ON THE VALUE FOR OLA-OPTIMAL STOPPING PROBLEM BY POTENTIAL THEORETIC METHOD

Masami Yasuda

Summary. The stopping problem on Markov process with OLA(One-stage Look Ahead) policy is considered. Its associated optimal value could be expressed explicitly by a potential for a charge of the positive part of the difference between the immediate reward and the one-period-after reward. As application to the best choice problem, the optimal value of three problems: the classical secretary problem, the problem with a refusal probability and the one with random number of objects are calculated.

1. Introduction. The optimal stopping problem is a special case of Markov decision processes. The decision maker can either select to stop, in which case he receives reward and the process terminates, or to pay cost and continue observing the state. If the decision is made to continue, then he proceeds to the next state according to the given transition probability. The objective is to choose the policy which maximizes the expected value. A policy for the decision process means to take the adaptation of a stopping time of the process.

Let x_n, $n=0,1,2,..$ be a Markov chain over a state space S in R^1. We assume that S is countable, but this is inessential for our discussion. In the last section 3 the cases of the unit interval are considered. The optimal stopping problem is to find a stopping time τ which maximizes the expectation of payoff $v(i;\tau)$ starting at i. Let us denote the optimal value by

$$(1.1)\qquad v(i) = \sup_{\tau<\infty} v(i;\tau) = \sup_{\tau<\infty} E[r(x_\tau) - \sum_{n=0}^{\tau-1} c(x_n) \mid x_0=i], \qquad i\in S$$

where $r(i)$ means an immediate reward and $c(i)$ a paying cost. The admissible class of the policies is the set of all finite stopping times. The detailed analyses are discussed by many authors such as Chow/Robbins/Siegmund[1], Shiryaev[11] and so on.

Consider the set of states for which stopping immediately is at least as good as stopping after exactly one more period. Denote this

---

Lecture Notes in Mathematics, 1299,
Probability Theory and Mathematical Statistics,
S.Watanabe and Yu.V.Prokhorov(Eds.),
Springer-Verlag, Berlin, 1988.

set by

$$(1.2)\qquad B = \{i\in S \;;\; r(i) \geq Pr(i) - c(i)\}$$

where $Pr(i) = \sum_{j\in S} P(i,j)r(j)$ and $P(i,j)$, $i,j\in S$, is a stationary transition probability. The policy, defined by the hitting time of this set B, is called a "One-stage Look Ahead"(abridged by OLA) policy by Ross[10], since it compares stopping immediately with stopping after one period. Under certain conditions, He shows that the OLA policy is optimal. The OLA policy is useful for many problems and also extended to the continuous parameter process by Prabhu[8].

Our aim of this note is to obtain, by applying the potential operator, the explicit optimal value associated with the OLA policy for stopping problems. With this result, we will give, in the section 3, the explicit solution of the various versions for the best choice problem in the asymptotic form. The first is the classical problem and the second is the case with the refusal probability. Also the solution for the case of random number of objects is calculated. The last case was reduced to a functional optimality equation by an ad hoc method in Yasuda[13].

The motivation for this approach arose in connection with the results of Darling[2] and that of Hordijk[5]. The former gave the upper bound of the optimal value by a potential operator and the latter gave a sufficient condition to find an optimal stopping time.

2. Markov potential and the optimal value. For a transition probability $P=P(i,j),i,j\in S$, a function $f=f(i)$, $i\in S$, is called a charge if $\lim\,[(I+P+...+P^k)f](i)$ for each $i\in S$, exists and is finite-valued. Function $g=g(i), i\in S$ is a potential if there exists a charge f such that $g(i) = \lim\,[(I+P+...+P^k)f](i)$, $i\in S$. We shall use the notation $g(i)= Nf(i)$, $i\in S$.

The relation between Markov Potential Theory and Dynamic Programming or Markov Decision Process is discussed by Hordijk[5]. Since, as in the section 1, the optimal stopping problem is a special case of Markov Decision Process, the optimality equation of the stopping problem is reduced as follows.

$$(2.1)\qquad v(i) = \max\{r(i),\ Pv(i) - c(i)\}, \qquad i\in S.$$

Fundamental in such an investigation of Dynamic Programming is the uniqueness of the equation (2.1) and the determination of the optimal policy and the optimal value.

