Optimal stopping and Gittins indices for piecewise deterministic evolution processes

Max-Olivier Hongler
Département de Microtechnique, (DMT)
Institut de Production Microtechnique, (IPM) *
E.P.F.L.
CH-1015 LAUSANNE.

March 1, 1999

Abstract

An optimal stopping problem involving a piecewise deterministic evolution processes is explicitly solved using the method of quasi-variational inequalities. The explicit solution derived offer the possibility to explicitly discuss the associated dynamic allocation problems by means of the Gittins indices.

Keywords: Dynamic allocation of jobs, optimal stopping, piecewise-deterministic processes, continuous time Gittins indices.

1 Introduction

In a very elegant contribution, I. Karatzas, [1], was able to explicitly derive the Gittins indices associated to projects which time evolution is given by diffusion processes. The result was obtained with the use of variation inequalities, a method pioneered by A. Bensoussan and P. L. Lyons, [2]. While in [1], the optimal stopping problems are restricted

*dirigé par le Professeur Jacques Jacot

to diffusive processes, the present note examines the behavior arising when a class of non-diffusive processes is introduced. The class of processes addressed here belong to the piecewise deterministic evolutions described in [4]. This class of processes arises for instance in production engineering when the flow of parts, delivered by a failure prone machine, is assimilated to a fluid, [5]. The processes $X(t)$ considered here are in fact solutions of stochastic differential equations in which the noise source $I(t)$ is a markovian alternating renewal process. For these processes, we observe that several calculations performed in [1] can be used rather directly. Observing that the process $X(t)$ alone is not markovian but the couple $\{X(t), I(t)\}$ is, it is clear that the optimal stopping time will explicitly depend on the initial state of the noise. Despite to this additional complexity, it is nevertheless simple to solve the problem completely.

Our paper is organized as follows. In section 1, we explicitly solve the optimal stopping problem for a scalar piecewise deterministic evolution with a retirement option. In section 2 the dynamic allocation problem is addressed. Section 3 mentions natural extensions of this study together with a short conclusion.

2 Optimal stopping for a simple, non-diffusive, stochastic process

Let us consider the scalar time evolution for $t \geq 0$ given by:

$$\frac{d}{dt}X(t) = I(t), \quad X(0) = x, \quad I(0) = i, \quad \text{and} \quad X(t) \in R, . \tag{1}$$

where $I(t)$ is an alternating renewal process taking the alternate constant values U, (up) and D, (down). Assume that the sojourn times in the U and D phases are exponentially distributed with parameters λ and μ respectively. We shall focus our attention to the cases for which we have:

$$D\lambda + U\mu > 0 \text{ with } D < 0 \text{ and } D + U > 0. \tag{2}$$

so that the asymptotic average of $X(t)$ increases with time.
Further, we introduce an instant reward function $h(x)$ which is a twice differentiable function. The function $h(x)$ is monotone increasing. Additionally, it has the following asymptotic properties:

$$\lim_{x\to\infty} h(x) = \alpha K, \quad \lim_{x\to-\infty} h(x) = \alpha k, \tag{3}$$

with $\alpha > 0$ and:

$$\lim_{|x|\to\infty} \frac{d}{dx} h(x) = 0. \tag{4}$$

Consider the cumulate reward $J(x, i)$ defined by:

$$J_{(x,i)} = E_{(x,i)} \int_0^{\tau} h\left(X(s)\right) e^{-\alpha s} ds + m e^{-\alpha\tau}, \quad m \in R^{+}. \tag{5}$$

where τ is a stopping time and m a terminal reward. The notation $E_{(x,i)}$ stands for the mathematical expectation when the initial conditions are $X(t = 0) = x$ and $I(t = 0) = i$. The problem is to find an optimal stopping time τ^* which maximizes the expected reward $J_{(x,i)}$.

To solve this problem, we use, following [1] and [2], the method of variation inequalities. This method is based on the observation that when the process $X(t), I(t)$ given in Eq.(1) is in a given state, we can either wait a certain period of time before stopping or stop immediately. The resulting rewards are calculated by assuming that the subsequent decisions are optimal. Accordingly, we can write:

- i) If at time t, the decision is to wait for an (infinitesimal) time ξ before stopping, the optimal reward $J_{(X(t),I(t))}$ is greater than the running reward from t to $t+\xi$ and the optimal reward from time $t+\xi$ onwards. This yields the inequality:

$$J_{(X(t),I(t))} \geq \int_t^{t+\xi} e^{-\alpha(s-t)} h\left(X(s)\right) ds + e^{-\alpha\xi} J_{(X(t+\xi),I(t+\xi))}. \tag{6}$$

- ii) If the decision is to stop immediately, the reward is precisely: $J_{(X(t),I(t))} = m$

