(B,C) = 2 \times 0,6 + 4 \times 0,4 = \boxed{2,8}$$

$$\bar{Z}_2(C,A) = 2,4$$

$$\bar{Z}_2(C,B) = \boxed{3,5}$$

$$\bar{Z}_2(C,C) = 2,8$$

L'espérance mathématique maximale indique, dans chaque cas, le choix optimal :

$$\bar{f}_2(A) = 3,5 \quad \text{avec} \quad x_2 = B$$
$$\bar{f}_2(B) = 2,8 \quad \text{avec} \quad x_2 = C$$
$$\bar{f}_2(C) = 3,5 \quad \text{avec} \quad x_2 = B.$$

Calculons, de la même manière, l'espérance $\bar{Z}_1(y_1, x_1)$:

$$\bar{Z}_1(A,A) = (3 + 3,5) \times 0,7 + (2 + 2,8) \times 0,2 + (1 + 3,5) \times 0,1 = 5,96$$

$$\bar{Z}_1(A,C) = (2 + 2,8) \times 0,2 + (5 + 3,5) \times 0,8 = \boxed{7,76}$$

$$\bar{Z}_1(B,A) = 5,96$$

$$\bar{Z}_1(B,B) = (4 + 3,5) \times 0,6 + (3 + 3,5) \times 0,4 = \boxed{7,10}$$

$$Z_1(C,B) = 7,10$$

$$Z_1(C,C) = \boxed{7,76}$$

Nous avons donc, si :

$$y_1 = A \ldots\ldots\ldots F_{MAX}(A) = 7{,}76 \quad \text{avec} \quad x_1 = C$$
$$y_1 = B \ldots\ldots\ldots F_{MAX}(B) = 7{,}10 \quad \text{avec} \quad x_1 = B$$
$$y_1 = C \ldots\ldots\ldots F_{MAX}(C) = 7{,}76 \quad \text{avec} \quad x_1 = C.$$

De sorte que les politiques optimales seront les suivantes :

En phase I		En phase II	
Si le joueur se trouve en	Il choisit	Si le joueur se trouve en	Il choisit
A B C	C B C	A B C	B C B

La première remarque qui s'impose, c'est que, dans cet exemple, nous avons dû rechercher les politiques optimales, par le critère de la valeur moyenne, *en remontant* du futur au passé. Il en est toujours ainsi qu'il soit question d'évolution de première ou seconde espèce.

Le théorème d'optimalité, relatif à un phénomène s'étendant sur N périodes s'énonce alors : "Une sous-politique optimale de N à N - n ne peut être formée que par une sous-politique optimale de N à N - n + 1".

On peut, d'autre part, considérer en avenir aléatoire :

a) des programmes discrets ;

b) des programmes continus ;

et, dans chacun des cas précédents :

1/ un horizon limité ;

2/ un horizon illimité.

De nombreux théorèmes de convergence et d'unicité ont été trouvés pour les situations les plus diverses et des critères plus ou moins adaptés.

Comme nous avons surtout en vue les applications au calcul automatique, nous examinerons le cas fréquent pour lequel on dispose de méthodes de traitement.

Si le processus stochastique de décision est markovien et discret, à tout changement d'état, $E_i \longrightarrow E_j$, $i, j = 1, 2, \ldots, M$, on peut faire correspondre une probabilité $p_{ij}^{(r)}$. Pour chaque état E_i, on a un ensemble m_i de *vecteurs stochastiques* $p_i^{(r)} = [p_{i_1}^{(r)}\ p_{i_2}^{(r)} \ldots p_{i_M}^{(r)}]$, le choix de r étant libre. A chaque probabilité $p_{ij}^{(r)}$ est associée une valeur du revenu $R_{ij}^{(r)}$, d'où des *vecteurs de revenu*, $R_i^{(r)} = [R_{i_1}^{(r)}\ R_{i_2}^{(r)} \ldots R_{i_M}^{(r)}]$.

