Markov decision processes and quasi-martingales

Groenewegen, L.P.J.; van Hee, K.M.

Published: 01/01/1976

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 21. Dec. 2018
Markov decision processes and Quasi-Martingales

by

L.P.J. Groenewegen and K.M. van Hee

Eindhoven, February 1976

The Netherlands
Markov decision processes and Quasi-Martingales

by

L.P.J. Groenewegen and K.M. van Hee

0. Abstract

It is showed in this paper that quasi-(super)martingales play an important role in the theory of Markov decision processes. For excessive functions (with respect to a charge) it is proved that the value of the state at time t converges almost surely under each Markov strategy, which implies that the value function in the state at time t converges to zero (a.s), if an optimal strategy is used. At last a characterization of the conserving and equalizing properties is formulated using martingale theory.

1. Introduction

In this section the framework of convergent dynamic programming, see Hordijk (1974a, 1974b), is sketched. A Markov decision process will be a triple \((S,P,r)\) where \(S\) is a countable set, called the state space, \(r\) is a real measurable function on \(S \times P\), called the reward function and \(P\) is a Borel subset of \(E\) where \(E\) is the set of all Markov transition functions on \(S\), i.e.
\[
P \in E, \text{ implies } P : S \times S \to [0,1], \sum_{j \in S} P(i,j) \leq 1 \text{ for all } i \in S.
\]
It is assumed that \(E\) is endowed with a metric such that \(E\) is a Polish space.

We will use the following notational convention for functions \(g\) on \(S \times P\):
\[
g_P(i) := g(i,P)
\]
and we assume that all functions on \(S \times P\) are measurable on \(P\).

It is assumed that \(P\) and \(r\) have the following (product) properties:
let \(P_1,P_2,P_3,\ldots \in P\) and \(A_1,A_2,A_3,\ldots \subset S\) then there is a \(P \in P\) such that for all \(i \in S\):
\[
P(i,:) = P_k(i,:)
\]
if \(i \in A_k\) and \(r_P(i) = r_{P_1}(i)\) if \(P_1(i,:) = P_2(i,:)\).

A Markov strategy is a sequence \((P_0,P_1,P_2,\ldots)\) with \(P_i \in P_i\); the set
of all Markov strategies is denoted by $M$. The state space $S$ is extended to $S^*$ by adding a state $p$ such that $S^* := S \cup \{p\}$, and all $P \in P$ are extended to $S^*$ by: $P(i,p) := 1 - \sum_{j \in S} P(i,j)$ and $P(p,p) := 1$.

All functions on $S$ are extended to $S^*$ by defining them 0 in $p$.

Let $F_n$ be the usual $\sigma$-field on $(S^*)^\infty$ generated by the first $n + 1$ coordinates of the paths, $n = 0, 1, 2, \ldots$, and $F$ is the $\sigma$-field generated by $\bigvee_{n=0}^{\infty} F_n$.

Now $\{X_n, n = 0, 1, 2, \ldots\}$ is a stochastic process on $((S^*)^\infty, F, P_{i,R})$ for each $(i,R)$, and for all $\omega \in (S^*)^\infty X_n(\omega)$ selects the $(n + 1)$-th coordinate of $\omega$.

The expectation with respect to $P_{i,R}$ is denoted by $E_{i,R}$. Any $(\Omega,F)$-measurable function $f$ is said to be integrable w.r.t. $F$ if at least one of the terms $E_{i,R}f^+$ and $E_{i,R}f^-$ is finite, and summable if both are finite.

If $f(X_n)$ is integrable w.r.t. $P_{i,R}$ the expectation may be evaluated as follows: $E_{i,R}[f(X_n)] = P_0 \ldots P_{n-1} f(i)$ for $R = P_0, P_1, \ldots$ (an empty product of Markov transition functions is defined as the identity operator).

Definition 1.1.

i) A function $g : S \times P \to \mathbb{R}$ is called a charge iff

$$E_{i,R} \left[ \sum_{n=0}^{\infty} |g_{P_n}(X_n)| \right] < \infty$$

for all $i \in S$, $R \in M$.