In this note we consider the optimal stopping problem, for which the OLA policy is optimal. To give a sufficient condition, we prepare the next two propositions. The subset B of S in (1.2) is closed if

(2.2) $\qquad P(i,j) = 0 \quad$ for $i \in B,\ j \in \bar{B}$

where $\bar{B}$ denotes the complement of the set B. The process is stable if the sequence defined by $v_0(i) = r(i)$,

$$v_n(i) = \max\{r(i),\ Pv_{n-1}(i) - c(i)\} \text{ for } n \geq 1,$$

converges uniformly and $\lim_{n\to\infty} v_n(i) = v(i)$ for $i \in S$.

Proposition 2.1 (Ross[10]). If the process is stable and the set B is closed, then the OLA policy is optimal.

Although this situation occurs in many applications and is useful to determine the stopping region, the proposition does not state the optimal value. The stability is somewhat less satisfactory to check in the application because the optimal value is unknown. So we impose assumption on a potential of the chain. Our aim is to calculate the optimal expected value under the closedness and the following equalization assumption in stead of the stability assumption, and express it explicitly by using a potential. We call the problem where the OLA policy is optimal as the OLA-optimal stopping problem.

Proposition 2.2 (Hordijk[5]). Suppose that

(2.3) $\qquad v(i) = Pv(i) - c(i) \quad$ when $\ i \notin \Gamma = \{i \in S\ ;\ v(i) = r(i)\}$.

If the value function is a potential, then the hitting time of set Γ becomes optimal.

Proof. This is a special case in Theorem 4.1 of Hordijk[5]. Q.E.D.

Let P_A denote the restriction to a subset A of the transition probability P, i.e.,

(2.4) $\qquad P_A(i,j) = P(i,j)1_A(j) \quad$ if $i,j \in S$,

where 1_A denotes the indicator of set A.

When the stability is dropped, as Ross shows, a stopping problem does not imply the optimality of the OLA policy. So we must impose a condition so as to preserve the OLA principle. The condition of equalizing for the reward function due to a potential notion is considered.

Assumption 2.1. We assume that

(2.5) $\qquad \lim_{k\to\infty}[(P_{\bar{B}})^k r](i) = 0 \qquad$ for $i \in \bar{B}$

where $\bar{B}$ denotes the complement of the set B defined by (1.2), and that the potential for $P_B r - c$ is finite-valued with respect to $P_{\bar{B}}$, that is,

(2.6) $\qquad [N_{\bar{B}}(P_B r - c)](i) < \infty \quad$ for $i \in \bar{B}$.

The assumption (2.5) is equivalent to

$$\lim_{k\to\infty} E[r(x_k)1_{\{\tau>k\}};x_n \notin B, n=1,\ldots,k-1] = 0$$

where τ is the hitting time of B. The property $\lim_{k\to\infty}(P_{\bar{B}})^k v = 0$ for the optimal value is called equalizing in Optimal Gambling(Dubins /Savage[3] or Hordijk[5]). One might say that here the actually received in the time period up to N and the promised earnings equalize as N tends to infinity. If the optimal value satisfies this property, the assumption (2.5) holds since $v(i) \geq r(i)$, $i \in S$ by (2.1).

Theorem 2.1. Under the assumptions (2.5) and (2.6), the OLA policy τ_B is optimal, that is, the optimal policy is the hitting time of B. The optimal value $v(i) = v(i;\tau_B)$ is given by

(2.7) $$v(i) = r(i) + N(Pr - r - c)^+(i) = \begin{cases} r(i) & \text{on } B \\ N_{\bar{B}}(P_B r - c)(i) & \text{on } \bar{B} \end{cases}$$

where + is the positive part of the function, and N and $N_{\bar{B}}$ are potentials with respect to P and $P_{\bar{B}}$ respectively.

