REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE, SÉRIE ROUGE

A. AUSLENDER
F. BRODEAU

Convergence d'un algorithme de Franck et Wolfe appliqué à un problème de contrôle

Revue française d'informatique et de recherche opérationnelle, série rouge, tome 2, n^{o} 1 (1968), p. 3-12.
<http://www.numdam.org/item?id=M2AN_1968__2_1_3_0>

R.I.R.O.
(2e année, N° 7, 1968, p. 3-12)

CONVERGENCE D'UN ALGORITHME DE FRANCK ET WOLFE APPLIQUE A UN PROBLEME DE CONTROLE

par A. AUSLENDER et F. BRODEAU (*)

Résumé. — « *On résoud un problème de contrôle optimal par l'extension à un espace fonctionnel d'un algorithme devenu classique dans les problèmes de minimisation dans les espaces de type* $\mathbf{R}^n$. »

INTRODUCTION

On se propose ici de résoudre un problème vectoriel de contrôle caractérisé par une loi d'évolution linéaire, un critère de type quadratique et des contraintes sur le contrôle. Ce problème est ramené à l'étude d'un opérateur défini dans un espace fonctionnel. Les propriétés de cet opérateur permettent d'utiliser un algorithme de résolution qui est une extension d'un algorithme proposé par Frank et Wolfe [1] dans le cas de la minimisation d'une fonction convexe sur un polyèdre convexe de $\mathbf{R}^n$.

Des conditions d'extension de cet algorithme ont déjà été données par Valadier [2], et par Demjanov et Rubinov [3], mais elles ne s'appliquent pas exactement à l'exemple étudié ici.

I. — POSITION DU PROBLEME

Soit T un nombre réel positif donné et r un entier positif donné; on désigne par $\mathcal{H}$ l'espace de Hilbert des fonctions définies sur $[0, T]$ à valeurs dans $\mathbf{R}^r$ qui sont de carré sommable.

* Service de Mathématiques Appliquées de la Faculté des Sciences de Grenoble.

On note ainsi le produit scalaire défini sur $\mathcal{H}$:

$$\langle f, g \rangle = \int_0^T f^*(t) \cdot g(t)\, \mathrm{d}t \ ; f, g \in \mathcal{H}$$

où $f^*(t)$ désigne la matrice transposée de la matrice associée au vecteur $f(t)$, et $g(t)$ la matrice associée au vecteur $g(t)$.

Soit C un sous-ensemble borné, fermé et convexe de $\mathcal{H}$; on appellera stratégie admissible un élément de C.

On étudie un système physique Σ sur l'intervalle de temps $[0, T]$. L'état de Σ à l'instant t est représenté par un vecteur $X(t)$ de $\mathbf{R}^m$, où m est un entier positif donné.

Si Y est la stratégie admissible appliquée à Σ alors $X(t)$ est solution de :

$$(1) \qquad X'(t) = A(t)X(t) + B(t)Y(t) + Z(t), X(0) = X_0,$$

où $A(t)$, $B(t)$, $Z(t)$ sont des matrices données, respectivement (m, m), (m, r), $(m, 1)$, définies et continues sur $[0, T]$.

Le problème que nous étudions est de minimiser sur C la fonctionnelle :

$$(2) \quad v[Y] = \frac{1}{2}\left[X^*(T)SX(T) + \int_0^T [X^*(t)Q(t)X(t) + Y^*(t)R(t)Y(t)]\, \mathrm{d}t\right],$$

où X correspond à Y par l'intermédiaire de (1); S est une matrice carrée symétrique d'ordre m définie positive, $Q(t)$ et $R(t)$ sont, pour toute valeur de t sur $[0, T]$, des matrices carrées symétriques données, respectivement d'ordre m et r, définies positives. De plus Q et R sont continues sur $[0, T]$. Le problème a effectivement un sens puisque les hypothèses assurent que, Y étant choisi, (1) admet une solution unique définie sur $[0, T]$; d'autre part on a $v[Y] \geqslant 0$.

II. — PROPRIETES DE LA FONCTIONNELLE v

Proposition 1 : *v est une application convexe de $\mathcal{H}$ dans* $\mathbf{R}$.