Let us focus on case i). We have to consider two possible rewards depending on the realizations of the process $I(t)$, namely: $I = U$ or $I = D$. To this two possibilities, we associate the rewards functions $J_{(x,U)}$ and $J_{(x,D)}$ respectively. With this notation and the markov character of the alternating process $I(t)$, the first order time expansion of Eq.(6) yields:

$$0 \geq h(x) - (\alpha + \lambda) J_{(x,U)} + U \frac{d}{dx} J_{(x,U)} + \lambda J_{(x,D)} \tag{7}$$

and

$$0 \geq h(x) - (\alpha + \mu) J_{(x,D)} + D \frac{d}{dx} J_{(x,D)} + \mu J_{(x,U)}. \tag{8}$$

The increasing nature of $h(x)$ suggests that the continuation regions of the stopping problem will be formed by two open intervals:

$\mathcal{O}_1 = (b_d, \infty]$ and $\mathcal{O}_2 = (b_u, \infty]$ which will be respectively relevant for the situations given by $I(0) = D$ or $I(0) = U$.
To calculate the position of the boundaries b_d and b_u, we consider the equalities in Eqs.(7) and (8) and by a direct elimination of $J_{(x,U)}$ and $J_{(x,D)}$ in Eqs.(7) and (8), we respectively obtain:

$$\mathcal{A}\left(J_{(x,U)}\right) + h_D(x) = 0 \tag{9}$$

and

$$\mathcal{A}\left(J_{(x,D)}\right) + h_U(x) = 0, \tag{10}$$

where the notation $\frac{d}{dx}h(x) = h'(x)$ is adopted and $\mathcal{A}$ stands for the second order differential operator:

$$\mathcal{A}(\cdot) = -UD\frac{d^2}{dx^2}(\cdot) + \delta\frac{d}{dx}(\cdot) - \alpha\rho(\cdot) \tag{11}$$

with

$$\delta = D(\alpha + \lambda) + U(\alpha + \mu) > 0, \quad \text{and} \quad \rho = (\alpha + \mu + \lambda). \tag{12}$$

and

$$h_D(x) = \rho h(x) - Dh'(x), \quad \text{and} \quad h_U(x) = \rho h(x) - Uh'(x). \tag{13}$$

Note that $\delta > 0$ follows directly from the definition Eq.(2).

From Eqs.(9) and (10), the variation inequalities takes the form, (see [1], [2]):

$$\left[\mathcal{A}_U J_{(x,U)} + h_D(x) = 0\right] \cdot \left[J_{(x,C)} - m = 0\right] = 0, \tag{14}$$

with $J_{(x,U)} \geq m, \ \forall x \in R$; similarly, we have:

$$\left[\mathcal{A}_D J_{(x,D)} + h_U(x) = 0\right] \cdot \left[J_{(x,D)} - m = 0\right] = 0, \tag{15}$$

again with $J_{(x,D)} \geq m, \ \forall x \in R$.
Soon, it will become clear that the form of the inhomogeneities in Eqs.(9) and (10), suggests to introduce a subclass of functions $\mathcal{H}$ defined by:

<u>Definition</u>. A function $h(x)$ with the properties given by Eqs.(3) and (4) belongs to the class $\mathcal{H}$ if in addition it satisfies simultaneously both inequalities:

$$h'_U(x) = \rho h'(x) - Uh''x \geq 0 \quad \text{and} \quad h'_D(x) = \rho h'(x) - Dh''x \geq 0. \tag{16}$$

Hence for functions $h(x) \in \mathcal{H}$, the inhomogeneities in Eqs.(9) and (10) are strictly increasing functions and we can directly follow the procedure exposed in [1]. Accordingly, the solutions of Eqs.(14) and (15) can be found explicitly. For $k \leq m \leq K$, they respectively read:

$$J_{(x,D)} = \begin{cases} A_d \exp\{-\gamma_+ x\} + p_d(x), & \text{if } x > b_d; \\ m & \text{if } x \leq b_d \end{cases} \tag{17}$$

and

$$J_{(x,U)} = \begin{cases} A_u \exp\{-\gamma_+ x\} + p_u(x), & \text{if } x > b_u; \\ m & \text{if } x \leq b_u \end{cases} \tag{18}$$

with A_d and A_u being integration constants and:

$$\gamma_\pm = \frac{-1}{DU}\left[\sqrt{\delta^2 - 2DU\rho} \pm \delta\right], \tag{19}$$

$$p_d(x) = \mathcal{N}\left[e^{-\gamma_+ x}\int_{-\infty}^{x} h_U(y)e^{\gamma_+ y}dy + e^{-\gamma_- x}\int_{-\infty}^{x} h_U(y)e^{\gamma_- y}dy\right], \tag{20}$$