Proof. The proof for the optimality of the OLA policy is similar to Ross[10] and Hordijk[5]. In generally if $v(i)$, $i \in S$ is a solution of the optimality equation(2.1), then $v(i) \geq r(i)$ and $v(i) \geq Pv(i) - c(i)$ for $i \in S$. So

$$v(i) \geq P^n r(i) - \sum_{k=0}^{n-1} P^k c(i) \text{ for each n.}$$

It yields that

$$v(i) \geq v(i;\tau) \text{ for any policy } \tau < \infty.$$

Therefore it is sufficient to assert the followings to show the optimality. One is that, for the OLA policy τ_B, its value equals the right hand side of (2.7), that is, $v(i;\tau_B) = r(i)$ on B and $N_{\bar{B}}(P_B r - c)(i)$ on $\bar{B}$, and the second is that it satisfies the optimality equation (2.1).

Because of the definition of the OLA policy, $v(i) = r(i)$ on B and $v(i) = Pv(i) - c(i) = P_{\bar{B}}v(i) + P_B r(i) - c(i)$ on $\bar{B}$. Hence we get the first assertion immediately.

Nextly we show that it satisfies the optimality equation (2.1). From $Pf(i) = P_B f(i)$, $i \in B$ for any $f = f(i)$, we have

$$Pv(i) - c(i) = P_B v(i) - c(i) = P_B r(i) - c(i) = Pr(i) - c(i)$$

on B. Hence

$$\max\{r(i),\ Pv(i) - c(i)\}$$
$$= \max\{r(i),\ Pr(i) - c(i)\}$$
$$= r(i) \quad \text{for } i \in B.$$

On the other hand, by substitution of (2.7),

$$Pv(i) - c(i) = P_{\bar{B}}v(i) + P_B v(i) - c(i)$$
$$= [P_{\bar{B}}N_{\bar{B}}(P_B r - c)](i) + P_B r(i) - c(i)$$
$$= [N_{\bar{B}}(P_B r - c)](i), \quad i \in S.$$

On $i \in \bar{B}$, that

$$r(i) < Pr(i) - c(i) = P_{\bar{B}}r(i) + P_B r(i) - c(i)$$

implies $r(i) - P_{\bar{B}}r(i) < P_B r(i) - c(i)$. And, by the assumption (2.6), $r(i) \leq [N_{\bar{B}}(P_B r - c)](i)$ for $i \in \bar{B}$. Combining the above assertions,

$$\max\{r(i),\ Pv(i) - c(i)\}$$
$$= \max\{r(i),\ [N_{\bar{B}}(P_B r - c)](i)\}$$
$$= [N_{\bar{B}}(P_B r - c)](i) \qquad \text{for } i \in \bar{B}.$$

Thus the value $v(i) = v(i;\tau_B)$ satisfies the optimality equation.

It now remains to calculate the potential $N(Pr-r-c)^+(i)$, $i \in S$. From the definition of the set B in (1.2), we have $(Pr-r-c)^+(i) = 0$ on B. So the support of charge is the complement of B and hence

$$N(Pr-r-c)^+(i) = 0 \quad \text{on } B.$$

On other hand, since $(Pr-r-c)^+(i) = (Pr-r-c)(i) = (P_{\bar{B}}r + P_B r - r - c)(i)$ for $i \in \bar{B}$, we have that

$$P(Pr-r-c)^+(i) = P_{\bar{B}}(Pr-r-c)(i)$$
$$= ((P_{\bar{B}})^2 r + P_{\bar{B}}P_B r - P_{\bar{B}}r - P_{\bar{B}}c)(i), \quad i \in \bar{B}.$$

Repeating this procedure to take the expectation up to k times and adding these,

$$(I+P+P^2+\ldots+P^k)(Pr-r-c)^+(i)$$
$$= (P_{\bar{B}})^k r(i) + [((P_{\bar{B}})^{k-1} + (P_{\bar{B}})^{k-2} + \ldots + I_{\bar{B}})(P_B r - c)](i) - r(i), \quad i \in \bar{B}.$$

Hence

$$N(Pr-r-c)^+(i) = [N_{\bar{B}}(P_B r - c)](i) - r(i), \quad i \in \bar{B}$$

follows immediately from the assumptions (2.5) and (2.6). Q.E.D.