Cette propriété découle facilement des deux faits suivants :

— Si X_1 et X_2 correspondent respectivement à Y_1 et Y_2 par (1), $aX_1 + (1 - a)X_2$ correspond à $aY_1 + (1 - a)Y_2$ (a est un nombre réel quelconque).

— x^*Sx est une application convexe de $\mathbf{R}^m$ dans $\mathbf{R}$,

— $x^*Q(t)x$ et $y^*R(t)y$ sont, pour toute valeur de t, des applications convexes respectivement de $\mathbf{R}^m$ et $\mathbf{R}^r$ dans $\mathbf{R}$.

Les propriétés de différentiabilité sont fondamentales pour l'utilisation de l'algorithme de Frank et Wolfe.

Proposition 2 : *v est une application différentiable au sens de Frechet et son gradient F est défini par :*

$$F[Y] = RY - B^*\psi, \tag{3}$$

où ψ est une application de $[0, T]$ *dans* $\mathbf{R}^m$ *définie par :*

$$\psi'(t) = - A^*(t)\psi(t) + Q(t)X(t),\ \psi(T) = - SX(T), \tag{4}$$

$X(t)$ correspondant à Y par (1).

Démonstration : Y et ΔY étant des éléments de $\mathcal{H}$ évaluons

$$v[Y + \Delta Y] - v[Y]$$

à Y et $Y + \Delta Y$ correspondent respectivement par (1) X et $X + \Delta X$, où ΔX satisfait au système différentiel suivant :

$$[\Delta X(t)]' = A(t)\,\Delta X(t) + B(t)\,\Delta Y(t),\ \Delta X(0) = 0. \tag{5}$$

On a alors :

$$v[Y + \Delta Y] - v[Y] = X^*(T)S\,\Delta X(T) + \int_0^T [X^*(t)Q(t)\,\Delta X(t) + Y^*(t)R(t)\,\Delta Y(t)]\,\mathrm{d}t + \omega(Y, \Delta Y)$$

où :

$$\omega(Y, \Delta Y) = \frac{1}{2}[\Delta X(T)]^*S[\Delta X(T)] + \frac{1}{2}\int_0^T [[\Delta X(t)]^*Q(t)[\Delta X(t)] \ldots + [\Delta Y(t)]^*R(t)[\Delta Y(t)]]\,\mathrm{d}t.$$

De (5) on déduit, en utilisant les hypothèses faites sur les matrices A, B, Q et R, qu'il existe une constante H telle que :

$$\|\omega(Y, \Delta Y)\,| \leqslant H\,\|\Delta Y\|^2, \tag{6}$$

où $\|\ \|$ désigne la norme de $\mathcal{H}$ associée au produit scalaire ; on a alors :

$$\frac{|\omega(Y, \Delta Y)|}{\|\Delta Y\|} \to 0 \text{ si } \|\Delta Y\| \to 0. \tag{7}$$

ψ étant défini par (4), on montre que :

$$X^*(t)Q(t)\,\Delta X(t) = [\psi^*(t)\,\Delta X(t)]' - \psi^*(t)B(t)\,\Delta Y(t),$$

d'où :

$$v[Y + \Delta Y] - v[Y] = \int_0^T [Y^*(t)R(t) - \psi^*(t)B(t)]\,\Delta Y(t)\,\mathrm{d}t + \omega(Y, \Delta Y). \tag{8}$$

Le résultat de la proposition 2 découle alors immédiatement de (7) et (8).

Il est aussi utile de montrer que v est fortement convexe.

Définition [4] : *f est une application fortement convexe d'un espace de Banach dans* **R** *si il existe un nombre réel $\gamma > 0$ tel que :*

$$\forall u_1, u_2 \in \mathcal{B}, \forall \alpha \in [0,1] f[\alpha u_1 + (1-\alpha)u_2] \leqslant \alpha f[u_1] + (1-\alpha)f[u_2] - \gamma\alpha(1-\alpha)\|u_1 - u_2\|^2,$$