Soit $\bar{q}_i^{(r)}$ l'espérance mathématique de la valeur d'une phase lorsqu'étant dans l'état E_i à la date x, on a choisi le vecteur stochastique $p_i^{(r)}$; on a :

$$\bar{q}_i^{(r)} = \sum_{j=1}^{M} p_{ij}^{(r)} R_{ij}^{(r)} \qquad i = 1, 2, \ldots, M.$$

Par suite :

$$\bar{v}_i(N-r, N) = \underset{r=1,2,\ldots,m}{\mathrm{MAX}} \left[\bar{q}_i^{(r)} + \sum_{j=1}^{M} p_{ij}^{(r)}\, v_j(N-n+1, N) \right],$$

si $\bar{v}_i(N-n, N)$ est l'espérance mathématique de la valeur totale sur n transitions de N - n à N et si à N - n le système est dans l'état E_i.

D'où les calculs successifs de :

$$\bar{v}_i(N-1,N) = \max_{r=1,2,\dots,m}\left[\bar{q}_i^{(r)} + \sum_{j=1}^{M} p_{ij}^{(r)}\,\bar{v}_j(N,N)\right]$$

$$\bar{v}_i(O,N) = \max_{r=1,2,\dots,m}\left[\bar{q}_i^{(r)} + \sum_{j=1}^{M} p_{ij}^{(r)}\,\bar{v}_j(1,N)\right]$$

Si toutes les matrices $\mathfrak{M}^{(r)}$ formées par M vecteurs quelconques $[p_i^{(r)}]$ sont complètement ergodiques, pour chaque décision r, lorsqu'on est dans l'état E_i à la date N - n, soit k + 1 périodes avant N, et si k est suffisamment grand, on montre que :

$$\bar{v}_i^{(r)}(k) = k\gamma + w_i^{(r)} \qquad i = 1,2,\dots,M,$$

avec :

$$\gamma = \sum_{i=1}^{M} p_i\bar{q}_i,$$

où $p_i (i = 1,2,\dots,M)$ sont les éléments des lignes de $[\tilde{\mathfrak{M}}]$, toutes identiques(1).

On posera donc :

$$\gamma + w_i^{(r)} = \max_{r=1,2,\dots,m}\left[\bar{q}_i^{(r)} + \sum_{j=1}^{M} p_{ij}^{(r)}\,w_j^{(r)}\right]$$

et, pour éviter des calculs trop volumineux, on pourra appliquer l'*algorithme itératif* de Howard(2).

Celui-ci peut se résumer comme suit :

Opération 1. Choisir r quelconque et résoudre le système linéaire

$$\gamma + w_i = \bar{q}_i + \sum_{j=1}^{M} p_{ij}\,w_j \qquad i = 1,2,\dots,M$$

avec les $\bar{q}_i$ et p_{ij} correspondants à r et w_1 par exemple égal à 0.

Opération 2. Avec les valeurs $w_1, w_2, \dots, w_M$ obtenues lors de l'opération 1, évaluer, pour toutes les décisions admissibles :

$$\bar{Z}_i^{(r)} = \bar{q}_i^{(r)} + \sum_{j=1}^{M} p_{ij}^{(r)}\,w_j \qquad \begin{matrix} i = 1,2,\dots,M \\ r = 1,2,\dots,m. \end{matrix}$$

Opération 3. Pour chaque état E_i choisir :

$$Z_i^{(r_i')} = \max_{r=1,2,\dots,m}[Z_i^{(r_i)}].$$

Opération 4. Si, remplaçant dans (1) $\bar{q}_i$ par $\bar{q}_i^{(r_i')}$ et p_{ij} par $p_{ij}^{(r_i')}$, on a $\gamma' = \gamma$, on a terminé.

Sinon, on reprend l'opération 2.

Remarque. L'optimisation par itération peut être modifiée et étendue au cas où la chaîne n'est pas complètement ergodique.