We remark that the upper bound on the optimal value in Theorem 3.6 of Darling[2] equals exactly the optimal value in this case. That is, the bound holds with equality when the OLA policy is optimal and it is equalizing. This explicit solution and the proposition 2.1, 2.2 determine the optimal value and policy completely in the OLA-optimal stopping problem.

3. The Best Choice Problem. In this section we apply the previous method to the typical stopping problem known as the best

choice problem. By taking a scale limit, the asymptotic form of the problem is considered in order to get an analytical explicit solution. In the asymptotic form the state space of the problem is not countable but the unit interval. However immediately the previous method can be applied to the case. Some of results are well known and the optimal values are already obtained by some method, for example, by using differential equation. Here we intend to illustrate it for the application which optimal value can be determined by the unified previous method. Each of these are cases whose problem is the OLA-optimal stopping one.

3.1. The Classical Secretary Problem. The secretary problem, variously called dowry problem or Googol, is an optimal stopping problem based on relative ranks for objects arriving in a random fashion; the objective is to find the stopping rule that maximizes the probability of stopping at the best object of the sequence. The optimality equation for the optimal value $v(i)$, $i \in S$, the maximal probability of win, on the state space $S = \{1,2,..,n\}$ is as follows.

$$v(i) = \max\{i/n,\ i\sum_{j=i+1}^{n} v(j)/((j-1)j)\}, \quad i \in \{1,2,\ldots,n-1\}, \qquad v(n) = 1$$

where there are no costs. This formulation of the problem as Markov Process is given by Dynkin and Yushkevitch[4]. Also refer to Chow/Robbins/Siegmund[1] or Shiryaev[11]. To consider the problem in the asymptotic form, take the scale limit by $i/n \to x$ as $i,n \to \infty$. It becomes that

$$v(x) = \max\{\ x\ ,\ x\int_x^1 v(y)y^{-2}dy\}, \qquad x \in S = [0,1]. \tag{3.1}$$

The solution of (3.1) is well known as

$$v(x) = \begin{cases} e^{-1} & \text{on } [0,\ e^{-1}], \\ x & \text{on } [e^{-1},\ 1]. \end{cases} \tag{3.2}$$

Now we extend the problem by changing the reward function $r(x) = x$, which means probability of choosing the best object in the classical secretary problem, to a general reward $r(x)$. To ensure the OLA policy is still optimal, we assume, for a function $h(x)$ defined by

$$h(x) = r(x) - x\int_x^1 r(y)y^{-2}dy \qquad \text{on } [0,1], \tag{3.3}$$

that

$r(x) = x$

$h(x) = x - x\int_x^1 \frac{dy}{y} = x + x\log x$

$x = e^{-1}$

i) each term of the function $h(x)$ is finite-valued on $[0,1]$, and

ii) it changes its sign from $-$ to $+$ only once as x increases 0 to 1 and so the equation $h(x)=0$ has a unique solution α.

The optimality equation (3.1) with the reward function $r(x)$ is

$$(3.4)\qquad v(x) = \max\{r(x),\ x\int_x^1 v(y)y^{-2}dy\},\qquad x\in[0,1].$$

provided that the underlying Markov chain is unchanged. The solution of this equation (3.4) is given by

$$(3.5)\qquad v(x) = \begin{cases} r(\alpha) & \text{on } [0,\alpha], \\ r(x) & \text{on } [\alpha,1] \end{cases}$$

where α is defined by (3.3ii).

In fact, one can show this (3.5) by applying the previous result. Straightforward calculation yields that the set B becomes $[\alpha,1]$ and it is closed with respect to the transition probability:

$$P(x,dy) = \begin{cases} xy^{-2}dy & \text{for } 0<x<y<1, \\ 0 & \text{otherwise.} \end{cases}$$

