The policy iteration method for the optimal stopping of a Markov chain with an application

by

K.M. van Hee

0. Summary

In this paper we study the problem of the optimal stopping of a Markov chain with a countable state space. In each state i the controller receives a reward r(i) if he stops the process or he must pay the cost c(i) otherwise. We show that, under the condition that there exists an optimal stopping rule, the policy iteration method, introduced by Howard, produces a sequence of stopping rules for which the expected return converges to the value function. For random walks on the integers with a special reward and cost structure, we show that the policy iteration method gives the solution of a discrete two point boundary value problem with a free boundary. We give a simple algorithm for the computation of the optimal stopping rule.

1. Introduction

Consider a Markov chain $\{X_n \mid n = 0,1,2,\ldots\}$ defined on the probability space $(\Omega,F,\mathbb{P})$. The state space S is countable. We suppose that $\mathbb{P}[X_0 = i] > 0$ for all $i \in S$. Hence $\mathbb{P}_i[A]$, the conditional probability of $A \in F$ given $X_0 = i$, is defined for all $i \in S$.

On S real functions r and c are defined, where r(i) is the reward if the process is stopped in state i and c(i) is the cost if the process goes on. We consider stopping times T (for a definition see [7]). For a nonnegative function g on S we define

$$\mathbb{E}_i[g(X_T)] := \int_{\{T<\infty\}} g(X_T)d\mathbb{P}_i \ .$$

Condition A. Suppose that the reward function r satisfies

$$\mathbb{E}_i[r^+(X_T)] + \mathbb{E}_i[r^-(X_T)] < \infty$$

for all $i \in S$ and all stopping times T.

(Note that: $r^+(i) := \max\{0,r(i)\}$, $r^-(i) := -\min\{0,r(i)\}$).

Let P be the transition matrix of the Markov chain, with components $P(i,j)$ for $i,j \in S$. If the function c on S is integrable for all $\mathbb{P}_i[.]$, we define the function Pc by

$$Pc(i) := \sum_{j \in S} P(i,j)c(j) ,$$

and with induction, if $P^{n-1}c$ is integrable for all $\mathbb{P}_i[.]$

$$P^n c := P(P^{n-1}c) .$$

We call a function c on S a *charge* (see [3]) if

$$\sum_{n=0}^{\infty} P^n|c| < \infty .$$

(Note that for function v and w on S: $v \le w$ if $v(i) \le w(i)$ for all $i \in S$ and $v < w$ if $v(i) < w(i)$ for all $i \in S$. Further $|v|$ is defined by $|v|(i) := |v(i)|$.

Condition B. Either the cost function c is a charge or r and c are nonnegative, both.

Throughout this paper we shall suppose that conditions A and B hold.

We call a function w on S *c-excessive* with respect to the cost function c if

1) $\quad w \ge -c + Pw$

2) $\quad w \ge -\sum_{n=0}^{\infty} P^n c .$

For a stopping time T the expected return $v_T(i)$, given the starting state i, is defined by

$$v_T(i) := \mathbb{E}_i[r(X_T) - \sum_{n=0}^{T-1} c(X_n)] .$$

The existence of the expected return $v_T(i)$ is guaranteed for all T since $|\mathbb{E}_i[r(X_T)]| < \infty$ for all i and c is either a charge or a nonnegative function. Note that $v_T(i) = -\infty$ is permitted.
The value function $v(i)$ is the supremum over all the stopping times T

$$v(i) := \sup_T v_T(i) .$$

Sometimes we need the following assumption.

Assumption C. There exists an optimal stopping time T^*, i.e. $v_{T^*}(i) = v(i)$ for all $i \in S$.

In the rest of this section we summarize some properties of stopping problems.

1.1. The value function v satisfies the functional equation

$$v(i) = \max\{r(i), -c(i) + \sum_{j\in S} P(i,j).v(j)\}$$

(see [2], [3] or [7]).

1.2. The value function v is the smallest c-excessive function dominating the reward function r (see [2] and [3]).

1.3. If an optimal stopping time exists the entrance time T_Γ in the set $\Gamma := \{i \mid r(i) = v(i)\}$ is optimal (see [2] and [6]).

1.4. If $\sup_{i\in S} |r(i)| < \infty$ and $\inf_{i\in S} c(i) > 0$ then there exists an optimal stopping time (see [2] and [7]).