$$p_u(x) = \mathcal{N}\left[e^{-\gamma_+ x}\int_{-\infty}^{x} h_D(y)e^{\gamma_+ y}dy + e^{-\gamma_- x}\int_{-\infty}^{x} h_D(y)e^{\gamma_- y}dy\right] \tag{21}$$

with

$$\mathcal{N} = \left[-DU(\gamma_+ + \gamma_-)\right]^{-1}. \tag{22}$$

Note that the functions $p_d(x)$ and $p_u(x)$ are particular solutions of Eqs.(9) and (10). They indeed result when stopping is done at time $\tau = \infty$, (i.e. no stopping). Accordingly, we clearly have $p_u(x) > k$ and $p_d(x) > k$.

Finally, the parameters b_u and b_d together with the integration constants A_u and A_d will be determined by the relations:

$$\frac{d}{dx}J(x,D)\mid_{x=b_d} = 0 \quad \text{and} \quad J(b_d, D) = m \quad \Rightarrow$$

$$\begin{cases} M_d(b_d) = m \\ A_d = \frac{p_d'(b_d)}{\gamma_+}\exp(\gamma_+ b_d) \end{cases} \tag{23}$$

and similarly:

$$\frac{d}{dx}J(x,U)\mid_{x=b_u} = 0 \quad \text{and} \quad J(b_u, U) = m \quad \Rightarrow$$

$$\begin{cases} M_u(b_u) = m \\ A_u = \frac{p_u'(b_u)}{\gamma_+}\exp(\gamma_+ b_u) \end{cases} \tag{24}$$

We explicitly obtain:

$$M_u(x) = M_{diff}(x) - D\int_0^\infty h'(x+\frac{z}{\gamma_-})e^{-z}dz, \tag{25}$$

$$M_d(x) = M_{diff}(x) - U\int_0^\infty h'(x+\frac{z}{\gamma_-})e^{-z}dz, \tag{26}$$

and $M_{diff}(x)$ being the index calculated in [1], namely:

$$M_{diff}(x) = \frac{1}{\alpha}\int_0^\infty h(x+\frac{z}{\gamma_-})e^{-z}dz. \tag{27}$$

Note that when $h(x) \in \mathcal{H}$ as defined by Eq.(16), both $M_d(x)$ and $M_u(x)$ are strictly increasing, (see Eq. (3.19) in [1] for the direct demonstration). Hence, the parameters b_d and b_u are the unique solutions of Eqs.(**??**) and (24) respectively.

Note that the following properties hold:

$$\lim_{x\to\infty} M_u(x) = \lim_{x\to\infty} M_d(x) = \lim_{x\to\infty} M_{diff}(x) = K \tag{28}$$

and similarly:

$$\lim_{x\to-\infty} M_u(x) = \lim_{x\to-\infty} M_d(x) = \lim_{x\to-\infty} M_{diff}(x) = k. \tag{29}$$

The behaviors given by Eqs.(28) and (29) are clearly consistent with intuition. Moreover, we also directly observe that:

$$\Delta(x) = M_u(x) - M_d(x) = (U-D)\int_0^\infty h'(x+\frac{z}{\gamma_-})e^{-z}dz \geq 0, \tag{30}$$

with equality in Eq.(30) being realized when $\mid x \mid\to\infty$

It is important to realize that the functions $M_d(x)$ and $M_u(x)$ represents the smallest values of the terminal reward m which makes immediate stopping profitable if the process is in the states (x, D) and (x, U) respectively. This is precisely the definition of the dynamic allocation index introduced by Gittins, [3]. In view of the above results, we can now formulate:

<u>Theorem</u>. Assume that the instant cost function $h(x)$ belong to the class $\mathcal{H}$ given by Eq.(16). Then for the initial conditions $X(t=0) = x$ and $I(0) = D$ respectively $I(0) = U$, the optimal stopping times are $\tau_d^*(x, m)$ and $\tau_u^*(x, m)$ read:

- a) $k < m < K$.

$$\tau_d^*(x,m) = \inf\{t \geq 0; X(t) \leq b_d\} \tag{31}$$

and

$$\tau_u^*(x,m) = \inf\{t \geq 0; X(t) \leq b_u\} \tag{32}$$

where the thresholds b_d and b_u are implicitely given in Eqs.(23) and (24).

- *b)* $m > K$.

In this case we have: $b_d = b_u = \infty$ and therefore $\tau_d^*(x,m) = \tau_u^*(x,m) = 0$.

In this case, the associated cost functions read $J(x,U) \equiv J(x,D) = m$

- *c)* $m < k$. In this cases we have: $b_d = b_u = -\infty$ and hence $\tau_d^*(x,m) = \tau_u^*(x,m) = \infty$.

