REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE

A. KAUFMANN
R. CRUON

Un tour d'horizon sur la programmation dynamique et ses applications

Revue française d'informatique et de recherche opérationnelle, tome 2, n° V3 (1968), p. 29-35

<http://www.numdam.org/item?id=RO_1968__2_3_29_0>

© AFCET, 1968, tous droits réservés.

L'accès aux archives de la revue « Revue française d'informatique et de recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

R.I.R.O.
(2e année, nº 14, 1968, p. 29-36)

UN TOUR D'HORIZON SUR LA PROGRAMMATION DYNAMIQUE ET SES APPLICATIONS (1)

par A. KAUFMANN (2) et R. CRUON (3)

1. — INTRODUCTION

L'expression de « programmation dynamique » désigne à la fois un type de modèle mathématique, et un algorithme de calcul adapté à ce type de problème. Nous nous proposons ici, après avoir caractérisé schématiquement la structure des programmes dynamiques et le principe de la méthode de résolution, de donner quelques indications sur les progrès et extensions réalisés depuis une quinzaine d'années au plan de la théorie, et sur les domaines d'application. On pourra également se reporter à l'article introductif de R. Howard [17], ainsi qu'à R.E. Larson [23] pour les méthodes de calcul.

Pour éviter d'introduire des notations mathématiques complexes, qui alourdiraient par trop l'exposé, notre description sera basée sur des concepts « physiques » plutôt que mathématiques.

Considérons un système, caractérisé à chaque instant $t \in T$ (où T est un ensemble donné, continu ou discret) par une variable d'état e_t définie dans un espace convenable. Ce système est susceptible d'évoluer sous l'influence de facteurs extérieurs, et décrit une « trajectoire ». En général, l'état initial e_0, l'horizon T et certaines conditions sur l'état final e_T sont données (4).

L'évolution du système peut être sous la dépendance d'une ou plusieurs volontés conscientes, agissant par l'intermédiaire de paramètres

(1) Exposé de synthèse présenté au Congrès européen de l'IMS-TIMS-ES-IASPS Amsterdam, 2-7 septembre 1968.

(2) Professeur à l'Institut polytechnique de Grenoble, Conseiller scientifique à la Cie Bull General Electric.

(3) Directeur de la Division Études au Centre Interarmées de Recherche Opérationnelle.

(4) Nous ne discuterons pas ici les diverses formulations auxquelles peut donner lieu la spécification des conditions aux limites.

de « commande » (« controls » en anglais) selon la terminologie du contrôle optimal [28] ou de « variables de décision » selon la terminologie de la programmation dynamique [3]. Les facteurs extérieurs, ou certains d'entre eux, peuvent également être aléatoires.

Le cas de systèmes soumis uniquement à des facteurs aléatoires est du ressort de la théorie des processus stochastiques. On sait l'importance que joue dans cette théorie la propriété de Markov. Parmi les processus stochastiques, les processus markoviens jouissent de propriétés remarquables. Rappelons, très schématiquement, qu'un processus est de Markov si l'évolution future du système, à partir de l'instant t, ne dépend (en probabilité) de l'histoire du processus jusqu'à t que par l'intermédiaire de l'état présent e_t.

2. — PROGRAMMES DYNAMIQUES DETERMINISTES

On peut mettre en parallèle avec le cas d'un système à évolution purement aléatoire, celui d'un système régi par ce qu'on peut appeler un « processus décisionnel », c'est-à-dire d'un système dont l'évolution est *déterminée* uniquement par la suite des valeurs données à une variable de décision. Parmi ces processus décisionnels, nous nous intéresserons particulièrement à ceux qui possèdent une propriété analogue à la propriété markovienne dans les processus stochastiques. Cette propriété se traduira par deux conditions. La première est que l'évolution future du système à partir de l'instant t ne dépende que de l'état présent e_t et des décisions prises à partir de l'instant t ; ceci implique que l'ensemble des décisions possibles à l'instant t (en général, la décision est soumise à un certain nombre de contraintes, qui définissent un ensemble de décisions possibles) ne dépend que de l'état présent e_t, et non de la façon dont le système y est parvenu.

