

## UTILISATION DE DENSITES DES PREMIER PASSAGE EN COMMANDE OPTIMALE STOCHASTIQUE

MARIO LEFEBVRE, \*Ecole Polytechnique de Montréal

### Abstract

A theorem that gives the optimal control of Gaussian processes using the mathematical expectation of a function of the time and the place where the uncontrolled processes hit the boundary of the stopping region for the first time is proved. The result obtained in this note is an extension of a theorem due to Whittle.

STOCHASTIC CONTROL; FIRST-PASSAGE DENSITIES; WIENER-LEVY PROCESS

### 1. Introduction

Dans cette note on considère des systèmes dynamiques définis par l'équation différentielle stochastique

$$(1) \quad dx(t) = a(x, t) dt + B(t)u(t) dt + dW(t),$$

où  $x(t)$  est un processus vectoriel de dimension  $n$ ,  $a(x, t)$  est un vecteur dans  $R^n$ ,  $B(t)$  est une matrice  $n \times m$  et la commande  $u(t)$  est un vecteur dans  $R^m$ . Le système est soumis à des perturbations aléatoires sous forme de bruit blanc gaussien; c'est-à-dire,  $W(t)$  est un processus de Wiener-Lévy à  $n$  dimensions défini par

$$(2) \quad \begin{aligned} E[W(t)] &= 0 \\ E[W(t)W'(t)] &= N(t) dt, \end{aligned}$$

où  $N(t)$  est une matrice carrée d'ordre  $n$  symétrique et définie positive.

Whittle (1982) a considéré la fonction de coût

$$(3) \quad J_0 = \int_0^T [u'(t)Q(t)u(t)/2] dt + K[x(T), T],$$

où  $Q(t)$  est une matrice carrée d'ordre  $m$  symétrique et définie positive,  $K$  est une fonction générale de coût final et  $T$  est le moment auquel le processus  $x(t)$  quitte pour la première fois une certaine région  $C$ , appelée région de continuation. Il a démontré (voir aussi Whittle et Gait (1970)) que s'il existe une constante positive  $c$  telle que l'on peut écrire

$$(4) \quad N = cBQ^{-1}B',$$

alors la commande optimale est donnée par

$$(5) \quad u^* = -Q^{-1}B'F_x,$$

où

$$(6) \quad F(x, t) = -c \ln \{E_{(x,t)}[\exp(-K(x(\tau), \tau)/c)]\}.$$

Reçu le 29 juin 1987; révision reçue le 15 octobre 1987.

\* Adresse postale: Département de mathématiques appliquées, Ecole Polytechnique de Montréal, Case postale 6079, succursale "A", Montréal, Québec, Canada H3C 3A7.

Recherche subventionnée par le Conseil de recherches en sciences naturelles et en génie du Canada. Subvention no A7989.

Dans l'équation (6) l'espérance mathématique est par rapport au moment  $\tau$  et à l'endroit  $x(\tau)$  où le processus non commandé

$$(7) \quad dx(t) = a(x, t) dt + dW(t),$$

parti de  $(x, t)$ , quitte la région  $C$  pour la première fois. En fait, en plus de la relation (4), il faut aussi que  $P[\tau < \infty] = 1$ .

Le résultat de Whittle est remarquable, puisqu'il permet de trouver la commande optimale en considérant uniquement le processus non commandé. De plus, le théorème est encore vrai même si  $B, N$  et  $Q$  dépendent aussi de  $x$ , en plus du temps. Toutefois, la fonction de coût qu'il a considérée est très particulière, en ce sens qu'elle ne pénalise pas les déviations de la variable d'état  $x(t)$  de la trajectoire idéale, sauf au temps final. Bien sûr, le fait que  $T$  représente le temps de *premier* passage à l'extérieur de la région de continuation  $C$  implique que c'est comme si un coût était effectivement encouru lorsque  $x(t)$  dévie de la trajectoire idéale; cependant, on peut se demander si le théorème de Whittle peut être étendu au cas où les déviations de la trajectoire optimale sont pénalisées explicitement. Dans cette note, nous donnons un cas où la réponse à cette question est affirmative.

### 2. Le résultat principal

On considère des systèmes dynamiques définis par l'équation (1). On remplace la fonction de coût définie à l'équation (3) par

$$(8) \quad J = \int^T [u'(t)Q(t)u(t)/2 + u'(t)Q(t)r(t) + s(x, t)] dt + K[x(T), T],$$

où  $r(t)$  est un vecteur dans  $R^m$  et  $s(x, t)$  est une fonction à valeurs réelles. La fonction  $J_0$  utilisée par Whittle est donc un cas particulier de la fonction  $J$ .

Soit  $F(x, t)$  le coût minimal moyen encouru à partir de  $(x, t)$ . Pour  $(x, t)$  dans la région  $C$  on obtient l'équation de programmation dynamique

$$(9) \quad 0 = \inf_u [u'Qu/2 + u'Qr + s + F_t + (a + Bu)'F_x + \text{tr}(NF_{xx})/2].$$

La condition frontière est

$$(10) \quad F(x, t) = K(x, t) \quad ((x, t) \in D),$$

où  $D$  est la complément de  $C$ . On trouve en dérivant que la commande optimale est donnée par

$$(11) \quad u^* = -[Q^{-1}B'F_x + r].$$

En remplaçant dans (9) on obtient

$$(12) \quad 0 = -F_x' B Q^{-1} B' F_x / 2 - r' Q r / 2 - r' B' F_x + s + F_t + a' F_x + \text{tr}(N F_{xx}) / 2.$$