2. Some preparations

A stopping rule f is a mapping from S to $\{0,1\}$ where $f(i) = 0$ means that the process is stopped in i and $f(i) = 1$ means that the process goes on in state i. The stopping rule f is equivalent with the entrance time T_f in the set $\Gamma_f := \{i \mid f(i) = 0\}$. The expected return under a stopping rule f is indicated by $v_f(i)$.

For a stopping rule f we define

2.1. $D_f := \{i \in S \mid f(i) = 1\}$, the go-ahead set.

$\Gamma_f := S\backslash D_f$, the stopping set.

2.2. P_f is the matrix with components

$$P_f(i,j) := \begin{cases} P(i,j) & \text{if } i \in D_f \\ 0 & \text{otherwise} . \end{cases}$$

2.3. d_f is a function on S with

$$d_f(i) := \begin{cases} r(i) & \text{if } i \in \Gamma_f \\ -c(i) & \text{otherwise} . \end{cases}$$

If assumption C holds, property 1.3 guarantees that the entrance time T_Γ in the set Γ is also optimal. In that case

2.4. $$v(i) = \mathbb{E}_i[r(X_{T_\Gamma}) - \sum_{n=0}^{T_\Gamma - 1} c(X_n)] .$$

According to the stopping time T_Γ we define the stopping rule f_* by

2.5. $$f_*(i) = 0 \quad \text{if and only if } i \in \Gamma .$$

Further let

$$D := S\backslash\Gamma,\ d := d_{f*} \text{ and } \tilde{P} := P_{f*} .$$

Lemma 1. For each stopping rule f with $v_f \geq r$ we have

1) $|v_f(i)| < \infty$

2) $v_f = \sum_{n=0}^{\infty} P_f^n d_f$

3) $\lim_{n\to\infty} P_f^n |d_f| = 0$ (pointwise convergence)

4) $v_f = d_f + P_f v_f$

5) $\lim_{n\to\infty} P_f^n |v_f| = 0$ (pointwise convergence) .

Proof. If r and c are nonnegative we have

$$0 \leq r(i) \leq v_f(i) \leq \mathbb{E}_i[r(X_{T_f})] < \infty \quad \text{for all } i \in S .$$

Since

$$v_f(i) = \mathbb{E}_i[r(X_{T_f})] - \mathbb{E}_i[\sum_{n=0}^{T_f-1} c(X_n)]$$

we may conclude

$$\mathbb{E}_i[\sum_{n=0}^{T_f-1} |c(X_n)|] < \infty \quad \text{for all } i \in S .$$

Note that if c is a charge this also true. Define:

2.6. $$w_f(i) := \mathbb{E}_i[|r(X_{T_f})|] + \mathbb{E}_i[\sum_{n=0}^{T_f-1} |c(X_n)|] .$$

So we have for both cases of B

$$|v_f(i)| \leq w_f(i) < \infty . \quad \text{(statement 1)}$$

We have the following representation

$$w_f = \sum_{n=0}^{\infty} P_f^n |d_f|$$

(note that $P_f^0(i,j) = 1$ if and only if $i = j$) and in the same way, by absolute convergence,

$$v_f = \sum_{n=0}^{\infty} P_f^n d_f . \quad \text{(statement 2)}$$

Because $w_f < \infty$ we may conclude $P_f^n|d_f| \to 0$ for $n \to \infty$ (statement 3)

$$v_f = \sum_{n=0}^{\infty} P_f^n d_f = d_f + \sum_{n=1}^{\infty} P_f^n d_f .$$

Since

$$\sum_{n=1}^{\infty} P_f^n |d_f|$$

is finite we may change the summation order, hence

$$v_f = d_f + P_f \sum_{n=0}^{\infty} P_f^n d_f = d_f + P_f v_f . \quad \text{(statement 4)}$$

In the same way

$$w_f = d_f + P_f w_f .$$

By iterating this equation we get

$$w_f = \sum_{n=0}^{N} P_f^n d_f + P_f^{N+1} w_f$$

from which it follows that $P_f^n w_f$ tends to 0 if n tends to ∞. Because $|v_f| \le w_f$ we have also

$$\lim_{n\to\infty} P_f^n |v_f| = 0 . \quad \text{(statement 5)} \qquad \square$$

<u>Corollary 1</u>. If C hold we have from 2.4 and lemma 1 that

$$|v(i)| < \infty \text{ for all } i \in S \text{ and } \lim_{n\to\infty} \tilde{P}^n |d| = 0 .$$

Define:

$$w := \sum_{n=0}^{\infty} \tilde{P}^n |d| .$$

By lemma 1 we have

2.7. $$\lim_{n\to\infty} \tilde{P}^n w = 0 .$$

In the next section we study expressions like $P_g^k v_f$, where f and g are stopping rules. We shall give sufficient conditions in lemma 2 for the finiteness of these expressions.