Let $g$ be such a charge.

ii) A function $f : S \to \mathbb{R}$ is called superharmonic w.r.t. (a charge) $g$ iff $f(X_n)$ is integrable w.r.t. $P_{i,R}$ for all $i,R,n$ and

$$f \geq g_P + Pf$$

for all $P \in P$.

iii) A function $f : S \to \mathbb{R}$ is called excessive w.r.t. (a charge) $g$ iff $f$ is superharmonic w.r.t. $g$ and

$$f(i) \geq \sum_{n=0}^{\infty} E_{i,R}[g_{P_n}(X_n)]$$

for all $i,R$. 

Assumption 1.2.
i) The reward function \( r \) is a charge and

\[ \sup_{R \in M} \mathbb{E}_{i,R} \left[ \sum_{n=0}^{\infty} r^{+}_p (X_n) \right] < \infty \]

(recall: \( x^+ := \max(0,x) \), \( x^- := (-x)^+ \)).

Definition 1.3.
i) The \textit{value function} \( v \) of \((S,P,r)\) is a real function on \( S \):

\[ v(i) := \sup_{R \in M} \mathbb{E}_{i,R} \left[ \sum_{n=0}^{\infty} r_p (X_n) \right] \quad \text{for all } i \in S \]

ii) A strategy \( R \in M \) is called \textit{optimal} if this supremum is attained for \( R \) in all \( i \in S \).

In this paper it will be shown, using some theory on supermartingales, that each superharmonic function \( f \) with w.r.t. charge \( g \) has the property that \( f(X_n) \) converges \( \mathbb{P}_{i,R} \) almost surely for all \( i \) and \( R \). Especially the value function \( v(X_n) \) converges almost surely to zero under each optimal strategy.

As last result we give a slight extension of the following theorem of Hordijk (1974a), which is based on a result of Dynkin and Juschkewitsch (1969): if \( \tau_1 \) and \( \tau_2 \) are stopping times for the sequence of \( \sigma \)-fields \( \{F_n, n = 0,1,\ldots\} \) then \( \tau_1 \leq \tau_2 \mathbb{P}_{i,R} \) a.s. implies:

\[ \mathbb{E}_{i,R}[v(X_{\tau_1^-}) + \sum_{k=0}^{\tau_1^-} r_p (X_k)] = \mathbb{E}_{i,R}[v(X_{\tau_2^-}) + \sum_{k=0}^{\tau_2^-} r_p (X_k)]. \]

In section 2 some theory on quasi-martingales is developed and in section 3 this theory is applied to Markov decision processes. Most of the lemmas used in this paper are well-known but the authors do not know of any place in the literature where the facts were combined to get the results mentioned above.
2. Quasi-martingales

Quasi-martingales have been introduced by Fisk (1965) as continuous time stochastic processes having a decomposition into the sum of a martingale and a process having almost all sample functions of bounded variation. In this paper we give essentially the same definition for the discrete time case.

Let $\mathbb{N}$ be the set $\{0,1,2,\ldots\}$ and let $(\mathcal{G},\mathcal{A},\mathbb{P})$ be a probability space and $(A_t, t \in \mathbb{N})$ an increasing sequence of $\sigma$-fields contained in $\mathcal{A}$. All stochastic processes in this section are defined on $(\mathcal{G},\mathcal{A},\mathbb{P})$ and have values in the set of real numbers with the Borel-$\sigma$-field on it. Moreover they are adapted to $(A_t, t \in \mathbb{N})$, i.e. the $\sigma$-field generated by the first $n+1$ coordinates of the paths is a subset of $A_n$, $n \in \mathbb{N}$ (The conditional expectation w.r.t. $A_t$ is denoted by $\mathbb{E}^A_t$).

**Definition 2.1.**

Let $\{B_t, t \in \mathbb{N}\}$ be a stochastic process such that $\sum_{t \in \mathbb{N}} |B_t| < \infty \mathbb{P}$-a.s. (super)martingale $\{S_t, t \in \mathbb{N}\}$ such that

$$V_t = S_t + \sum_{k=0}^{t-1} B_t \quad \mathbb{P}$-a.s.$$

In Fisk's paper the process $\{B_t, t \in \mathbb{N}\}$ of definition 2.1 is called a process of bounded variation.