On fonde beaucoup d'espoir sur les études entreprises à la Faculté de Grenoble par Ramon Companys pour décomposer les matrices de décision en matrices booléiennes et aussi pour trouver les conditions d'application d'un algorithme récurrent. R. Companys s'est particulièrement intéressé aux processus HD relatifs à la gestion des équipements industriels complexes, comportant diverses machines sujettes à pannes et susceptibles d'entretien préventif.

----

(1) On appelle $[\tilde{\mathfrak{M}}]$ la $\lim_{n\to\infty} [\mathfrak{M}]^{(n)}$.

(2) R. A. Howard. *Dynamic programming and Markovprocess*. Ed. Wiley, 1960.

INTERVALLE D'ANTICIPATION - PROCESSUS ADAPTATIFS -

Pour déterminer ce court exposé, nous désirerions éviter de passer sous silence les processus adaptatifs.

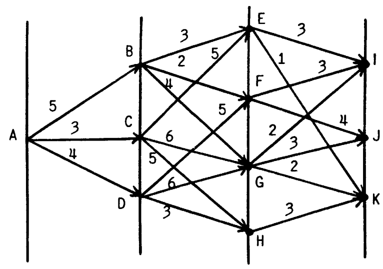

Fig. 4

Dans le *cas déterministe* représenté par la figure 4, si l'intervalle ϑ d'anticipation n'est que d'une phase, le meilleur chemin est AC(3) ; si ϑ s'étend sur deux phases, ABF et ADH sont minimaux équivalents avec un coût de 7 ; enfin, si l'intervalle d'anticipation est de trois phases, le problème peut être résolu complètement : ABEK est le meilleur chemin, avec un coût de 9.

Naturellement, plus ϑ est grand, meilleur est le chemin choisi, le coût de celui-ci devant décroître monotonement lorsque ϑ augmente d'une unité.

En réalité, il est rare que l'avenir soit connu formellement. Et, lorsque les phases sont temporelles, non seulement on ne connaît l'évolution qu'en *probabilité*, mais encore cette connaissance est, en général, d'autant plus floue qu'on anticipe sur une époque plus lointaine.

La situation pourrait être dépeinte en admettant qu'un chien cherche à rattraper son maître parcourant une certaine route (C). La courbe décrite par un chien qui ne relèverait la tête que tous les 200 mètres serait du type ABCDEFGHIJK = 2 000 m. Celle d'un chien intelligent qui, dans les mêmes conditions, anticiperait d'une période sur la future position de son maître serait A'B'C'D' E'F'G' = 1 200 m.

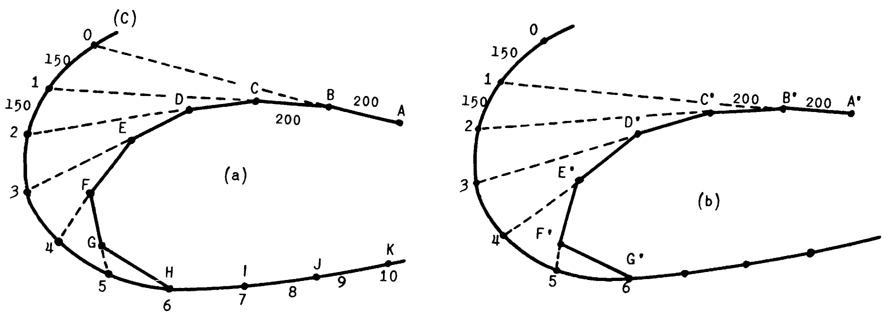

Fig. 5

Mille variantes sont permises : maître dont la trajectoire et la vitesse se modifient, chien dont la faculté d'anticipation est plus ou moins grande et qui reçoit une information discrète ou continue, etc.

L'idée essentielle est ici la *réadaptation* qui a lieu à chaque phase.

Un tout petit exemple, se rapportant à un système économique pouvant prendre seulement deux états E_1 et E_2, selon des probabilités de transition inconnues, et pour lequel existent deux décisions D_1 et D_2 entraînant des coûts déterminés, fait l'objet du graphe de la figure 6 et du tableau ci-après.