<u>Lemma 2</u>. Let f and g are stopping rules. Suppose $v_f \ge r$. Then $P_g^k |v_f|$ is finite for $k = 1,2,3,\ldots$.

<u>Proof</u>. Let $T := T_f + k$. Using the same arguments as in lemma 1, we derive for c a charge:

$$\mathbb{E}_i[|r(X_T)| + \sum_{n=0}^{T-1} |c(X_n)|] < \infty .$$

Note that

$$\mathbb{E}_i[|r(X_T)| + \sum_{n=0}^{T-1} |c(X_n)|] = \sum_{n=0}^{k-1} P^n|c|(i) + P^k w_f(i)$$

(w_f is defined in 2.6).
Hence

$$|P_g^k v_f| \le P_g^k|v_f| \le P^k|v_f| \le P^k w_f < \infty .$$

Now let r and c be nonnegative.
$P^k v_f$ is defined because $v_f \ge r \ge 0$. Hence $P^k v_f \ge P^k r \ge 0$

$$0 \le P^k v_f(i) = \sum_{j\in S} P^k(i,j)\mathbb{E}_j[r(X_{T_f}) - \sum_{n=0}^{T_f-1} c(X_n)] \le$$

$$\le \sum_{j\in S} P^k(i,j)\mathbb{E}_j[r(X_{T_f})] = \mathbb{E}_i[r(X_T)] < \infty .$$

Define vectors c_f and r_f by

$$c_f(i) := c(i) \text{ if } i \in D_f, \quad r_f(i) := r(i) \text{ if } i \in \Gamma_f$$
$$:= 0 \quad \text{otherwise} \qquad := 0 \quad \text{otherwise} .$$

Note that $|d_f| = r_f + c_f$. It is easy to verify that

$$\sum_{j\in S} P^k(i,j)\mathbb{E}_j[r(X_{T_f})] = P^k \sum_{n=0}^{\infty} P_f^n r_f(i)$$

and

$$\sum_{j\in S} P^k(i,j)\mathbb{E}_j[\sum_{n=0}^{T_f-1} c(X_n)] = P^k \sum_{n=0}^{\infty} P_f^n c_f(i) .$$

Hence $P^k w_f = P^k \sum_{n=0}^{\infty} P_f^n\{r_f + c_f\} < \infty$. Reasoning like before, we see that $P_g^k|v_f| < \infty$.□

3. <u>Policy iteration method</u>

Let f be a stopping rule, such that $\sum_{j\in S} P(i,j)v_f(j)$ is defined. For f we define the *improved stopping rule* g by

3.1. $$g(i) := 0 \quad \text{if } r(i) \ge -c(i) + \sum_{j\in S} P(i,j)v_f(j)$$
$$:= 1 \quad \text{otherwise.}$$

Lemma 3. Let g be the improved stopping rule of f and let $v_f \geq r$. Then

1) $D_g \subset D$

2) $v_f \leq d_g + P_g v_f$.

Proof. We first prove 1).
If $g(i) = 1$ then

$$r(i) < -c(i) + \sum_{j\in S} P(i,j)v_f(j) \leq -c(i) + \sum_{j\in S} P(i,j)v(j) \leq v(i)$$

hence

$$D_g = \{i \mid g(i) = 1\} \subset \{i \mid v(i) > r(i)\} = D .$$

We proceed with 2).
Note that $P_g v_f$ is finite (by lemma 2). Let $i \in D_g$ then $g(i) = 1$, $d_g(i) = -c(i)$, $P_g(i,.) = P(i,.)$ and so

$$r(i) < -c(i) + \sum_{j\in S} P(i,j)v_f(j) = d_g(i) + \sum_{j\in S} P_g(i,j)v_f(j) .$$

Since either

$$v_f(i) = -c(i) + \sum_{j\in S} P(i,j)v_f(j)$$

or $v_f(i) = r(i)$ the statement is true for $i \in D_g$.
If $i \in \Gamma_g$ then $g(i) = 0$, $d_g(i) = r(i)$ and $P_g(i,.) = 0$ and since

$$r(i) \geq -c(i) + \sum_{j\in S} P(i,j)v_f(j)$$

it is true for $i \in \Gamma_g$. □

Lemma 4. Assume C. If g is the improved stopping rule of f and if $v_f \geq r$ then $v_g \geq v_f$.

