Math Meth Oper Res (1999) 50 : 421–448

Mathematical Methods
of Operations Research

Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards

Arie Hordijk[1], **Alexander A. Yushkevich**[2]

[1] Department of Mathematics and Computer Science, Leiden University, 2300 RA Leiden, The Netherlands (e-mail: hordijk@math.leidenuniv.nl)
[2] Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA (e-mail: aayushke@email.uncc.edu)

Abstract. This paper is the second part of our study of Blackwell optimal policies in Markov decision chains with a Borel state space and unbounded rewards. We prove that a stationary policy is Blackwell optimal in the class of all history-dependent policies if it is Blackwell optimal in the class of stationary policies.

We also develop recurrence and drift conditions which ensure ergodicity and integrability assumptions made in the previous paper, and which are more suitable for applications. As an example we study a cash-balance model.

Key words: Markov decision chains, Blackwell optimality, drift and recurrence conditions

1. Introduction

In the preceding paper Hordijk and Yushkevich (1999), further on referred to as HY, we proved the existence of Blackwell optimal policies in the class of stationary policies in Markov decision chains (MDC) with a Borel state space, unbounded rewards, and transition densities (with respect to a reference measure) satisfying compactness-continuity, uniform ergodicity and uniform integrability conditions (the full list of assumptions is presented in Section 2). Our first goal in this paper is to prove under the same assumptions that a policy Blackwell optimal in the class of stationary policies is at the same time Blackwell optimal in the class of all history-dependent policies (Theorem 4.1), and thus to establish the existence of Blackwell optimal policies in that broader sense (Theorem 2.2). The proof follows the main patterns of a similar proof for countable models developed in Dekker and Hordijk (1988), with the addi-

Manuscript received: October 1998

tional use of the weak-strong topology in the class of stationary policies, as it is done in the case of Borel models with bounded rewards in Yushkevich (1997). However, it is more technical and requires additional results on this topology.

The uniform geometric ergodicity condition for Markov chains generated by stationary policies, and the uniform integrability condition for nonstationary Markov policies assumed in HY, are nonpractical for a straightforward verification in specific MDCs with a noncompact Borel state space and an unbounded reward function. Our second goal is to substitute these two conditions by simpler recurrence and drift (Lyapunov-type) assumptions more suitable for applications (Theorems 5.1 and 5.2). This part of the paper is based on the ideas developed in Hordijk and Spieksma (1992) and Hordijk, Spieksma and Tweedie (1995).

Our third goal is to apply the obtained results to a cash-balance control model (Theorem 6.1).

In Section 2 we recall the model, definitions and assumptions from HY. In Section 3 we prove additional properties of the weak-strong topology in the space of stationary policies needed for our goals. In Section 4 we extend the Blackwell optimality from stationary to all policies. In Section 5 we replace the uniform integrability and ergodicity assumptions by simpler recurrence and drift conditions. The cash-balance model is treated in Section 6.

2. Model and assumptions

In this section we give a brief review of the model, notations and assumptions from HY, refering to that paper for more details. For convenience, we preserve the numeration of assumptions given in HY.

A Markov decision chain (MDC) is defined by a state space X, an action space A, action sets $A_x = A(x)$, a transition probability function $P(x, a, B)$ and a reward function $r(x, a)$. Let

$$K = \{(x, a) : \ a \in A_x, \ x \in X\}.$$

Everywhere below measurability means Borel measurability, and $\mathscr{B}_E$ denotes the Borel σ-algebra in a Borel space E.

Assumption 2.1. *The state space X is a Borel space, the action space A is a topological Borel space, the action sets A_x are nonempty compact x-sections of a Borel measurable set K in $X \times A$. The reward function $r(x, a)$ and the transition probability $P(x, a, B)$, $B \in \mathscr{B}_X$ are measurable functions of (x, a) on K.*

By $\Phi \subseteq \Sigma \subseteq M \subseteq \Pi$ we denote, respectively, the sets of deterministic stationary, all stationary, Markov, arbitrary history-dependent policies. A policy from Φ is determined by (and identified with) a selector, i.e. a measurable mapping $\varphi : X \to A$ with its graph in K. A policy $\sigma \in \Sigma$ is a (measurable) stochastic kernel $\sigma(x, da)$ from X to A with the property $\sigma(x, A_x) = 1$. A policy $\pi \in M$ is a sequence $\{\sigma_1, \sigma_2, \ldots\}, \sigma_t \in \Sigma$.