$\{(V_t, t \in \mathbb{N})$ is said to be a QSPM w.r.t. $\{B_t, t \in \mathbb{N}\}$).

**Lemma 2.2.**

Let $\{B_t, t \in \mathbb{N}\}$ and $\{V_t, t \in \mathbb{N}\}$ be stochastic processes with $\sum_{t \in \mathbb{N}} |B_t| < \infty \mathbb{P}$-a.s. Then:

$\{V_t, t \in \mathbb{N}\}$ is a QSPM w.r.t. $\{B_t, t \in \mathbb{N}\}$ iff $\sum_{t \in \mathbb{N}} |B_t| < \infty \mathbb{P}$-a.s.

**Proof:**

Define $S_t := V_t - \sum_{k=0}^{t-1} B_t$, $t \in \mathbb{N}$. 

$$E^{A_t} V_{t+1} \leq B_t + V_t \quad \mathbb{P}$-a.s.$$

Define $\sum_{k=0}^{t-1} B_t$, $t \in \mathbb{N}$. 

$$E^{A_t} V_{t+1} \leq B_t + V_t \quad \mathbb{P}$-a.s.$$$
i) Suppose
\[
A_t V_{t+1} \leq V_t + B_t
\]
Then
\[
E S_{t+1} = E V_{t+1} - \sum_{k=0}^{t} B_k \leq V_t - \sum_{k=0}^{t-1} B_k = S_t
\]
Hence \( \{ S_t, t \in \mathbb{N} \} \) is a supermartingale.

ii) Conversely, suppose \( \{ V_t, t \in \mathbb{N} \} \) is a QSPM w.r.t. \( \{ B_t, t \in \mathbb{N} \} \).
Then
\[
E V_{t+1} = E S_{t+1} + \sum_{k=0}^{t} B_k \leq V_t + B_t \quad P\text{-a.s.}
\]
For a quasi-martingale the same characterization holds with equality. In lemma 2.3 it is shown that quasi-(super)martingales converge \( P\text{-a.s.} \) under a condition analogous to that for supermartingales.

Lemma 2.3.

Let \( \{ B_t, t \in \mathbb{N} \} \) and \( \{ V_t, t \in \mathbb{N} \} \) be as in lemma 2.2 with \( E V_{t+1} \leq B_t + V_t, P\text{-a.s.} \)
Assume furthermore

i) \( \limsup_{t \in \mathbb{N}} E V_t^- < \infty \)

ii) \( E \sum_{k=0}^{\infty} B_k^+ < \infty \)

then \( V_t \) converges \( P\text{-a.s.} \).

Proof.

Let
\[
S_t := V_t - \sum_{k=0}^{t-1} B_k, \quad t \in \mathbb{N}.
\]
\[
\limsup_{t \in \mathbb{N}} E S_t^- \leq \limsup_{t \in \mathbb{N}} E V_t^- + \limsup_{t \in \mathbb{N}} E \sum_{k=0}^{t-1} B_k^+ < \infty.
\]
Since \( \{S_t, t \in \mathbb{N}\} \) is a supermartingale it follows from the convergence theorem on supermartingales (see e.g. Neveu (1972), IV -1-2) that \( S_t \) converges \( \mathbb{P} \)-a.s. Also \( \sum_{k=0}^{t-1} B_k \) converges a.s., hence \( V_t \) does.

Remark.

From the proof of the cited theorem of Neveu it can be seen that Neveu's condition: \( \sup_{t \in \mathbb{N}} \mathbb{E} V_t^- < \infty \) can be replaced by \( \limsup_{t \in \mathbb{N}} \mathbb{E} V_t^- < \infty \).

The next lemma is not really used in the rest of the paper, but it shows that the requirements of lemma 2.3 almost imply that \( \mathbb{E} \left| \sum_{n=0}^\infty B_n \right| < \infty \).

Lemma 2.4.

Assume in addition to the assumptions of lemma 2.3 that \( V_0 \) is summable, then it holds that

\[
\mathbb{E} \sum_{n=0}^\infty |B_n| < \infty.
\]

Proof.