Proof. From lemma 2 it follows that $P_g^k|v_f|$ exists and is finite for all k. By lemma 3 is $v_f \leq d_g + P_g v_f$. Hence

$$\sum_{k=0}^{N} P_g^k v_f \leq \sum_{k=0}^{N} P_g^k d_g + \sum_{k=1}^{N+1} P_g^k v_f$$

and therefore

$$v_f - P_g^{N+1} v_f \leq \sum_{k=0}^{N} P_g^k d_g .$$

We shall prove that $P_g^N v_f^+ \to 0$ for $N \to \infty$. Consider first the case that $r \geq 0$ and $c \geq 0$.
Since $0 \leq r \leq v_f \leq v$ and $D_g \subset D$

$$0 \le P_g^N v_f \le P_g^N v \le \tilde{P}^N v$$

by corollary 1 $\tilde{P}^N v \to$ for $N \to \infty$.

Suppose now that c is a charge:

$$v_f^+ \le v^+ \le w$$

(w is defined in corollary 1) hence

$$P_g^N v_f^+ \le P_g^N w \le \tilde{P}^N w \ .$$

By 2.7

$$\tilde{P}^N w \to 0 \quad \text{for } N \to \infty \ .$$

Therefore

$$v_f \le \sum_{k=0}^{\infty} P_g^k d_g = v_g \ . \qquad \square$$

We define a sequence of stopping rules $\{f_0, f_1, f_2, \dots\}$ by

3.2. $f_0(i)$ is a stopping rule with $v_{f_0} \ge r$ (for example $f_0(i) = 0$ for all $i \in S$)

f_n is the improved stopping rule of f_{n-1}, $n \ge 1$ (see 3.1).

The method of approximating the optimal stopping rule and its expected return by the sequence 3.2 is called the *policy iteration method*. This method was introduced by Howard [4] for decision processes with a finite state space and discounted rewards. In theorem 1 some properties of the sequence $\{f_0, f_1, f_2, \dots\}$ are derived. In theorem 2 we study the convergence of v_{f_n} to v.

Most of Howards results carry over to our situation. Call

1) 1) $v_n := v_{f_n}$, 2) $d_n := d_{f_n}$,

3) $D_n := D_{f_n}$, 4) $\Gamma_n := \Gamma_{f_n}$.

Theorem 1. Assume C. The following assertions hold

1) $f_n(i)$ and $v_n(i)$ are nondecreasing in n

2) if $f_n(i_0) < f_{n+1}(i_0)$ then $v_n(i_0) < v_{n+1}(i_0)$.

Proof. It follows from lemma 4 that $v_{n+1} \ge v_n$ for $n \ge 0$, since $v_0 \ge r$. If $f_n(i) = 1$ then

$$r(i) < -c(i) + \sum_{j\in S} P(i,j)v_{n-1}(j) \le -c(i) + \sum_{j\in S} P(i,j)v_n(j), \quad \text{for } n \ge 1$$

hence $f_{n+1}(i) = 1$, which proves assertion 1. Suppose $f_n(i_0) = 0$ and $f_{n+1}(i_0) = 1$, then

$$v_n(i_0) = r(i_0) < -c(i_0) + \sum_{j\in S} P(i,j)v_n(j) \le$$

$$\le -c(i_0) + \sum_{j\in S} P(i_0,j)v_{n+1}(j) = v_{n+1}(i_0) \ . \qquad \square$$

Theorem 2. Assume C.

1) If, either $v_{n_0} \ge -\sum_{k=0}^{\infty} P^k c$ or $v_{n_0} \ge 0$ for some n_0, then $\lim_{n\to\infty} v_n = v$.
2) If, in addition to 1, $f_n = f_{n+1}$, for some $n \ge n_0$ then v_n is optimal.