To every initial state $x \in X$ and policy $\pi \in \Pi$ there corresponds a probability distribution $\mathbf{P}_x^\pi$ in the space of all infinite-horizon trajectories $x_0 a_1 x_2 a_2 \ldots$ (we follow here the enumeration of states and actions used in

Dynkin and Yushkevich (1979)). Let $\mathbf{E}_x^\pi$ be the corresponding expectation. For any discount factor $\beta \in (0, 1)$, the expected discounted reward $v_\beta(x, \pi)$, $x \in X$, $\pi \in \Pi$ is defined by

$$v_\beta(x, \pi) := \mathbf{E}_x^\pi \sum_{t=0}^{\infty} \beta^t r(x_t, a_{t+1}) = \sum_{t=0}^{\infty} \beta^t \mathbf{E}_x^\pi r(x_t, a_{t+1}). \tag{2.1}$$

The convergence of (2.1) is guaranteed by the bounding Assumption 2.2 below. To formulate this assumption, we introduce the following notations.

Given a possibly unbounded strictly positive function μ on X, we denote by V_μ the Banach space of measurable real-valued functions f on X with the finite μ-norm

$$\|f\|_\mu := \sup_{x \in X} \frac{|f(x)|}{\mu(x)};$$

for an operator T: $V_\mu \to V_\mu$ we set

$$\|T\|_\mu := \sup_{f : \|f\|_\mu \leq 1} \|Tf\|_\mu.$$

Also, for any real-valued function f on K, let

$$\hat{f}(x) := \sup_{a \in A_x} |f(x, a)|, \quad x \in X. \tag{2.2}$$

Assumption 2.2. *A measurable bounding function $\mu \geq 1$ on X is given, and the reward function r and the transition operator P satisfy the conditions:*

(a) $\|\hat{r}\|_\mu \leq 1$;
(b) *for some constant $C > 0$*

$$P\mu(x, a) \leq C\mu(x), \quad (x, a) \in K,$$

where the operator P transforms functions f on X into functions Pf on K by the formula

$$Pf(x, a) := \int_X f(y) P(x, a, dy), \quad (x, a) \in K$$

(the bound 1 in (a) is taken to simplify formulas; all results are trivially extended to any finite constant instead of 1).

Next is the continuity assumption.

Assumption 2.3. (a) *The reward function $r(x, a)$, $(x, a) \in K$ is continuous in a.*
(b) *$Pf(x, a)$ is continuous in a for every $f \in V_\mu$.*

In the uniform ergodicity assumption we consider all Markov chains on the space X generated by stationary policies $\sigma \in \Sigma$. For such policies, formula

(2.1) reduces to

$$v_\beta(x,\sigma) = \sum_{t=0}^{\infty} \beta^t P_\sigma^t r_\sigma(x), \quad x \in X, \quad \sigma \in \Sigma, \quad \beta \in (0,1) \tag{2.3}$$

where

$$\begin{aligned} P_\sigma f(x) &:= \int_A Pf(x,a)\sigma(x,da), \\ r_\sigma(x) &:= \int_A r(x,a)\sigma(x,da), \quad x \in X, \quad \sigma \in \Sigma. \end{aligned} \tag{2.4}$$

Another form of the operator P_σ defined in (2.4) is given in terms of the transition function $P_\sigma(x,B)$, $B \in \mathscr{B}_X$ of the considered Markov chain:

$$\begin{aligned} P_\sigma f(x) &= \int_X f(y)P_\sigma(x,dy), \\ P_\sigma(x,B) &:= \int_A P(x,a,B)\sigma(x,da), \quad x \in X. \end{aligned} \tag{2.5}$$

Assumption 2.4. *For every $\sigma \in \Sigma$, the t-th convolution $P_\sigma^t(x,B)$ of the transition function $P_\sigma(x,B)$ defined in (2.5) converges to a stochastic transition function $\bar{P}_\sigma(x,B)$ from X to X in the following sense: there exist positive numbers C and $\gamma < 1$ such that*

$$\|P_\sigma^t - \bar{P}_\sigma\|_\mu \le C\gamma^t, \quad \sigma \in \Sigma, \quad t = 0,1,2,\ldots$$

where $\bar{P}_\sigma$ is the operator in V_μ defined by a formula similar to (2.5).

The next assumption is about the existence of a transition density and its properties. Mention that it covers Assumptions 2.2(b) and 2.3(b) (see HY).