Let \( S_t \) be defined as in the proof of lemma 2.3.

Note that

\[
S_0 \geq \mathbb{E} S_t = \mathbb{E} V_t - \mathbb{E} \sum_{k=0}^{t-1} B_k.
\]

Let \( M := \limsup_{t} \mathbb{E} V_t^- \) and note that \( \limsup_{t} \mathbb{E} V_t + M \geq 0 \).

Then

\[
\mathbb{E} S_0 + \mathbb{E} \sum_{k=0}^{t-1} B_k^+ + M \geq \mathbb{E} (V_t + M) + \mathbb{E} \sum_{k=0}^{t-1} B_k^-
\]

hence, by taking the limsup of both sides we have:

\[
\mathbb{E} S_0 + \mathbb{E} \sum_{k=0}^{\infty} B_k^+ + M \geq \mathbb{E} \sum_{k=0}^{\infty} B_k^- \text{ for all } t \in \mathbb{N}
\]
Since \( \sum_{k=0}^{\infty} B_k < \infty \) it holds that \( \sum_{k=0}^{\infty} B_k < \infty \) which proves the lemma.

This section ends with a property on regular supermartingales. The definition of regularity given below is equivalent to the usual one (see e.g. Neveu (1972) IV-5-24).

Let \( \{S_t, t \in \mathbb{N}\} \) be a supermartingale and let \( \tau_1 \) and \( \tau_2 \) be two stopping times w.r.t. \( \{A_t, t \in \mathbb{N}\} \). \( A_\infty \) is the \( \sigma \)-field generated by \( \cup_{n \in \mathbb{N}} A_n \) and

\[
A_\tau := \{ B \in A_\infty | B \cap \{\tau_1 = n\} \in \mathbb{N}, n \in \mathbb{N}\}
\]

**Definition 2.5.**

The supermartingale \( \{S_t, t \in \mathbb{N}\} \) is called regular iff the sequence \( \{S_t, t \in \mathbb{N}\} \) converges in \( L^1 \)-sense.

**Property 2.6.** Let \( \{S_t, t \in \mathbb{N}\} \) be a regular supermartingale and let \( \tau_1, \tau_2 \) be stopping times. It holds that \( S_{\tau_1} \) and \( S_{\tau_2} \) are integrable and

\[
A_{\tau_1} \geq E_{\tau_1} S_{\tau_2} \text{ P-a.s. on } \{\tau_1 \leq \tau_2\}
\]

(for a proof see Neveu (1972) IV-5-25).


In this section we return to the model described in section 1. We first give a quick survey of the properties of this model which are relevant to our exposition here.

**Properties.**

3.1. \[
\sup_{i \in \mathbb{R}} \sum_{n=0}^{\infty} r_{i, n} (X_n) = \sup_{i \in \mathbb{R}} \sum_{n=0}^{\infty} r_{i, n} (X_n)
\]
where \( \Pi \) is the set of all strategies (see Van Hee (1975) for the definition of \( \Pi, \mathbb{E}_i \), and the proof).

For this property assumption 1.2 (ii) is required.

3.2. The value function \( v \) satisfies Bellman's optimality equation:

\[
v(i) = \sup_{P \in \mathcal{P}} \{ r_P(i) + P v(i) \}.
\]

This statement is a standard consequence of 3.1: the proof is similar to those in Ross (1970), th. 6.1 or Hordijk (1974\textsuperscript{a}) th. 3.1 and 3.5.

3.3. Let \( g \) be a charge and \( f \) a superharmonic function w.r.t. \( g \) then \( \lim_{n \to \infty} \mathbb{E}_i \mathbb{R} f(X_n) \) exists for all \( R \in M, i \in S \) and the following assertions are equivalent:

i) \( f \) is excessive w.r.t. \( g \)

ii) \( \lim_{n \to \infty} \mathbb{E}_i f(X_n) = 0 \) for all \( R \in M, i \in S \)

iii) \( \lim_{n \to \infty} \mathbb{E}_i f(X_n) = 0 \) for all \( R \in M, i \in S \).