We have $P_B r(x) = r(\alpha)(x/\alpha)$ and $(P_{\bar{B}})^n P_B r(x) = r(\alpha)(x/\alpha)\log^n(\alpha/x)/n!$ for $x\in[0,\alpha]$, $n=0,1,2,\ldots$ Hence

$$N_{\bar{B}}P_B r(x) = r(\alpha)(x/\alpha)\{1+\log(\alpha/x)+2^{-1}\log^2(\alpha/x)+\ldots +(n!)^{-1}\log^n(\alpha/x)+\ldots\}$$
$$= r(\alpha)$$

for $x\in[0,\alpha]$ and thus the assumption (2.6) is satisfied. Also we can check the condition of (2.5):

$$(P_{\bar{B}})^n r(x) = x\int_x^{\alpha}\{r(y)y^{-2}\log^{n-1}(y/x)/(n-1)!\}dy$$

tends to zero as $n\longrightarrow\infty$ for $x\in[0,\alpha]$.

3.2. A Problem with Refusal Probability. A variant of the best choice problem is a case with a refusal probability discussed by Smith[12]. We can also formulate the problem as optimal stopping on a Markov chain. The asymptotic form of the transition probability is obtained immediately and we have

$$P(x,dy) = \begin{cases} py^{-1}(x/y)^p dy & 0<x<y<1, \\ 0 & \text{otherwise} \end{cases}$$

where p is a given parameter $0 < p \le 1$, which quantity $1-p$ means a probability of the refusal. When there is no refusal, i.e., $p=1$, it

$r(x)=x$

$h(x) = x - px^p\int_x^1 \frac{1}{y^p}dy$

$= x - px^p\left[\frac{y^{-p+1}}{-p+1}\right]_{y=x}^{y=1}$

$= x - \frac{p}{1-p}x^p(1-x^{-p+1})$

$= x - \frac{p}{1-p}(x^p - x)$

$h'(x)=0$

$1 - \frac{p}{1-p}(x^{p-1}-1)=0$

$\frac{1}{p} = x^{p-1}$

$x^{1-p} = p$

$x = p^{\frac{1}{1-p}}$

$\therefore\ \alpha = p^{\frac{1}{1-p}}$

$V(x) = \begin{cases} p^{\frac{1}{1-p}} & 0<x<\alpha \\ x & \alpha<x<1 \end{cases}$

reduces to the classical secretary problem discussed in the section 3.1. Similarly as before, the optimality equation for the stopping problem with refusal probability p is

$$(3.6)\qquad v(x) = \max\{r(x),\ \int_x^1 py^{-1}(x/y)^p v(y)dy\},\qquad x\in[0,1].$$

Define a function $h(x)$ by

$$(3.7)\qquad h(x) = r(x) - \int_x^1\{py^{-1}(x/y)^p r(y)\}dy,\qquad x\in[0,1].$$

Under the same assumptions as (3.3i) and (3.3ii), we obtain the optimal value with refusal probability as follows.

$$(3.8)\qquad v(x) = \begin{cases} r(\alpha) & \text{on } [0,\alpha], \\ r(x) & \text{on } [\alpha,1] \end{cases}$$

where α is a unique solution of $h(x)=0$ in (3.7).

Another method to solve the best choice problem is given by Mucci[7], which method reduces the value to the solution of a differential equation.

Let

$$V(x) = \int_x^1\{py^{-1}(x/y)^p r(y)\}dy,\qquad x\in[0,1]$$

which means a conditional optimal value. This satisfies

$$(3.9)\qquad \begin{cases} dV(x)/dx = -px^{-1}(r(x)-V(x))^+, & x\in[0,1] \\ V(1)=0. \end{cases}$$

The optimal value at the beginning $v^* = V(0)$ equals $r(\alpha)$.

3.3. A Problem with Random Number of Objects. The discussion of the problem for variant on the random number of objects is given by Presman/Sonin[9]. The random environment in the problem means that there is a distribution $\phi(x)$ over $x\in[0,1]$ which denotes the random number of objects. If we adapt the approach by the differential equation(3.9), the following functional equation is obtained by Yasuda[13]. For $x\in[0,1]$,

$$(3.10)\qquad \begin{cases} dV(x) = V(x)\{1-\phi(x)\}^{-1}d\phi(x) - x^{-1}(R(x)-V(x))^+dx, \\ V(1) = 0 \end{cases}$$

where we set

$\bar{g}(x) > x$

$R(x) = \frac{x}{1-x}(-\log x)$

(3.11) $$R(x) = x(1-\Phi(x))^{-1} \int_x^1 y^{-1} d\Phi(y).$$

When the distribution is absolutely continuous, (3.10) reduces to a differential equation such as (3.9). Let us define

$$g(x) = \int_x^1 y^{-1} d\Phi(y) \quad \text{and} \quad h(x) = g(x) - \int_x^1 y^{-1} g(y)dy \quad \text{for } x\epsilon[0,1].$$