Assumption 2.5. *A σ-finite reference measure m on the space X is given, and*

(a) *the transition probabilities are defined by means of a given transition density function $p(x,a,y)$ ($(x,a) \in K$, $y \in X$) so that*

$$P(x,a,B) = \int_B p(x,a,y)m(dy), \quad (x,a) \in K, \quad B \in \mathscr{B}_X,$$

where p is measurable, nonnegative and its integral over X is identically 1;

(b) *$p(x,a,y)$ is continuous in a;*

(c) *the transition density p is related to the bounding function μ by the condition*

$$\int_X \hat{p}(x,y)\mu(y)\,dy \le C\mu(x), \quad x \in X \tag{2.6}$$

where, similar to (2.2)

$$\hat{p}(x,y) = \max_{a \in A_x} p(x,a,y), \quad (x,y) \in X \times X. \tag{2.7}$$

Where it is not confusing, we write dx in place of $m(dx)$.

In the last, uniform integrability condition we consider all Markov policies $\pi = \{\sigma_1, \sigma_2, \ldots\} \in M$. For such a policy, the definition (2.1) of the discounted reward takes on the following form similar to (2.3):

$$v_\beta(x,\pi) = \sum_{t=0}^{\infty} \beta^t Q_\pi^{(t)} r_{\sigma_{t+1}}(x), \quad x \in X, \quad \pi \in M, \quad \beta \in (0,1) \tag{2.8}$$

where

$$Q_\pi^{(0)} := I, \quad Q_\pi^{(t)} := P_{\sigma_1} P_{\sigma_2} \ldots P_{\sigma_t} \quad \text{if } t \geq 1. \tag{2.9}$$

We need some uniquely determined densities corresponding to the operators defined in (2.5) and (2.9) (except $Q^{(0)} = I$). We set

$$\begin{aligned} p_\sigma(x,y) &= \int_A p(x,a,y)\sigma(x,da), \quad q_\pi^{(1)}(x,y) = p_{\sigma_1}(x,y), \\ q_\pi^{(t+1)}(x,y) &= \int_X q_\pi^{(t)}(x,z) p_{\sigma_{t+1}}(z,y)\,dz, \quad (x,y) \in X \times X. \end{aligned} \tag{2.10}$$

Assumption 2.6. *The part (a) of Assumption 2.5 holds, and for every initial state $x_0 \in X$ and every Markov policy $\pi \in M$, the t-step transition densities defined in (2.10) satisfy together with the bounding function μ the following condition: for every $\varepsilon > 0$ there exists a set $X' \subseteq X$ with $m(X') < \infty$ and a constant $B > 0$ such that simultaneously for all $t = 1, 2, 3, \ldots$*

$$\int_{X \setminus X'} \mu(x) q_\pi^{(t)}(x_0, x)\,dx < \varepsilon \tag{2.11}$$

and

$$\mu(x) q_\pi^{(t)}(x_0, x) \leq B \quad \textit{for } x \in X'. \tag{2.12}$$

We use the following definition of Blackwell optimal policies.

Definition 2.1. *For any set $\Pi' \in \Pi$ we say that a policy $\pi^* \in \Pi'$ is Blackwell optimal within the class Π', if for every $x \in X$ and $\pi \in \Pi'$ there exists a number $\beta_0(x,\pi) < 1$ such that $v_\beta(x,\pi^*) \geq v_\beta(x,\pi)$ for all $\beta \in (\beta_0(x,\pi), 1)$. In the case of $\Pi' = \Pi$ we say that π^* is Blackwell optimal.*

The two main existence results of our work are summarized in the following theorems.

Theorem 2.1. *In MDC satisfying Assumptions 2.1–2.5 there exists a deterministic stationary policy φ^* which is Blackwell optimal in the class Σ of stationary policies.*

Theorem 2.2. *In MDC satisfying Assumptions 2.1–2.6 there exists a deterministic stationary policy φ^* which is Blackwell optimal in the class Π of all policies; namely, if φ^* is Blackwell optimal within Σ, then φ^* is Blackwell optimal within Π as well.*

Theorem 2.1 is proved in HY. Theorem 2.2 is a consequence of Theorem 2.1 and Theorem 4.1 of this paper.

3. More on the weak-strong topology

We first recall the needed definitions and results from HY, Sections 5 and 6. In this section only the Assumption 2.1 and the σ-finite measure m on X from Assumption 2.5 are used.