La seconde condition est relative à la « fonction d'évaluation » attachée au système. Parler de décision suppose, en effet, qu'on ait défini un critère de décision. Nous supposerons donc donnée une application de l'ensemble des trajectoires possibles du système sur la droite numérique ; cette fonction d'évaluation définit un ordre total dans l'ensemble des trajectoires, c'est-à-dire qu'elle indique si une trajectoire est préférable ou non à une autre. L'objectif du décideur est alors de maximiser (ou de minimiser) la valeur associée à la trajectoire qu'il impose au système. La condition évoquée plus haut est que, le système se trouvant dans l'état e_t à l'instant (actuel) t, la trajectoire future optimale soit la même quelle que soit la façon dont le système est parvenu à e_t. Cette condition est remplie notamment (1) si la fonction d'évaluation est définie comme une intégrale calculée le long de la trajectoire (2).

(1) Voir [20], chapitre 5, au sujet des extensions possibles.

(2) De façon plus précise, les deux conditions peuvent s'exprimer ainsi : soient A et A' deux portions de trajectoire définies sur l'ensemble $\{\tau/\tau \in T,\ \tau \leqslant t\}$, et aboutissant à e_t, et B une portion de trajectoire définie sur l'ensemble

$$\{\tau/\tau \in T, \tau \geqslant t\},$$

Ces deux conditions assurent la validité d'un « principe d'optimalité » [3] que nous énoncerons sous la forme suivante :

« Si une trajectoire est optimale et passe par le point (τ, e_τ), la portion de cette trajectoire correspondant à $t > \tau$ est elle-même optimale. »

Ce principe d'optimalité donne lieu à des théorèmes qu'on démontre aisément par l'absurde, dans chacun des cas que nous évoquerons ci-dessous, et qui satisfont les conditions générales que nous avons indiquées schématiquement. Il fournit une condition nécessaire d'optimalité qui permet de déterminer la trajectoire optimale sans qu'il soit nécessaire d'énumérer toutes les trajectoires. On y parvient en considérant le problème initialement posé comme faisant partie d'une famille de problèmes, dont chacun est caractérisé par les paramètres t et e_t.

L'expression mathématique du problème diffère sensiblement selon que l'évolution temporelle du système est discrète ou continue. Le caractère dénombrable ou non de l'ensemble des états possibles à un instant donné joue également un rôle, bien que moins important.

Le cas d'un ensemble d'états dénombrable (ou même fini) et d'une évolution temporelle discrète (T dénombrable, par exemple $T = \{0, 1, 2, \ldots\}$) est de loin le plus fréquent dans les applications qui ont été faites de la programmation dynamique. Les nécessités du calcul numérique obligent d'ailleurs souvent à s'y ramener. Le problème est alors celui de la recherche du plus court chemin dans le graphe des évolutions possibles du système, à chaque arc duquel on associe la valeur indiquée par la fonction d'évaluation. L'algorithme de la programmation dynamique est l'une des nombreuses variantes de l'algorithme de marquage classiquement utilisé dans ce problème du plus court chemin, avec cette particularité que la structure « séquentielle » du graphe permet de n'examiner qu'une seule fois chacun de ses sommets (voir [20], p. 21). Un rapprochement est également à faire avec la théorie des programmes mathématiques. La programmation dynamique permet, en effet, de trouver l'optimum global des programmes mathématiques qu'on peut appeler « séquentiels », c'est-à-dire qui ont la structure indiquée plus haut, mais qui ne sont soumis à aucune condition de linéarité, ni même de convexité ; de plus, la condition que les variables du programme (c'est-à-dire les variables de décision dans la terminologie de la programmation dynamique) soient à valeurs entières n'est pas gênante, bien au contraire. De telles contraintes limitent en effet le nombre des trajectoires à examiner. C'est une propriété que la programmation dynamique possède en commun avec la méthode SEP (Séparation et Évaluation progressives : branch and bound) ; toutes deux rentrent dans la catégorie générale des algorithmes