We assume conditions on the distribution $\Phi(x)$ so that these functions $R(x)$, $g(x)$ and $h(x)$ are well defined. The following result is obtained already by Yasuda[13].

If $h(x)$ changes its sign only once from - to + as x varies from 0 to 1, and if $\Phi(x)$ is continuous for $0 < x < 1$, then the optimal value at the beginning $v^* = V(0)$ is given by

(3.12) $$v^* = (1-\Phi(a))V(a) = ag(a)$$

where a is a unique solution of $h(x)=0$ for $x\epsilon[0,1]$.

This can be also obtained by applying the previous method. Since the optimality equation with random number of objects in the asymptotic form is, for $x\epsilon[0,1]$,

(3.13) $$v(x) = \max[R(x),\ x(1-\Phi(x))^{-1} \int_x^1 (1-\Phi(y))y^{-2}v(y)dy],$$

we have

(3.14) $$v(x) = \begin{cases} R(a) & \text{on } [0,a], \\ R(x) & \text{on } [a,1]. \end{cases}$$

In fact, it is equivalent to the classical secretary problem

$$w(x) = \max[r(x),\ x\int_x^1 y^{-2} w(y)dy)$$

$w(x) = \max\{-x\log x,\ x\int_x^1 \frac{w(y)}{y^2}dy\}$

with the function

(3.15) $$w(x) = (1-\Phi(x))v(x)$$

$w(x) = (1-x)v(x)$

and with the reward function

(3.16) $$r(x) = x\int_x^1 y^{-1} d\Phi(y).$$

$r(x) = -x\log x$

random number
refusal prob

Hence the solution (3.14) is immediately obtained by the result of (3.5). This method is simpler than the ad hoc treatment of the functional equation (3.10).

References

[1] Chow,Y.S., Robbins,H. & Siegmund,D. : Great Expectations: The Theory of Optimal Stopping, Houghton Mifflin, Boston (1971).

[2] Darling, D.A. : Contribution to the optimal stopping problem, Z. Wahr. verw. Gebiete 70 (1985), 525-533.

[3] Dubins,L.E. and Savage,L.J. : How to Gamble if You Must: Inequalities for Stochastic Processes, McGrow-Hill, New York (1965).

[4] Dynkin,E.B. & Yushkevitch,A.A. : Theorems and Problems on Markov Processes, Translation: Plenum Press, New York (1969).

[5] Hordijk, A. : Dynamic Programming and Markov Potential Theory, Mathematisch Centrum, Amsterdam (1974).

[6] Kemeny,J.G., Snell,J.L. & Knapp,A,W. : Denumerable Markov Chains, 2nd eds., Springer-Verlag, Berlin (1976).

[7] Mucci, A.G. : Differential equations and optimal choice problem, Ann. Statist. 1 (1973), 104-113.

[8] Prabhu, N.U. : Stochastic control of queueing systems, Nav. Res. Logist. Q. 21 (1974), 411-418.

[9] Presman,E.L. & Sonin,I.H. : The best choice problem for a random number of objects, T. Prob. Appl. 17 (1972), 657-668.

[10] Ross, S.M. : Introduction to Stochastic Dynamic Programming, Academic Press, New York (1983).

[11] Shiryaev, A.N. : Statistical Sequential Analysis, Translation: American Mathematical Society, Providence (1973).

[12] Smith, M.H. : A secretary problem with uncertain employment, J. Appl. Prob. 12 (1975), 620-624.

[13] Yasuda, M. : Asymptotic results for the best-choice problem with a random number of objects, J. Appl. Prob. 21 (1984), 521-536.

College of General Education
Chiba University
Chiba
260 Japan