Proof. Since $D_n \subset D$ for all n (lemma 3) and since $f_n(i)$ is nondecreasing in n (theorem 1) there exists a set $E \subset S$ such that

$$\lim_{n\to\infty} D_n = E \subset D \ .$$

And, in the same way, since $v_n(i) \le v(i)$ for all n and since $v_n(i)$ is nondecreasing in n, there exists a function z such that

$$z(i) = \lim_{n\to\infty} v_n(i) \ .$$

Fix some $i \in E$. For all n sufficiently large is $i \in D_n$ and so:

$$r(i) \le v_n(i) = -c(i) + \sum_{j\in S} P(i,j)v_n(j) \le -c(i) + \sum_{j\in S} P(i,j)v(j) = v(i) \ .$$

Since $v_n(i) \uparrow z(i)$ we have by monotone convergence

$$-c(i) + \sum_{j\in S} P(i,j)\{v_n(j) - r(j)\} \uparrow -c(i) + \sum_{j\in S} P(i,j)\{z(j) - r(j)\} \ ,$$

hence

$$z(i) = -c(i) + \sum_{j\in S} P(i,j)z(j) \le v(i) \ .$$

Fix some $i \in S\backslash E$. For all n it holds that $i \in \Gamma_n$ hence

$$v_n(i) = r(i) \ge -c(i) + \sum_{j\in S} P(i,j)v_n(j)$$

and therefore (again by monotone convergence)

$$z(i) = r(i) \ge -c(i) + \sum_{j\in S} P(i,j)z(j) \ .$$

So z satisfies the functional equation:

$$z(i) = \max\{r(i), -c(i) - \sum_{j\in S} P(i,j)z(j)\} .$$

Now, suppose $v_{n_0} \geq -\sum_{n=0}^{\infty} P^n c$. Then $z \geq -\sum_{n=0}^{\infty} P^n c$ and since z satisfies the functional equation, z is a c-excessive function dominating r. Because v is the smallest function with this property it must hold that v = z. If $v_{n_0} \geq 0$ it must hold that $z \geq 0$ and $v \geq 0$. We now prove that $v = z$ on Γ. Let $i \in \Gamma$:

$$0 \leq v(i) - z(i) \leq r(i) - r(i) = 0 .$$

Let, now $i \in D$:

$$0 \leq v(i) - z(i) \leq \sum_{j\in S} P(i,j)\{v(j) - z(j)\} .$$

Hence $0 \leq v - z \leq \tilde{P}(v - z)$.
Iterating this inequality gives

$$0 \leq v - z \leq \tilde{P}^n(v - z) \leq \tilde{P}^n v \to 0 \quad \text{for } n \to \infty$$

which proves v = z. The first assertion is proved.
Suppose $f_n = f_{n+1}$ for some $n \geq n_0$. Then $v_n = v_{n+1}$ and therefore $f_{n+2} = f_{n+1}$. By induction it follows that $z = v_n$ which proves the theorem. □

<u>Lemma 5</u>. Let c be a charge. Let f be the stopping rule defined by f(i) = 1 for all $i \in S$ and let g be the improved stopping rule, then

$$v_g \geq v_f \quad \text{and} \quad v_g \geq r .$$

If $v_g = v_f$ then f is optimal.

<u>Proof</u>. Since $v_f = -\sum_{n=0}^{\infty} P^n c$ it holds that Pv_f and $P_g^k v_f$ are finite. Following exactly the proof of lemma 3 we have $v_f \leq d_g + P_g v_f$ and from the proof of lemma 4 it follows, since $P_g^k v_f$ is finite, that

$$v_f - P_g^{n+1} v_f \leq \sum_{k=0}^{n} P_g^k d_g .$$

Note that

$$P_g^n|v_f| \leq P^n|v_f| = P_f^n|v_f| .$$

Since c is a charge:

$$w_f := \sum_{n=0}^{\infty} P^n|c| < \infty .$$

Hence $w_f = |c| + Pw_f$ and therefore $P_f^n w_f$ tends to 0 if n tends to ∞. Because $w_f \geq |v_f|$ we may conclude

$$\lim_{n\to\infty} P_g^n|v_f| = 0 .$$

Hence

$$v_f \le \sum_{k=0}^{\infty} P_g^k d_g = v_g .$$

If $g(i) = 0$ then $v_g(i) = r(i)$ and if $g(i) = 1$ then

$$r(i) < -c(i) + \sum_{j\in S} P(i,j)v_f(j) = v_f(i) \le v_g(i) .$$

Hence $v_g \ge r$.
Now, suppose $v_g = v_f$, then

$$r \le v_f = -c + P_f v_f = -c + Pv_f$$

hence v_f is c-excessive and dominates r. Because $v_f \le v$ and the fact that v is the least function with this property, we have $v = v_f$. □

Corollary 2.