Proposition 3 : *v est une application fortement convexe.*

Comme v est convexe il suffit de montrer que l'opérateur défini par :

$$g[Y] = \int_0^T Y^*(t)R(t)Y(t)\,\mathrm{d}t$$

est fortement convexe. Or pour $\alpha \in [0, 1]$, Y_1, $Y_2 \in \mathcal{H}$, on a :

$$(9) \quad g[\alpha Y_1 + (1-\alpha)Y_2] = \alpha g[Y_1] + (1-\alpha)g[Y_2] - \alpha(1-\alpha)\int_0^T (Y_1(t) - Y_2(t))^* R(t)(Y_1(t) - Y_2(t))\,\mathrm{d}t.$$

Les hypothèses faites sur la matrice R entraînent l'existence d'une constante $a > 0$ telle que :

$$\forall Y \in \mathcal{H}, \int_0^T Y^*(t)R(t)Y(t)\,\mathrm{d}t \geqslant a\|Y\|^2.$$

Le résultat s'obtient alors par l'utilisation de (9).

III. — ALGORITHME DE RESOLUTION

Il faut d'abord s'assurer que le problème admet une solution : en effet v est convexe et continu, donc faiblement semi-continu inférieurement. D'autre part, C est faiblement compact; par conséquent v atteint son minimum en au moins un élément de C.

L'algorithme de Frank et Wolfe consiste à construire une suite $\{Y_n\}$ d'éléments de C de la façon suivante :

Y_0 est choisi arbitraire

Y_n étant connu on détermine Y_{n+1} en deux étapes.

1. On détermine $\bar{Y}_n$ rendant minimal sur $C \langle Y, F[Y_n] \rangle$, considéré comme fonction de Y.

2. Y_{n+1} est alors un élément de C réalisant le minimum de :

$$v[Y_n + \alpha(\bar{Y}_n - Y_n)],$$

considéré comme fonction de α définie sur l'intervalle [0, 1].

Cet algorithme, noté (A), s'arrête si $\langle \bar{Y}_n - Y_n, F[Y_n] \rangle = 0$, car on verra plus loin que, dans ce cas, Y_n est solution du problème.

Ce procédé permet de déterminer effectivement une suite Y_n. En effet $\langle Y, F[Y_n] \rangle$ est, comme fonction de Y, une application linéaire continue de $\mathcal{H}$ dans $\mathbf{R}$; la phase 1 permet donc d'affirmer l'existence d'au moins un élément $\bar{Y}_n$.

$\bar{Y}_n$ étant déterminé on montre que $v[Y_n + \alpha(\bar{Y}_n - Y_n)]$ est, comme fonction de α un polynôme du second degré. Le minimum de cette expression sur [0 1,] existe donc, ce qui assure l'existence de Y_{n+1} (cf. § IV).

On rappelle le résultat suivant extrait de [3].

Lemme 1. *Si* $\lim \langle \bar{Y}_n - Y_n, F[Y_n] \rangle = 0$, $\{ Y_n \}$ *est une suite minimisante; c'est-à-dire que* $\lim v[Y_n] = \min_{Y \in C} v[Y]$.

Démonstration : Soit Y_0 un élément de C où v atteint son minimum. La propriété de convexité de v permet de montrer que :

$$v[Y] - v[Y_n] \geqslant \langle Y - Y_n, F[Y_n] \rangle, \qquad \forall Y \in \mathcal{H}.$$

On en déduit :

$$\begin{aligned} v[Y_0] - v[Y_n] \geqslant \langle Y_0 - Y_n, F[Y_n] \rangle &\geqslant \operatorname*{Min}_{Y \in C} \langle Y - Y_n, F[Y_n] \rangle \\ &= \langle \bar{Y}_n - Y_n, F[Y_n] \rangle \end{aligned}$$

Finalement :

$$0 \leqslant v[Y_n] - v[Y_0] \leqslant \langle Y_n - \bar{Y}_n, F[Y_n] \rangle$$

ce qui démontre le lemme.