For a proof see Hordijk (1974\textsuperscript{a}) th. 2.17

Remark 3.4.

It is obvious from 3.2 that the value function \( v \) is superharmonic w.r.t \( r \) and by its definition it is clear that

\[
v(i) \geq \mathbb{E}_i \mathbb{R} [ \sum_{n=0}^{\infty} r_P(X_n) ] \text{ for all } i \in S, R \in M,
\]

hence \( v \) is excessive w.r.t. \( r \).

Remark 3.5.

If \( f \) is excessive w.r.t. a charge \( g \) it holds that

\[
\lim_{n \to \infty} \mathbb{E}_i \mathbb{R} |f(X_n)| = \lim_{n \to \infty} \mathbb{E}_i [f(X_n)] \text{ for all } i \in S, R \in M
\]

by 3.3 (ii).
Lemma 3.6.

Let $f$ be a superharmonic function w.r.t. $g$. Then for all $R \in M, t \in \mathbb{N}$ and $i \in S$ it holds that

$$f(X_t) \geq g_P(X_t) + \mathbb{E}_{i,R}^t f(X_{t+1})$$

Proof.

It is clear that $(P_t f)(X_t)$ is $F_t$-measurable and since

$$\mathbb{E}_{i,R}^t [P_t f(X_t)] = \sum_{k=0}^t P_t f(i) = \mathbb{E}_{i,R}^t [f(X_{t+1})]$$

it holds that

$$\mathbb{E}_{i,R}^t [f(X_{t+1})] = P_t f(X_t) \mathbb{P}_{i,R}^t - a.s.$$  

Hence, by the superharmonicity of $f$ the statement follows. 

The main results of this paper are easy to prove now.

Theorem 3.7.

Let $f$ be an excessive function w.r.t a charge $g$.

For any $i \in S, R \in M$ \{f(X_t), t \in \mathbb{N}\} is a quasi-supermartingale w.r.t. \{g_k(X_k), k \in \mathbb{N}\} and $f(X_t)$ converges $\mathbb{P}_{i,R}^t - a.s. (for t \to \infty)$.

Proof.

Fix $i \in S, R \in M$. By lemma 3.6. we have $f(X_t) \geq g_P(X_t) + \mathbb{E}_{i,R}^t f(X_{t+1})$.

Since $g$ is a charge we have

$$\sum_{k=0}^\infty |g_k(X_k)| < \infty \mathbb{P}_{i,R}^t - a.s.$$  

So lemma 2.2 shows that \{f(X_t), t \in \mathbb{N}\} is QSPM w.r.t. \{g_k(X_k), k \in \mathbb{N}\}. From $g$ being a charge and property 3.3 ii) it follows that all conditions of lemma 2.3 are fulfilled, which proves the theorem.
Theorem 3.8.

Let $f$ be an excessive function w.r.t. a charge $g$.

The supermartingale

$$\{f(X_t) + \sum_{k=0}^{t-1} g_p(X_k), t \in \mathbb{N}\}$$

is regular.

Proof.

Fix $i \in S$, $R \in M$ and let $S_t := f(X_t) + \sum_{k=0}^{t-1} g_p(X_k)$.

By theorem 3.7 we have: $\{S_t, t \in \mathbb{N}\}$ is a supermartingale, so we only have to check that $S_t^-$ converges in $L^1$-sense.

$$S_t^- \leq f^-(X_t) + \sum_{k=0}^{t-1} g_p^-(X_k)$$

hence

$$S_t^- - f^-(X_t) \leq \sum_{k=0}^{t-1} g_p^-(X_k) \leq \sum_{k=0}^{\infty} |g_p(X_k)|$$

On the other hand, since $(a + b)^- \geq a^- - b^+$

$$S_t^- = [f(X_t) + \sum_{k=0}^{t-1} g_p(X_k)]^- \geq f^-(X_t) - [\sum_{k=0}^{t-1} g_p(X_k)]^+$$

$$\geq f^-(X_t) - \sum_{k=0}^{t-1} g_p^+(X_k)$$

hence

$$S_t^- - f^-(X_t) \geq - \sum_{k=0}^{\infty} |g_p(X_k)|$$

By the dominated convergence theorem we have the $L^1$-convergence of $S_t^- - f^-(X_t)$. By property 3.3 ii) we have the $L^1$-convergence of $f^-(X_t)$.
to zero. This implies the $L^1$-convergence of $S_t^-$. 