---

et partant de e_t ; soit $\mathfrak{T}_A$ l'ensemble des trajectoires (définies sur T) contenant A ; enfin, soit (A, B) la trajectoire formée par A et B. Alors, on doit avoir, pour tout A, A', et B répondant aux définitions ci-dessus :

$$(A, B) \in \mathfrak{T}_A \Leftrightarrow (A', B) \in \mathfrak{T}_{A'}$$

$$(A, B) \text{ optimal dans } \mathfrak{T}_A \Leftrightarrow (A', B) \text{ optimal dans } \mathfrak{T}_{A'}$$

qui procèdent par « tamisage » ou « filtrage », en restreignant de façon séquentielle l'ensemble des solutions qui peuvent être optimales [15].

Dans la pratique, la principale limitation aux applications de la programmation dynamique concerne le nombre d'états possibles qu'on peut prendre en compte [5]. La programmation dynamique conduit, en effet, à des calculs assez lourds, qui peuvent cependant être exécutés sur ordinateur en des temps très acceptables, à condition que les résultats de calcul intermédiaires puissent être conservés en mémoire centrale ; or le nombre de ces résultats est proportionnel au nombre d'états possibles du système à un instant donné. Pour la plupart des ordinateurs actuels, le nombre d'états possibles à un instant donné ne peut pas dépasser quelques dizaines de milliers. Différentes méthodes peuvent être utilisées pour remédier partiellement à ces inconvénients (voir par exemple [23]).

Dans le cas d'une évolution temporelle continue, la programmation dynamique a des connexions profondes avec le calcul des variations [12] et avec le principe du maximum de Pontryagin [28] [6]. C'est cette dernière approche qui semble actuellement la plus fructueuse (1). Il faut reconnaître cependant que de gros progrès restent à faire ; pour l'instant, les applications pratiques restent limitées, malgré l'ampleur des besoins contrôle des processus de production continus, notamment dans les industries chimiques, pétrolières et métallurgiques, problèmes de guidage d'engins, etc...

3. — PROGRAMMES DYNAMIQUES STOCHASTIQUES

Venons-en maintenant aux systèmes dont l'évolution dépend à la fois de décisions et de facteurs aléatoires. Définissons une « stratégie » comme spécifiant, pour chaque état possible du système à tout instant futur, la décision à prendre. A chaque stratégie correspond alors un processus stochastique qui gouverne l'évolution du système lorsqu'on adopte cette stratégie. Le principe d'optimalité s'étend facilement à ce cas, si l'on prend comme critère l'espérance mathématique des valeurs associées aux trajectoires possibles pour une stratégie donnée.

Lorsque l'évolution temporelle est discrète et les états possibles dénombrables, à chaque stratégie correspond une chaîne de Markov. Howard [16] a étudié en détail ces problèmes, et indiqué un algorithme en horizon infini (voir aussi [7] [33] et [26]). Kaufmann, Cruon et Companys [20] se sont penchés sur le cas particulier important des programmes « sous forme décomposée », où décision et hasard interviennent alternativement. De façon un peu différente, Denardo [10] a défini des problèmes « séparables » (voir aussi [13]). D'autre part, signalons l'extension aux processus semi-markoviens [16] [30], qui revient à peu près à considérer le temps comme une composante de la variable d'état. Nous

---

(1) On trouvera dans Pallu de la Barrière [27] et dans Connors et Teichroew [8] des exposés récents de la théorie du contrôle optimal.

renvoyons sur ce sujet à Denardo et Fox [11] et surtout à de Ghellinck et Peeters [14], où l'on trouvera un exposé d'ensemble. Le cas très difficile des processus markoviens en temps continu a été abordé par de Lève [24].

On peut déplorer que peu d'efforts aient été consacrés aux problèmes d'études paramétriques et d'études de sensibilité. On peut citer cependant le paramétrage du coefficient d'actualisation dans les problèmes économiques [31], et l'exploration du voisinage de l'optimum dans les programmes dynamiques discrets [21] [22] [9].