Ce résultat nous permet d'obtenir le théorème essentiel :

Théorème 1 : *Toute suite* $\{ Y_n \}$ *construite par* (A) *est minimisante et on peut d'ailleurs extraire de cette suite une sous-suite convergente faiblement vers une solution du problème.*

Démonstration : De la façon dont est construite $\{ Y_n \}$ on déduit les deux conséquences :

(10) $$\langle \bar{Y}_n, F[Y_n] \rangle \leqslant \langle Y, F[Y_n] \rangle, \qquad \forall Y \in C$$

(11) $$v[Y_{n+1}] \leqslant v[Y_n + \alpha(\bar{Y}_n - Y_n)], \qquad 0 \leqslant \alpha \leqslant 1.$$

Il résulte de (11) que $v[Y_{n+1}] \leqslant v[Y_n]$, $n = 0, 1, 2, \ldots$ Puisque $v[Y] \geqslant 0$, $\forall Y \in C$, on peut affirmer que la suite $\{ v[Y_n] \}$ converge vers une limite notée v_0. D'autre part C étant borné on peut extraire de $\{ Y_n \}$ une sous-suite notée $\{ Y_{n1} \}$ convergeant faiblement vers un élément Y_e de $\mathcal{H}$. Comme C est convexe et fermé, Y_e est élément de C. Puisque v est faiblement semi-continu inférieurement, on a donc :

(12) $$\lim v[Y_{n1}] = v_0 \geqslant v[Y_e].$$

L'utilisation de (11) permet d'écrire :

$$v[Y_{n1} + \alpha(\bar{Y}_{n1} - Y_{n1})] - v[Y_{n1}] \geqslant v[Y_{n1+1}] - v[Y_{n1}]$$

et puisque $v[Y_{n1+1}] - v[Y_{n1}] \to 0$ si $n_1 \to \infty$:

$$\underline{\lim} \{ v[Y_{n1} + \alpha(\bar{Y}_{n1} - Y_{n1})] - v[Y_{n1}] \} \geqslant 0,$$

d'où d'après (8) :

$$(13) \quad \underline{\lim} \{ \alpha \langle \bar{Y}_{n1} - Y_{n1}, F[Y_{n1}] \rangle + \omega(Y_{n1}, \alpha(\bar{Y}_{n1} - Y_{n1})) \} \geqslant 0.$$

Or d'après (6) et le fait que C est borné, il existe une constante $K > 0$ telle que :

$$|\omega(Y_{n1}, \alpha(\bar{Y}_{n1} - Y_{n1}))| \leqslant K\alpha^2.$$

L'inégalité (13) entraîne alors, en utilisant une valeur de α suffisamment petite :

$$(14) \qquad \underline{\lim} \langle \bar{Y}_{n1} - Y_{n1}, F[Y_{n1}] \rangle \geqslant 0.$$

Or d'après (10) :

$$(15) \qquad \langle \bar{Y}_{n1} - Y_{n1}, F[Y_{n1}] \rangle \leqslant 0.$$

La comparaison de (14) et (15) prouve que :

$$\underline{\lim} \langle \bar{Y}_{n1} - Y_{n1}, F[Y_{n1}] \rangle = 0.$$

On peut donc extraire de $\{ Y_{n1} \}$ une sous-suite, notée $\{ Y_{n2} \}$, telle que :

$$\lim \langle \bar{Y}_{n2} - Y_{n2}, F[Y_{n2}] \rangle = 0.$$

D'après le lemme 1 $\{ Y_{n2} \}$ est une suite minimisante : $v[Y_{n2}]$ converge vers la valeur minimale v_m. D'après la propriété de monotonie de la suite $\{v[Y_n]\}$, $\{ v[Y_n] \}$ converge aussi vers v_m et l'on a $v_0 = v_m$. D'après (12) on a $v[Y_e] = v_m$; Y_e est donc solution du problème, d'où le théorème.

Montrons comment les résultats de ce théorème peuvent être améliorés en utilisant les propriétés de convexité forte de l'opérateur v. On énonce dans [3] le résultat suivant :

Lemme 2 : *Si f est une application fortement convexe d'un espace de Banach $\mathcal{B}$ dans* **R** *et admet en tout point de $\mathcal{B}$ un gradient F, on a, $\forall u_1, u_2 \in \mathcal{B}$:*

$$(16) \qquad f(u_1) - f(u_2) \geqslant \langle F[v_2] \cdot (u_1 - u_2) \rangle + \gamma \|u_2 - u_1\|^2$$

Ce lemme a pour conséquence le résultat suivant :

Théorème 2 : *la suite $\{ Y_n \}$ converge fortement vers Y_e; Y_e est solution unique du problème.*