**Corollary 3.9.**

Let $f$ be an excessive function w.r.t. a charge $g$, and let $\tau_1$ and $\tau_2$ be stopping times w.r.t. $\{F_t, t \in \mathbb{N}\}$

then

$$
\sum_{k=0}^{\tau_1 - 1} g_{P_k}(X_k) + f(X_{\tau_1}) + \mathbb{E}_{i,R} \left[ \sum_{k=0}^{\tau_2 - 1} g_{P_k}(X_k) + f(X_{\tau_2}) \right] \geq \mathbb{E}_{i,R} \left[ \sum_{k=0}^{\tau_1 - 1} g_{P_k}(X_k) + f(X_{\tau_1}) \right]
$$

$P_{i,R}$-a.s. on $\{\tau_1 \leq \tau_2\}$ for all $i \in S$, $R \in M$.

Note that 3.9 is a direct consequence of property 2.6 and theorem 3.8.

If $P_{i,R}[\tau_1 \leq \tau_2] = 1$ integration w.r.t. $P_{i,R}$ gives the theorem of Hordijk mentioned in the introduction.

4. Some remarks.

1). A strategy $R = (P_0, P_1, \ldots) \in M$ is called *conserving* if $v = r_{P_t} + P_t v$

for all $t \in \mathbb{N}$ ($v$ is the value function) and $R$ is called *equalizing* if

$$
\lim_{t \to \infty} \mathbb{E}_{i,R}[v(X_t)] = 0 \quad \text{for all } i \in S.
$$

It is well-known (see the proof of th. 4.6 in Hordijk (1974a)) that $R \in M$ is optimal iff $R$ is equalizing and conserving.

For each equalizing strategy $R$ we may conclude that $v(X_n) \to 0$ $P_{i,R}$-a.s. (for all $i \in S$) since by th. 3.7 $v(X_n)$ converges $P_{i,R}$-a.s. and by 3.5 we know that this limit must be zero.

In Groenewegen (1975) this result has been proved for optimal strategies.

2). For a conserving strategy $R = (P_0, P_1, \ldots)$ it holds that

$$
v(X_t) = r_{P_t}(X_t) + P_t v(X_t)
$$

hence $\{v(X_t), t \in \mathbb{N}\}$ is a quasi-martingale in this case.
3). Let \( g : S \times P \to \mathbb{R} \) and \( f_t : S \to \mathbb{R} \) \( t \in \mathbb{N} \); suppose we do not know whether \( g \) is a charge or not. Assume

i) \( f_t(X_t) \geq g_p(X_t) + \mathbb{E}_{t+1} (X_{t+1}) \mathbb{P}_{t+1} \) as for all \( i \in S, t \in \mathbb{N} \), \( R \in M \) and with \( \mathbb{E}_{t+1} f_t(X_{t+1}) \) well-defined.

ii) \( \limsup_{t \to \infty} \mathbb{E}_{i,R} f_t(X_t) < \infty \) for all \( i \in S, R \in M \) and \( t \in \mathbb{N} \)

iii) \( \sum_{k=0}^{\infty} |g_p(X_k)| < \infty \) \( \mathbb{P}_{i,R} \)-a.s. for all \( R \in M \)

iv) \( \mathbb{E}_{i,R} \sum_{k=0}^{\infty} |g_p(X_k)| < \infty \) for all \( i \in S, R \in M \).

Under these conditions, similar to those in lemma 2.3, we have that \( f_t(X_t) \) converges \( \mathbb{P}_{i,R} \)-a.s. for each \( i \in S, R \in M \). If in addition \( f_0 \) is finite, this implies also by lemma 2.4 that \( g \) is a charge. So if \( f_t = f \) for all \( t \in \mathbb{N} \), \( f \) is superharmonic w.r.t. \( g \).