Supposons maintenant que la relation (4) est vérifiée. Alors si l'on pose

$$(13) \quad F(x, t) = -c \ln \{G(x, t)\} + x'(t)v,$$

où  $v$  est un vecteur constant dans  $R^n$  tel que

$$(14) \quad B'v = -Qr,$$

et si l'on prend

$$(15) \quad s = -a'v,$$

on trouve que l'équation (12) devient

$$(16) \quad 0 = (G/c)[v'Nv/2c + r'Qr/2 + r'B'v] + G_t + a'G_x + \text{tr}(NG_{xx})/2.$$

Finalement, la relation (14) implique aussi que le terme entre crochets dans l'équation (16) est nul, de sorte que l'on peut écrire

$$(17) \quad 0 = G_t + a'G_x + \text{tr}(NG_{xx})/2.$$

C'est-à-dire,  $G$  satisfait à l'équation inverse de Kolmogorov pour le processus non commandé

$$(18) \quad dx(t) = a(x, t) dt + dW(t).$$

De plus, la condition frontière (10) devient

$$(19) \quad G(x, t) = \exp[-(K(x, t) - x'(t)v)/c] \quad ((x, t) \in D).$$

Donc, si le processus non commandé (18), parti de  $(x, t)$ , est certain d'atteindre la région d'arrêt  $D$ , on a le théorème qui suit.

*Théorème.* Considérons le système dynamique stochastique défini par (1) et la fonction de coût  $J$  définie à l'équation (8). Supposons qu'il existe une constante positive  $c$  telle que la relation (4) est vérifiée et que le processus non commandé, en partant de  $(x, t)$ , est certain d'atteindre la région d'arrêt  $D$ . Alors si l'on choisit  $r$  et  $s$  tels que

$$(20) \quad \text{et} \quad \begin{aligned} Qr &= -B'v \\ s &= -a'v, \end{aligned}$$

où  $v$  est un vecteur constant quelconque dans  $R^n$ , la commande optimale est donnée par

$$(21) \quad u^* = -[Q^{-1}B'F_x + r],$$

où  $F$  s'obtient de

$$(22) \quad \exp\{-[F(x, t) - x'v]/c\} = E_{(x,t)}\{\exp[(-K(x(\tau), \tau) + x'(\tau)v)/c]\}.$$

Dans l'équation (22), l'espérance mathématique est par rapport au temps  $\tau$  et à l'endroit  $x(\tau)$  où le processus non commandé (18), à partir de  $(x, t)$ , frappe la frontière de la région de continuation  $C$  pour la première fois.

### 3. Le cas invariant dans le temps

Lorsque  $a, B$  et  $N$  dans le système défini par (1), et  $Q, r, s$  et  $K$  dans la fonction de coût  $J$  sont indépendants du temps, on peut généraliser le théorème de la section précédente. On obtient le corollaire suivant (voir Lefebvre (1983)).

*Corollaire.* Sous les mêmes hypothèses que pour le théorème, si le problème est invariant dans le temps alors la fonction de coût

$$(23) \quad J_1 = \int^T [u'Qu/2 + u'Qr + s + p] dt + K[x(T)],$$

où  $p$  est un paramètre réel, est minimisée par la commande

$$(24) \quad u^* = -[Q^{-1}B'F_x + r].$$

La fonction  $F$  est obtenue à partir de l'équation

$$(25) \quad \exp\{-[F(x) - x'v]/c\} = E_x\{\exp[(-p\tau - K(x(\tau)) + x'(\tau)v)/c]\}.$$

*Preuve.* Dans ce cas l'équation (17) devient

$$(26) \quad 0 = -p(G/c) + a'G_x + \text{tr}(NG_{xx})/2 \quad (x \in C)$$

avec la condition frontière

$$(27) \quad G(x) = \exp [(-K(x) + x'v)/c] \quad (x \in D),$$

et l'on a

$$(28) \quad F(x) = -c \ln \{G(x)\} + x'v.$$

De là, on déduit l'interprétation (25).

Le théorème et le corollaire nous montrent que le résultat de Whittle peut être étendu au cas où la variable d'état  $x$  apparaît explicitement dans la fonction de coût. Il s'agit de poser la fonction  $s$  dans la fonction  $J$  (ou  $J_1$ ) égale à  $-a'v$ , où  $v$  est un vecteur constant dans  $R^n$ . Par exemple, dans le cas linéaire

$$(29) \quad dx(t) = (Ax(t) + Bu(t)) dt + dW(t),$$

la fonction  $s$  est simplement une combinaison linéaire des composantes  $x_i(t)$  de la variable d'état  $x(t)$ .

Le cas invariant dans le temps est intéressant car, selon la valeur du paramètre  $p$ , notre but peut être soit de quitter la région de continuation le plus rapidement possible, soit de maximiser le temps que le processus passe dans la région de continuation. Dans les deux cas, il faut bien sûr prendre les coûts de commande en considération.

Finalement, mentionnons que si l'on peut choisir  $c$  égal à 1 dans la relation (4), c'est-à-dire si l'on a

$$(30) \quad N = BQ^{-1}B',$$

alors un théorème semblable à celui obtenu dans cette note, mais en utilisant une fonction de coût exponentielle négative, peut être démontré (voir Lefebvre (1986)).

## References

- LEFEBVRE, M. (1983) *Models of Hazard Survival*. Thèse de doctorat, Université de Cambridge, Cambridge, Angleterre.
- LEFEBVRE, M. (1986) Optimal stochastic control of a class of processes with an exponential cost function. *Ann. sc. math. Québec* **10**, (2), 181–187.
- WHITTLE, P. (1982) *Optimization over Time*, Vol. I. Wiley, Chichester.
- WHITTLE, P. ET GAIT, P. A. (1970) Reduction of a class of stochastic control problems. *J. Inst. Maths Applics* **6**, 131–140.