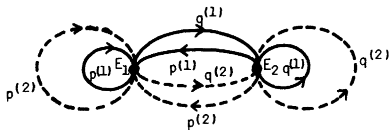

Fig. 6

Transition	Décision D_1		Décision D_2	
	prob.	coût	prob.	coût
$E_1 \longrightarrow E_1$	$p^{(1)}$	5	$p^{(2)}$	10
$E_1 \longrightarrow E_2$	$q^{(1)}$	8	$q^{(2)}$	5
$E_2 \longrightarrow E_1$	$p^{(1)}$	7	$p^{(2)}$	4
$E_2 \longrightarrow E_2$	$q^{(1)}$	6	$q^{(2)}$	8

On a :

$$q^{(1)} = 1 - p^{(1)}$$

et :

$$q^{(2)} = 1 - p^{(2)}.$$

On désire, en admettant que le processus soit stationnaire et que le problème se rapporte à un grand nombre de périodes, trouver la politique optimale.

N'ayant, à l'origine, aucun renseignement sur les valeurs de $p^{(1)}$ et $p^{(2)}$, on peut toujours imaginer que :

$$p^{(1)} = p^{(2)} = q^{(1)} = q^{(2)} = \frac{1}{2}$$

et l'on en déduit facilement quė, si l'on se trouve en E_1, il faut choisir la décision D_1, et, si l'on se trouve en E_2, la décision D_2.

En effet :

$$E_1 \begin{cases} \text{espérance mathématique décision } D_1 : & 5 \times \frac{1}{2} + 8 \times \frac{1}{2} = \boxed{6,5} \\ \text{espérance mathématique décision } D_2 : & 10 \times \frac{1}{2} + 5 \times \frac{1}{2} = 7,5 \end{cases}$$

$$E_2 \begin{cases} \text{espérance mathématique décision } D_1 : & 7 \times \frac{1}{2} + 6 \times \frac{1}{2} = 6,5 \\ \text{espérance mathématique décision } D_2 : & 4 \times \frac{1}{2} + 8 \times \frac{1}{2} = \boxed{6}. \end{cases}$$

Admettons maintenant que, nous étant trouvés en E_1, nous ayons choisi, d'après ce qui précède, la décision D_1, et que nous nous trouvions aussitôt en E_2.

Selon l'hypothèse de Laplace-Bayes, $q^{(2)}$ et $p^{(2)}$ demeurant inchangés, nous pouvons réévaluer $q^{(1)}$ et $p^{(1)}$:

$$q^{(1)} = \frac{1+1}{1+2} = \frac{2}{3} \quad , \quad p^{(1)} = 1 - \frac{2}{3} = \frac{1}{3},$$

ce qui nous fournit :

position	décision	espérance
E_1	D_1 D_2	$\boxed{7}$ 7,5
E_2	D_1 D_2	6,33 $\boxed{6}$

Il n'y a pas de modification de notre politique, aussi choisissons-nous D_2. A la phase suivante, nous nous trouvons, par exemple, en E_1. Nous pouvons alors réévaluer $p^{(2)}$ et $q^{(2)}$:

$$p^{(2)} = \frac{1+1}{1+2} = \frac{2}{3} \quad , \quad q^{(2)} = \frac{1}{3}$$

et les coûts moyens les plus faibles correspondent encore à D_1 pour $E_1(7)$ et à D_2 pour $E_2(5,33)$.

Dans ces conditions, nous choisirons D_1. Si l'état E_2 se présente alors, nous calculerons :

$$q^{(1)} = \frac{2+1}{2+2} = \frac{3}{4} \quad \text{et} \quad p^{(1)} = \frac{1}{4}$$

ce qui nous amènera à choisir toujours D_1, que nous soyons en E_1 (coût : 7,25) ou en E_2 (coût : 5,25).