4). Let \( N \in \mathbb{N} \) and \( Q_0, Q_{N+1}, \ldots \in P \) and let \( g \) be a charge.

Define \( R := \{ R = P_0 P_1 \ldots \in M | P_k = Q_k \text{ for } k \geq N \} \) and

\[
v_k := \begin{cases} \sum_{n=k}^{\infty} Q_k Q_{k+1} \ldots Q_{n-1} g_{Q_n} & k \geq N \\ \sup_{P \in P} \{ g_P + P v_{k+1} \} & 0 \leq k \leq N - 1 \end{cases}
\]

It is easy to check that the assumptions i), iii) and iv) in the above remark are satisfied for \( R \in R \). And assumption ii) is satisfied by the observation that \( v(n, X_n) := v_n(X_n) \) is excessive w.r.t. \( g \) for the space-time process, where the only allowed strategies are elements of \( R \). Hence \( v_n(X_n) \) converges \( \mathbb{P}_{i,R} \)-a.s.
5). Let $f$ be an excessive function w.r.t. a charge $g$. From the 3.8 we know that $f(X_t)$ converges $\mathbb{P}_{i,R}$-a.s. for $t \to \infty$ and from property 3.3 (ii) we know that $f(X_t)$ converges in $L^1$-sense for $t \to \infty$. The following counterexample shows that in general $f^+(X_t)$ does not converge in $L^1$-sense for $t \to \infty$.

Example:

$S := \{0, 1, 2, \ldots\}$. $P$ and $Q$ are Markov transition functions with $P(0, 0) = 1$, $P(i, i + 1) = i / (i + 1)$ for $i \geq 1$, $P(i, 0) = 1 / (i + 1)$ for $i \geq 1$, $Q(i, 0) = 1$. $P$ is the collection of Markov transition functions which can be generated from $P$ and $Q$ by using the product property. Furthermore $r_P \equiv 0$ and $r_Q(i) = i$.

It can be verified easily that the conditions 1.2 i) and ii) are fulfilled, that $v(i) = i$, that $\lim v(X_t) = 0$ $\mathbb{P}_{i,R}$-a.s. for all $R$ and that

$$\lim_{t \to \infty} \mathbb{E}^t_{i,R} v(X_t) = 1$$

for $R = PPP\ldots$

But is well-known that the $L^1$-limit and the a.s.-limit should be equal, if both exist. So $v(X_t)$ does not converge in $L^1$-sense for $t \to \infty$.

6). In Mandl (1974) a martingale is considered in connection with the average cost criterion for the optimal control of a Markov chain. Using his construction for the total return criterion we get the following. Define a real function on $S \times \mathbb{P}$:

$$\phi(i, P) := r_P(i) + P v(i) - v(i)$$

and a random variable

$$Y_n := r_P(i_n) + v(X_{n+1}) - v(X_n) - \phi(X_n, P_n)$$

It is easy to see that

$$\mathbb{E}^n_{Y_n} = 0,$$

so $M_n := \sum_{k=0}^{n-1} Y_k$

is a martingale. Now this martingale becomes

$$M_n = \sum_{m=0}^{n-1} r_{P_m}(X_m) + v(X_n) - v(X_0) - \sum_{m=0}^{n-1} \phi(X_m, P_m).$$

Note that

1. $v(X_0) = v(i) \mathbb{P}_{i,R}$-a.s. for all $R \in M$

2. by (3.2) $\phi(X_m, P_m) \leq 0$ $\mathbb{P}_{i,R}$-a.s.
Hence
\[ n-1 \sum_{m=0}^{n-1} r_p(X_m) + v(X_n) \]
is the supermartingale treated in section 3. In view of the conserving and equalizing strategies, mentioned in remark 4.1 it is worthwhile to note that
\[ v(i) - E_{i,R}[\sum_{n=0}^{\infty} r_p(X_n)] = \lim_{n \to \infty} E_{i,R}[v(X_n)] + E_{i,R}[\sum_{n=0}^{\infty} \delta(X_n,P_n)] \]
for all \( R \in M \).
From this equality it is easy to see that a \( R \in M \) is optimal if and only if it is equalizing and conserving.

**Literature**