1) If r is nonnegative, we have for $f_0 \equiv 0$ $v_{f_0} \ge r \ge 0$, hence the sequence v_n converges to v.
2) If c is a charge we may start with $f_{-1}(i) := 1$ for all $i \in S$ and try to improve this stopping rule by f_0. If no improvement is possible (i.e. $v_{f_0} = v_{f_{-1}}$) we have already the optimal stopping rule. Otherwise f_0 satisfies

   a) $v_0 = v_{f_0} \ge r$

   b) $v_0 \ge -\sum_{n=0}^{\infty} P^n c$

   hence v_n converges to v.

Counterexamples.

1) There exists a stopping problem satisfying assumptions A, B and C where the policy iteration method does not converge to the optimal stopping rule. Let $S = \{1,2\}$; $r(1) = r(2) = -1$, $c(1) = c(2) = 0$ and $P(1,1) = \alpha = 1 - P(1,2)$, $P(2,2) = \beta = 1 - P(2,1)$. The optimal stopping rule is $f(1) = f(2) = 1$ and $v(1) = v(2) = 0$. The cost function is a charge and $\mathbb{E}_i[|r(X_T)|] \le 1$. Note that $r(1) = \alpha r(1) + (1-\alpha)r(2)$ and $r(2) = \beta r(2) + (1-\beta)r(1)$ so that $r \ge c + Pr$ hence $f_n = f_0 \equiv 0$.
2) There exists a stopping problem satisfying assumptions A and B where the improved policy of f_0 is not at least as good as f_1. Let

   $$S = \{0,1,2,3,\dots\} \cup \{x\}, \quad 1 > \varepsilon > 0 .$$

   For $i = 0,1,2,3,\dots$:

$$P(i,i+1) = 1-\varepsilon,\ P(i,x) = \varepsilon,\ r(i) = \frac{1}{(1-\varepsilon)^i},\ c(i) = 0\ .$$

Further:

$$P(x,x) = 1,\ r(x) = 1,\ c(x) = 1\ .$$

Note that r and c are nonnegative both (condition A). We shall examine the stopping time $T_n \equiv n$:

$$v_{T_n}(i) = (1-\varepsilon)^n \frac{1}{(1-\varepsilon)^{i+n}} + \{1-(1-\varepsilon)^n\}\ .$$

Hence

$$w(i) := \sup_n v_{T_n}(i) = \frac{1}{(1-\varepsilon)^i} + 1\ .$$

This function w satisfies the functional equation

$$w(i) = \max\{r(i), -c(i) + \sum_{j\in S} P(i,j)w(j)\}$$

and $w \geq -\sum_{n=0}^{\infty} P^n c$, hence $w = v$ so that $v(i) < \infty$ from which it follows that $\mathbb{E}_i[|r(X_T)|] < \infty$ for all i and all T (condition B).
For $i = 0,1,2,3,\ldots$:

$$r(i) = \frac{1}{(1-\varepsilon)^i} < (1-\varepsilon)\frac{1}{(1-\varepsilon)^{i+1}} + \varepsilon = -c(i) + \sum_{j\in S} P(i,j)r(j)$$

and $r(x) = 1 > -c(x) + r(x)$.
Hence $f_1(i) = 1$ for $i \in \{0,1,2,3,\ldots\}$ and $f_1(x) = 0$ so that $v_1(i) = 1$ for all i, but $v_0(i) = \frac{1}{(1-\varepsilon)^i} > 1$ for $i = 1,2,3,\ldots$.

4. An application

We shall study in this section the optimal stopping of a random walk on the integers with a special cost and reward structure, to illustrate the computational aspects of the policy iteration method. For simplicity we shall not formulate the results as general as possible.