Les applications de la programmation dynamique stochastique sont nombreuses, notamment dans la gestion des stocks, la régulation simultanée de la production et des stocks, de nombreux problèmes d'investissement et d'allocation de ressources, la fiabilité et l'entretien des équipements.

4. — PROGRAMMES DYNAMIQUES ADAPTATIFS

L'hypothèse que l'avenir est connu en probabilité constitue une approximation souvent insuffisante. Même en supposant qu'il existe une structure probabiliste sous-jacente, et qu'elle est immuable, les paramètres des lois de probabilité qui interviennent dans le phénomène sont, en général, mal connus. L'évolution même du système apportant à chaque instant des informations nouvelles, il est naturel d'en profiter pour améliorer continuellement l'estimation de ces paramètres, dans l'esprit de la théorie de Wald.

D'un point de vue purement mathématique, les programmes dynamiques adaptatifs [4] auxquels on est alors conduit se ramènent à des programmes stochastiques ; il suffit, en effet, d'inclure dans la variable d'état du système une ou plusieurs composantes représentant un résumé exhaustif (au sens de la statistique) des observations antérieures. Le modèle comporte alors, outre la description des lois d'évolution physique du système, celle des règles de mise à jour des informations obtenues, règles qui sont généralement du type bayesien [1] [32] [25].

Cette théorie a eu jusqu'à présent des applications limitées, car elle conduit à une variable d'état multi-dimensionnelle, et la résolution numérique se heurte alors à la difficulté mentionnée au paragraphe 2 concernant le nombre d'états possibles.

D'autre part, le monde économique évolue rapidement ; dans ce domaine, on n'est jamais assuré que la structure probabiliste du problème, ou même la loi d'évolution du système et la fonction d'évaluation resteront immuables. Il faut alors recourir à des procédures heuristiques, du type « apprentissage et adaptation ». Ces procédures peuvent sommairement être décrites de la façon suivante : partant d'un corps d'hypothèses sur la structure de l'« univers » auquel on a affaire, on construit un modèle stochastique « adaptatif » au sens indiqué plus haut (nous parlerons plutôt d'apprentissage) ; on détermine une stratégie optimale

sur un certain intervalle d'anticipation (dont le choix fait partie de la construction du modèle), et on exécute la *première* décision de cette stratégie. On observe alors l'évolution du système et de son environnement, et on recommence en modifiant éventuellement le modèle initialement adopté. La chaîne cybernétique du comportement est donc la suivante :

modèle stochastique sur un horizon fixé — optimisation sur cet horizon — décision conforme à la stratégie optimale — boîte noire — apprentissage (effets bayesiens) — adaptation (remise en question du modèle) — nouveau modèle.

Un tel processus, qui fait constamment appel à l'initiative humaine, nous paraît seul susceptible de répondre aux nécessités de l'action dans un monde incertain.

REFERENCES

[1] M. Aoki, *Optimization of stochastic systems. Topics in discrete time systems*, Academic Press (1967).

[2] R. Aris, *Discrete dynamic programming*, Blaisdell Publ. Co. (1964).

[3] R. Bellman, *Dynamic programming*, Princeton Univ. Press, Princeton (N. J.), 342 p. (1957).

[4] R. Bellman, *Adaptive control processes : a guided tour*, Princeton Univ. Press, Princeton (N. J.), (1961).

[5] R. Bellman et S. Dreyfus, *Applied dynamic programming*, Princeton Univ. Press, Princeton (N. J.), 363 p. (1962).

[6] R. Bellman et R. Kalaba, *Dynamic programming and modern control theory*, Academic Press (1965).

[7] D. Blackwell, « Discrete dynamic programming », *Ann. Math. Stat.* **33**, 719-26 (1962).

[8] M. Connors et D. Teichroew, *Optimal Control of dynamic Operations Research models*, International Textbook Co, Scranton, Pa, 118 p. (1967).

[9] A. Deledicq, « Programmation dynamique discrète : k-optimums d'un problème séquentiel », *Revue française d'Informatique et de R. O.*, **II**-V2, 13-32 (août 1968).

[10] E. V. Denardo, « Separable Markovian decision problems », *Management Science*, **14**, 7, 451-62 (mars 1968).