In this case, the associated cost functions read:

$$J(x,D) = p_d(x) > k \quad \text{and} \quad J(x,U) = p_u(x) > k.$$

Remark: When $k < m < K$, Eqs.(32) and (31) implies that at the stopping time $\tau_d^*(x,m)$ or $\tau_u^*(x,m)$, the process $I(t)$ will be in its D phase. We shall namely have:

$$I(\tau_d^*(x,m) = D \quad \text{or} \quad I(\tau_u^*(x,m) = D \tag{33}$$

3 Dynamic allocation problem

Let us now consider the dynamic system characterized by:

$$\frac{d}{dt}X_j(t) = \mathbf{1}_{[\chi(t)=j]} I_j(t), \quad \chi, j = 1,2,...,d. \tag{34}$$

where $\mathbf{1}_{[\chi(t)=j]}$ is the indicator function, (i.e. $\chi(t) \in \{1,2,,...d\}$) and $I_j(t)$, $j \in \{1,2,...d\}$ are independent alternating renewal processes similar to the one introduced in Eq.(1) with the states U_j and D_j, $j \in \{1,2,...d\}$. In other words, the process $\chi(t)$ indicates which project is engaged at time t. Remember that when a project is not "played" it remains "frozen". The initial conditions for Eq.(34) read:

$$\vec{x} = (X_1(t=0), X_2(t=0), ..., X_d(t=0))$$

and

$$\vec{i} = (I_1(t=0), I_2(t=0), ..., I_d(t=0)). \tag{35}$$

The problem is to find an optimal allocation policy $\mathcal{P}^* = \{\chi^*(t)\}$ which maximizes the expected reward given by:

$$J(\vec{x}, \vec{i}; \mathcal{P}^*) = E_{\vec{x},\vec{i}} \int_0^\infty e^{-\alpha t} h\left(\chi^*(t), X_{\chi^*(t)}(t)\right) dt. \tag{36}$$

The reward functions $h(i, x)$, for $i = 1, 2, ..., d$, are assumed to belong the the class $\mathcal{H}$ for $j = 1, 2, ..., d$. If $\chi(t) = j$, one acquires an instant reward $h(j, X_j(t))$ per unit time, discounted by a factor $e^{-\alpha t}$.
Using the fact M_d and M_u are dynamic allocation indices, it follows directly that the optimal allocation policy $\mathcal{P}^*$ will be given by a "play the leader policy" of the form:

$$\chi^*(t) = \min_{\{1 \leq q \leq d\}} \{\mathcal{M}_q (X_q(t))\}, \tag{37}$$

with

$$\mathcal{M}_q (X_q(t)) = \begin{cases} \max_{1 \leq j \leq d} M_{(j,d)} (X_j(t)) & \text{if } I(t) = D; \\ \max_{1 \leq j \leq d} M_{(j,u)} (X_j(t)) & \text{if } I(t) = U \end{cases} \tag{38}$$

4 Conclusion and perspectives

Quite few are the random evolution processes for which an explicit solution of the optimal stopping problems of the type Eq.(5) are possible. Such class of models are worth to be worked out. This is even more so when the underlying dynamics is intimately related to actual applications. The stochastic differential equation of the type given by Eq.(1) arises naturally in the fluid modeling of the production flows delivered by failure prone machines. Hence, using the Gittins indices, the model discussed here offer several possibilities to explicitly discuss the dynamic allocation problem for flexible production flows in the fluid modeling context.
Among several possibilities to extent the study proposed in this short and prospective note, we would like to mention two research directions:

- a) We can relax the markovian character of alternating renewal process which models the noise source in (1). This can be achieved by introducing the method of the supplementary variable commonly used in the context of queueing theory

- b) The class $\mathcal{H}$ of instant reward function $h(x)$ defined by Eq.(13) should be enlarge to include all strictly increasing bounded functions as considered in [1].

Acknowledgement R. Dalang and T. Mountford are thanked for their stimulating interest in this problem.

5 References

[1]. I. Karatzas. "Gittins indices in the dynamic allocation problem for diffusion processes". Ann. of Probab. **12**, (1984), 173-192.

[2]. A. Bensoussan and J. L. Lyons. *"Applications des Inéquations Variationnelles en Contrôle Stochastique"*. Dunod, (1978).

[3]. J. C. Gittins. "Bandit processes and dynamic allocation indices", (with discussion). J. Roy. Stat. Soc. Ser. **B 41**, (1979), 148-177.

[4]. M. A Pinsky. *Lectures on Random Evolution.* World Scientific, (1991).

[5] Ph. Ciprut, M.-O. Hongler and Y. Salama. " On the variance of the production output of transfer lines". IEEE Trans. on Robotics and Automation, in press.