Definition of the decision process

Consider a random walk on the set of integers (Z). Let the transition matrix P be defined by

4.1. $\quad P(i,i+1) := p_i,\ P(i,i) := s_i,\ P(i,i-1) = q_i$

with $p_i, q_i > 0$, $s_i \geq 0$ and $p_i + q_i + s_i = 1$. The reward function

4.2. $\quad 0 \leq r(i) \leq M, \quad i \in Z\ .$

The cost function

4.3. $c(i) \geq \delta > 0, \quad i \in Z$.

Further we assume the existence of integers d, e, $d \leq e$, such that:

4.4. $r(i) < -c(i) + p_i r(i+1) + q_i r(i-1) + s_i r(i)$

if and only if $d \leq i \leq e$. Call $H := \{i \in Z \mid d \leq i \leq e\}$.

Assumption 4.4 says that for $i \in Z \backslash H$ immediately stopping is more profitable than making one more transition. In statistical sequential analysis there are examples of random walks where this assumption is fulfilled in a natural way (compare [5]). In lemma 6 we collect some properties of this process.

Lemma 6. For the sequence of stopping rules $f_0, f_1, f_2, \ldots$ defined in 3.2 with $f_0(i) = 0$ for all $i \in Z$ it holds that

1) there exist numbers $k_n, \ell_n \in Z$ such that

$$D_n = \{i \in Z \mid k_n \leq i \leq \ell_n\}, \quad n = 0,1,2,\ldots .$$

2) $k_n \geq k_{n+1} \geq k_n - 1$ and $\ell_n \leq \ell_{n+1} \leq \ell_n + 1$.
3) for some n f_n is optimal.

Proof. Since $0 \leq r(i) \leq M$ and $c \geq 0$ A and B are satisfied. By 1.4 we know that the entrance time in Γ is optimal, hence the assumption C is fulfilled. By theorem 1 we have $D_n \subset D_{n+1}$ for $n = 0,1,2,3,\ldots$ and by theorem 2 we have $\lim_{n\to\infty} v_n(i) = v(i)$. We shall prove 1 and 2 with induction.

D_0 is empty. It is easy to verify that $f_1(i) = 1$ if and only if $i \in H$, hence $k_1 = d$ and $\ell_1 = e$. Suppose 1 hold for n = m. For $i < k_m - 1$ and $i > \ell_m + 1$ it holds that $f_{m+1}(i) = 0$ because $v_m(i) = r(i)$ and $i \in Z\backslash H$. Therefore it can happen only in the points $i = k_m - 1$ and $i = \ell_m + 1$ that $f_{m+1}(i) > f_m(i)$. Since $D_m \subset D_{m+1}$ 1 and 2 are proved. Now the last assertion.
Note that $0 \leq r(i) \leq M$ and $c(i) \geq \delta > 0$ for all $i \in Z$. Choose $1 > \varepsilon > 0$ and a natural number k such that $(1-\varepsilon)k > \frac{M}{\delta}$. Let f be the optimal stopping rule. We shall prove $\mathbb{P}_i[T_f \leq k] \geq \varepsilon$. Suppose the contrary, i.e. let $\mathbb{P}_i[T_f \leq k] < \varepsilon$. Then

$$v_f(i) \leq M - \delta\mathbb{E}_i[T_f] \leq M - \delta(1-\varepsilon)k < 0$$

which is a contradiction.
Hence for all $i \in Z$ Γ must be reachable in at most k steps, so that $D \subset \{i \mid d-k \leq i \leq e+k\}$. Since $D_{n-1} \subset D_n \subset D$ and because D_{n-1} is a proper subset of D_n if $f_{n-1}(i) \neq f_n(i)$ for at least one i we may conclude that $f_{n-1} = f_n$ for some n.
□

Computational aspects

In our case v is the smallest solution of

$$v(i) = \max\{r(i), -c(i) + p_i v(i+1) + s_i v(i) + q_i v(i-1)\} .$$

Because we know the structure of D we may say v is the smallest function x which has the following properties.