Il nous faut alors préférer D_1 et, si nous trouvons en E_2 à la phase suivante, nous avons :

$$q^{(1)} = \frac{3+1}{3+2} = \frac{4}{5} \quad \text{et} \quad p^{(1)} = \frac{1}{5} ;$$

nous sommes ramenés à D_1 pour E_1 (coût : 7,40) et D_2 pour E_2 (coût : 5,33).

Partant de E_2, et adoptant D_2, nous nous trouverons, par exemple, en E_2 et nous en déduirons :

$$q^{(2)} = \frac{2+1}{2+2} = \frac{3}{4} \quad , \quad p^{(2)} = \frac{1}{4},$$

d'où la politique :

D_2 si E_1 (coût : 5,25)

D_1 si E_2 (coût : 6,20).

Et ainsi de suite.

Au bout d'un nombre suffisant de phases, nous aboutissons, par exemple, aux évaluations suivantes :

$$q^{(1)} = 0,78 \quad , \quad p^{(1)} = 0,22$$
$$q^{(2)} = 0,80 \quad , \quad p^{(2)} = 0,20,$$

lesquelles nous permettent de définir la politique optimale permanente :

D_2 si E_1 (coût : 6)

D_1 si E_2 (coût : 6,22)

Le jeu qui nous a occupés pourraît décrire les ordres à terme en bourse sur une catégorie de valeurs en hausse ou en baisse.

Plus généralement, les processus adaptatifs, dont la théorie reste à faire, ont déjà été étudiés par Bellman, Vaszonyi, etc.

CONCLUSION -

Les méthodes de programmation dynamique conviennent parfaitement, on le voit, aux processus de nature *séquentielle*. C'est dire qu'elles ont leur application, en ce qui concerne l'économie, dans de nombreux problèmes relatifs aux investissements, aux stocks, au réapprovisionnement et à l'entretien des équipements.

En passant, elles ont le grand mérite de mettre en lumière la notion d'*horizon économique*, si fondamentale dans ces sortes de questions.

D'autre part, le célèbre problème de l'âne, à qui l'on se propose de faire parcourir la distance maximale, étant données sa charge maximale, sa consommation kilométrique et une certaine réserve de nourriture (supérieure à la charge admissible) en un point initial, est encore justiciable de la programmation dynamique. La recherche des étapes optimales n'est pas sans rapport avec celle du meilleur étagement des fusées : une récréation mathématique très ancienne est revenue à l'ordre du jour, mais la richesse des procédés de résolution est désormais plus grande.

Enfin, l'étude des *processus adaptatifs* n'en est encore qu'à ses tout débuts et pourtant semble déjà prometteuse de résultats intéressants.

Pour l'application des méthodes de programmation dynamique, nous disposons déjà de quelques algorithmes, il est vrai rarement généraux, mais sans doute les progrès de la théorie faciliteront-ils la simplification et l'unification des procédés d'emploi.

Sous la plume de Jacques Chevalier (1), on peut dire que Blaise Pascal méprisait, entre autres, "*les algorithmes que les hommes ont inventés pour l'économie de la pensée*".

On peut mettre en doute la généralité de ce jugement, si l'on veut bien considérer que plusieurs de ses oeuvres scientifiques - ne serait-ce que celle sur les carrés magiques (De numéricarum potestatum ambitibus) - ont abouti à des méthodes *pratiques* de résolution, c'est-à-dire, au sens que nous accordons à ce mot, à de véritables algorithmes (2).

Quoi qu'il en soit, c'est bien aux génies de Pascal et de son grand ami Fermat, aux inventeurs de la notion d'espérance mathématique et de celle d'optimalité, que nous devons les fondements des méthodes variées, désormais en pleine évolution, de la programmation dynamique.

----

(1) Pascal. *Oeuvres complètes*, Préface. Editions de la Pléiade, Gallimard, 1957.

(2) *Oeuvres de Messire A.Arnauld*, Paris-Lausanne, 1781. La partie relative aux carrés magiques, et notamment la "méthode pour disposer magiquement le carré naturel", est reproduite dans les *Oeuvres complètes* de Pascal, éd. de la Pléiade.