L'inégalité (16) appliquée à v nous donne :

(17) $v[Y]-v[Y_e] \geqslant \langle Y-Y_e, F[Y_e] \rangle + \gamma \| Y-Y_e \|^2, \qquad \forall Y \in C.$

Or d'après le théorème 1 de [3], puisque Y_e est solution du problème,

$$\min_{Y \in C} \langle Y-Y_e, F[Y_e] \rangle = 0.$$

(17) entraîne donc :

(17 b) $$v[Y]-v[Y_e] \geqslant \gamma \| Y-Y_e \|^2.$$

Cette inégalité appliquée à Y_n s'écrit :

$$v[Y_n]-v[Y_e] \geqslant \gamma \| Y_n-Y_e \|^2,$$

puisque :

$$v[Y_n] \to v[Y_e] \qquad \text{quand} \qquad n \to \infty$$

$$\| Y_n-Y_e \| \to 0 \qquad \text{quand} \qquad n \to \infty$$

D'autre part, l'inégalité (17 b) étant valable pour tout Y_e solution du problème, on a l'unicité de Y_e.

Ce théorème admet comme conséquence que la suite $\{ X_n \}$, associée à $\{ Y_n \}$ par (1), *converge uniformément* vers X, associé à Y.

IV. — RESOLUTION NUMERIQUE

Trois exemples de sous-ensembles C seront envisagés. Nous désignerons par Y^i ($i = 1, ..., r$) les composantes de l'élément Y de $\mathcal{H}$.

a) C est le sous-ensemble C_1 des éléments Y tels que, pour presque toute valeur de t sur $[0, T]$:

$$|Y^i(t)| \leqslant \lambda \qquad (i = 1, ..., r),$$

b) C est le sous-ensemble C_2 des éléments Y tels que, pour presque toute valeur de t sur $[0, T]$:

$$\left[\sum_{i=1}^{r} |Y^i(t)|^2 \right]^{1/2} \leqslant \mu,$$

c) C est le sous-ensemble C_3 des éléments Y tels que :

$$\| Y \| \leqslant \nu,$$

λ, μ, ν sont des constantes positives données.

Il est facile de vérifier que C_1, C_2, C_3 sont des ensembles convexes, bornés et fermés. Montrons comment, pour chacun de ces exemples, la construction d'une suite minimisante $\{ Y_n \}$ peut être entreprise. Supposons connu Y_n.

1. Détermination de $\bar{Y}_n$

Il s'agit de minimiser :

$$\int_0^T \sum_{i=1}^{r} Y^i(t) \cdot F^i[Y_n(t)]\, dt.$$

Dans le cas *a)* et *b)* le problème est équivalent à la minimisation de l'expression :

$$\sum_{i=1}^{r} Y^i(t) \cdot F^i[Y_n(t)] \tag{18}$$

pour presque toute valeur de t, dans la mesure où cette minimisation permet d'obtenir un élément de $\mathcal{H}$.

Cas a : le résultat est immédiat; on peut prendre :

$$\bar{Y}_n^i(t) \begin{cases} = -\lambda & \text{si} \quad F^i[Y_n(t)] > 0 \\ = +\lambda & \text{si} \quad F^i[Y_n(t)] < 0 \\ \text{quelconque} & \text{si} \quad F^i[Y_n(t)] = 0 \end{cases}$$

Cas b : on pose $Y = kF[Y_n] + W$ où k est une constance réelle et W un élément de $\mathcal{H}$ orthogonal à $F[Y_n]$ (18) s'écrit alors :

$$k \sum_{i=1}^{r} |F^i[Y_n(t)]|^2 .$$

cette quantité est minimale pour la plus petite valeur de k autorisée; or on doit avoir :

$$k^2 \sum_{i=1}^{r} |F^i[Y_n(t)]|^2 + \sum_{i=1}^{r} |W^i|^2 \leqslant \mu^2$$

la plus petite valeur de k est donc obtenue pour $\sum_{i=1}^{r} |W^i|^2 = 0$ et est égale à :

$$\frac{-\mu}{\left\{ \sum_{i=1}^{r} |F^i[Y_n(t)]|^2 \right\}^{1/2}},$$

d'où :

$$\bar{Y}_n^i(t) = \frac{-\mu F^i Y_n(t)}{\left\{ \sum_{i=1}^{r} |F^i[Y_n(t)]|^2 \right\}^{1/2}}, \quad \text{si} \quad \sum_{i=1}^{r} |F^i[Y_n(t)]|^2 > 0,$$

sinon $\bar{Y}_n(t)$ est indéterminé.