For some $k \le d$ and some $\ell \ge e$, $i, k, \ell \in Z$:

1) $x(i) = -c(i) + p_i x(i+1) + s_i x(i) + q_i x(i-1), \quad k \le i \le \ell$

2) $x(i) = r(i), \quad i > \ell, \; i < k$

3) $r(k-1) \ge -c(k-1) + p_{k-1} x(k) + s_{k-1} r(k-1) + q_{k-1} r(k-2)$

$$r(\ell+1) \ge -c(\ell+1) + p_{\ell+1} r(\ell+2) + s_{\ell+1} r(\ell+1) + q_{\ell+1} x(\ell) .$$

This is a *two point boundary value problem* with a *free boundary*. We shall show that for fixed k and ℓ the function x is completely determined by 1 and 2.

Define, for function on Z, the difference operator Δ as usual by

4.5. $\Delta x(i) := x(i+1) - x(i)$.

Consider the difference equation, derivated from 1,

4.6. $p_i \Delta x(i) - q_i \Delta x(i-1) = c(i)$.

Call:

$$z_i := \Delta x(i), \; a_i := \frac{q_i}{p_i} \text{ and } b_i := \frac{c(i)}{p_i} .$$

Hence 4.6 becomes

$$z_i - a_i z_{i-1} = b_i .$$

With induction on m it is easy to verify that for $k \le m \le \ell$

4.7. $z_m = z_{k-1} \prod_{i=k}^{m} a_i + \sum_{i=k}^{m} \{b_i \prod_{j=i+1}^{m} a_j\}$

(an empty product has the value 1, an empty sum the value 0).

Because $x(\ell+1) = r(\ell+1)$ and $x(k-1) = r(k-1)$ it holds that

$$r(\ell+1) - r(k-1) = \sum_{m=k-1}^{\ell} z_m$$

hence

$$4.8. \qquad z_{k-1} = \frac{r(\ell + 1) - r(\ell - 1) - \sum_{m=k-1}^{\ell} \sum_{i=k}^{m} \{b_i \prod_{j=i+1}^{m} a_j\}}{\sum_{m=k-1}^{\ell} \prod_{i=k}^{m} a_i} .$$

From 4.7 and 4.8 one can compute $z_k, z_{k+1}, \ldots, z$ and even so $x(k), x(k+1), \ldots, x(\ell)$, which shows that the function x is completely determined.
The boundary conditions 3 can be formulated as follows

$$4.9. \qquad \begin{aligned} & z_{k-1} - a_{k-1}\Delta r(k-2) \le b_{k-1} \\ & \Delta r(\ell + 1) - a_{\ell+1} z_\ell \le b_{\ell+1} , \end{aligned}$$

which shows that we only have to compute the differences z_k to check 3 and not the function x itself.
It is easy to verify that the sums and products in 4.7 and 4.8 can be computed recursively. We shall formulate an algorithm to compute the optimal stopping rule and the value function v.

Algorithm

1. $k := d$, $\ell := e$,
2. compute z_{k-1} (by 4.8) and z_ℓ (by 4.7), set $i := 0$,
3. if $z_{k-1} - a_{k-1} \cdot \Sigma r(k-2) > b_{k-1}$ then $k := k - 1$ and $i := 1$,
4. if $\Delta r(\ell + 1) - a_{\ell+1} z_\ell > b_{\ell+1}$ then $\ell := \ell + 1$ and $i := 1$,
5. if $i = 0$ then goto 6, else goto 2,
6. D is the set $\{i \in Z \mid k \le i \le \ell\}$ and v can be compute by 4.7.

Acknowledgement

The author whishes to express his gratitude to Dr. A. Hordijk for valuable remarks.

Literature

[1] DYNKIN, E.B., JUSCHKEWITSCH, A.A.; Sätze und Aufgaben über Markoffsche Prozesse. Springer-Verslag (1969).

[2] HORDIJK, A., POTHARST, R., RUNNENBURG, J. Th.; Optimaal stoppen van Markov ketens. MC-syllabus19 (1973).

[3] HORDIJK, A.; Dynamic programming and Markov potential theory. MC tract (1974).

[4] HOWARD, R.A.; Dynamic programming and Markov processes. Technology Press, Cambridge Massachusetts (1960).

[5] VAN HEE, K.M., HORDIJK, A.; A sequential sampling problem solved by optimal stopping. MC-rapport SW 25/73 (1973).

[6] VAN HEE, K.M.; Note on memoryless stopping rules. COSOR-notitie R-73-12, T.H. Eindhoven (1974).

[7] ROSS, S.; Applied probability models with optimization applications. Holden-Day (1970).