[11] E. V. Denardo et B. L. Fox, « Multichain Markov renewal programs », *SIAM J. Appl. Math.*, **16**, 3, 468-87 (mai 1968).

[12] S. E. Dreyfus, *Dynamic programming and the calculus of variations*, Academic Press, 248 p. (1965).

[13] G. de Ghellinck et G. D. Eppen, « Linear programming solutions for separable Markovian decision problems », *Manag. Science*, **13**, 5, 371-94 (janvier 1967).

[14] G. de Ghellinck et L. Peeters, « Markov programming » Invited paper, European Meeting 1968, IMS-TIMS-ES-IASPS, Amsterdam, 2-7 sept. 1968.

[15] Ph. Hervé, « Les procédures arborescentes d'optimisation », *Revue Franç. d'Inf. et de Rech. Op.*, n° 14-V3, pp. 69-80 (1968).

[16] R. Howard, *Dynamic programming and Markov processes*, The Technology Press of the M.I.T. et J. Wiley, 136 p. (1960).

[17] R. Howard, « Dynamic programming », *Management Science*, **12**, 5, 317-48 (janvier 1966).

[18] O. R. L. Jacobs, *An introduction to dynamic programming*, Chapman et Hall, London, 124 p. (1967).

[19] W. S. JEWELL, « Markov-Renewal programming », I : Formulation, Finite return models. II : Infinite return models, Example », *Operations Research*, **11**, 6, 938-71 (1963).

[20] A. KAUFMANN et R. CRUON, *La programmation dynamique. Gestion scientifique séquentielle*, Dunod, Paris, 273 p. (1965).
Dynamic programming : Sequential scientific Management, translated by H. Sneid, Academic Press (1967).

[21] A. KAUFMANN et R. CRUON « Étude de la sensibilité en programmation dynamique : politiques k-optimales en avenir certain », *Revue Française de Recherche Opérationnelle*, 32, 293-302 (4e trimestre 1964).

[22] A. KAUFMANN et R. CRUON, « Stratégies k-optimales dans les programmes dynamiques stochastiques finis », 4e Congrès international de l'IFORS, Boston, 20 août-2 septembre 1966.

[23] R. E. LARSON, « A survey of dynamic programming computational procedures », *IEEE Trans. on Automatic control*, **AC-12**, 6, 767-74 (déc. 1967).

[24] G. de LEVE, *Generalized Markovian decision processes.* I. *Model and Method.* II : *Probabilistic background*, Mathematisch Centrum, 2nd Boerhaavestraat 49, Amsterdam, 128 + 123 p. (1964).

[25] J. J. MARTIN, *Bayesian decision problems and Markov chains*, Wiley (1967).

[26] B. L. MILLER, « Finding optimal policies in discrete dynamic programming », RAND memorandum RM-5601-PR, 11 p. (avril 1968).

[27] R. PALLU DE LA BARRIÈRE, *Cours d'automatique théorique*, Dunod, Paris (1966).
Optimal Control Theory, Saunders, Philadelphie (1967).

[28] L. S. PONTRYAGIN, V. BOLTYANSKII, R. GAMKRELIDZE et E. MISHCHENKO, *The mathematical theory of optimal processes*, Wiley (1962).

[29] S. M. ROBERTS, *Dynamic programming in chemical engineering and process control*, Academic Press (1964).

[30] P. J. SCHWEITZER, *Perturbation theory and Markovian decision processes*, Techn. Report nº 15, Contract Nonr-1841-(87), NR 042-230, Mass. Institute of Techn. (AD 618-406), 316 p. (juin 1965).

[31] R. D. SMALLWOOD, « Optimum policy regions for Markov processes with discounting », *Operations Research*, **14**, 4, 658-69 (juillet-août 1966).

[32] D. SWORDER, *Optimal adaptive control systems*, Academic Press (1966).

[33] A. F. VEINOTT, Jr., « On finding optimal policies in discrete dynamic programming with no discounting », *Ann. Math. Stat.*, **37**, 1284-95 (1966).