Cas c : ce cas se traite comme le cas *b*. On trouve :

$$\bar{Y}_n^i(t) = \frac{-\nu F^i[Y_n(t)]}{\|F[Y_n]\|}, \qquad \text{si} \qquad \|F[Y_n]\| > 0 .$$

sinon $\bar{Y}_n$ est indéterminé.

2. Détermination de Y_{n+1}

Soit U_1 et U_2 deux éléments de $\mathcal{H}$ auxquels correspondent par (1) V_1 et V_2. On pose :

$$G(U_1, U_2) = \frac{1}{2}\left[V_1^*(T)SV_2(T) + \int_0^T \{ V_1^*(t)Q(t)V_2(t) + U_1^*(t)R(t)\, U_2(t) \}\, \mathrm{d}t\right].$$

Un calcul élémentaire montre que :

$$(19) \quad v[Y_n + \alpha(\bar{Y}_n - Y_n)] = v[Y_n] + 2\alpha[-v[Y_n] + G(\bar{Y}_n, Y_n)] + \ldots + \alpha^2(v[\bar{Y}_n] + v[Y_n] - 2G(\bar{Y}_n, Y_n)).$$

Comme S, $Q(t)$, $R(t)$ sont des matrices définies positives, on constate d'ailleurs que le coefficient de α^2 ne s'annule que lorsque $\bar{Y}_n = Y_n$, d'où deux cas :

1. $\langle \bar{Y}_n - Y_n, F[Y_n] \rangle = 0$, alors Y_n est solution du problème d'après le théorème 1 de [3] et la convexité de la fonction $v(y)$; il est inutile de poursuivre.

2. $\langle \bar{Y}_n - Y_n, F[Y_n] \rangle \neq 0$, alors on peut facilement déterminer la valeur α_n de α rendant minimale sur [0, 1] le second membre de (19).

Si on pose :

$$\bar{\alpha}_n = \frac{v[Y_n] - G(\bar{Y}_n, Y_n)}{v[\bar{Y}_n] + v[Y_n] - 2G(\bar{Y}_n, Y_n)},$$

on a : $\alpha_n = \min(1, \bar{\alpha}_n)$.

Le cas $\alpha_n = 0$ étant impossible, puisqu'on aurait alors

$$\langle \bar{Y}_n - Y_n, F[Y_n] \rangle = 0.$$

On peut ainsi envisager de résoudre certains exemples numériques. Il est alors important de savoir arrêter la détermination des Y_n lorsqu'une précision fixée à l'avance pour v_m a été atteinte. Ceci découle de l'inégalité :

$$0 \leqslant v[Y_n] - v_m \leqslant \langle Y_n - \bar{Y}_n, F[Y_n] \rangle,$$

obtenue dans la démonstration du lemme 1. A chaque étape le calcul de $\langle Y_n - \bar{Y}_n, F[Y_n] \rangle$ permet ainsi de savoir si la poursuite de l'algorithme doit ou ne doit pas être envisagée.

Nous nous proposons d'entreprendre effectivement la résolution numérique d'exemples. Bien évidemment la méthode exposée ici peut s'étendre à d'autres exemples de problèmes de contrôle.

(mai 1967)

BIBLIOGRAPHIE

[1] M. FRANK and P. WOLFE, *An algorithm for quadratic programming*, Naval Research Logistics Quaterly (1956).

[2] M. VALADIER, Extension d'un algorithme de Frank et Wolfe, *Revue française de Recherche Opérationnelle* n° **36** (1965).

[3] V. F. DEM'JANOV and A. M. RUBINOV, On the problem of minimization of a smooth functional with convex constraints, *Soviet Mathematics*, vol. **6**, n° 1, 1965.

[4] B. T. POLJAK, Existence theorems and convergence of minimizing sequences in extremum problems with restrictions, *Soviet Math.*, 1966, t. **166**, n° 2.