Definition 3.1. (a) *A real-valued function $f(x,a)$, $a \in A_x$, $x \in X$ on K is called a Carathéodory function if f is (Borel) measurable and is continuous in the variable a everywhere on K.*

(b) *The class $Car_0(K)$ consists of all bounded Carathéodory functions on K.*
(c) *The class $Car_m(K)$ consists of Carathéodory functions on K satisfying the condition*

$$\int_X \hat{f}(x)m(dx) < \infty.$$

Definition 3.2. (a) *The space $S = S(K,m)$ consists of all measures s on K satisfying the condition*

$$\mathrm{Pr}_X s = m.$$

(b) *For any positive c, the class $S_c = S_c(K,m)$ consists of all measures s on K satisfying the condition*

$$\mathrm{Pr}_X s \leq cm.$$

(Here $(\mathrm{Pr}_X s)(B) = s((B \times A) \cap K)$ for every $B \in \mathscr{B}_X$.)

Definition 3.3. *The weak-strong topology (ws-topology) in S (or in S_c) is the coarsest topology in which the mapping $s \to \int_K f\, ds$ is continuous in s for every $f \in Car_m(K)$, so that $s_n \xrightarrow{ws} s_\infty$ if and only if*

$$\lim_{n\to\infty} \int_K f\, ds_n = \int_K f\, ds_\infty \quad \text{for every } f \in Car_m(K). \tag{3.1}$$

Lemma 3.1. *The spaces S and S_c are sequentially compact in the ws-topology.*

Lemma 3.2. *There exists a unique mapping $j : \Sigma \to S = S(K, m)$ such that if $s = j(\sigma)$, $\sigma \in \Sigma$ then*

$$s(dx\,da) = \sigma(x, da)m(dx)$$

so that

$$\int_X \int_{A_x} f(x, a)\sigma(x, da)\,dx = \int_K f\,ds,$$

for every measurable and absolutely integrable function f on K.
Moreover, j maps Σ on S.

In this paper we need in addition the n-dimensional analogue of Lemma 3.1 and a specific convergence result.

Consider n copies (X_i, A_i, K_i, m_i) of (X, A, K, m), $i = 1, 2, \ldots, n$. The product spaces $\overline{X} = X_1 \times \cdots \times X_n$, $\bar{A} = A_1 \times \cdots \times A_n$ and $\overline{K} = K_1 \times \cdots \times K_n$ also satisfy Assumption 2.1 (in $\bar{A}$ we consider the product topology). The product measure $\bar{m} = m_1 \times \cdots \times m_n$ also is σ-finite. Let $\overline{S} := S(\overline{K}, \bar{m})$ (see Definition 3.2(a)). Similar to (2.2) and (2.7), for any real-valued function $f(x_1, a_1, \ldots, x_n, a_n)$ on $\overline{K}$ (here $a_i \in A(x_i)$, $x_i \in X_i$) we define

$$\hat{f}(x_1, \ldots, x_n) = \sup |f(x_1, a_1, \ldots, x_n, a_n)|$$

where each a_i runs over the set $A(x_i)$. Applying to $\overline{X}$, $\bar{A}$, $\overline{K}$, $\bar{m}$ and $\overline{S}$ the Definitions 3.1, 3.3 and Lemma 3.1, we have the following result.

Lemma 3.3. *If f is continous in $\bar{a} = (a_1, \ldots, a_n)$ (in the product topology in $\bar{A}$), and if*

$$\int_{\overline{X}} \hat{f}(x_1, \ldots, x_n)\,d\bar{m} < \infty,$$

then the integral

$$J(\bar{s}) = \int_{\overline{K}} f(x_1, a_1, \ldots, x_n, a_n)\,d\bar{s}, \quad \bar{s} \in \overline{S}$$

is continuous in $\bar{s}$ on the compact $\overline{S}$ (in the ws-topology).

For the convergence result we need one more definition.

Definition 3.4. *The space $S_{\text{fin}}(K, m)$ consists of all finite measures s on K with the property: $\text{Pr}_X s$ is absolutely continuous with respect to the measure m on X.*

Lemma 3.4. *Let $s_t \in S_{\text{fin}}(K, m)$, $t = 1, 2, \ldots$, and let u_t be fixed nonnegative Radon-Nikodym derivatives of $\text{Pr}_X s_t$ with respect to m. If for every $\varepsilon > 0$ there exists a set $X' \in \mathscr{B}_X$* with *$m(X') < \infty$ and a constant $B > 0$ such that*

$$\text{Pr}_X s_t(X \backslash X') < \varepsilon, \quad t = 1, 2, \ldots \tag{3.2}$$

and

$$u_t(x) \le B, \qquad x \in X', \quad t = 1, 2, \ldots, \tag{3.3}$$

then there exists a measure $s_\infty \in S_{\mathrm{fin}}(K, m)$ *and a subsquence* $\{s_i'\}$ *of the sequence* $\{s_t\}$ *such that*

$$\lim_{i \to \infty} \int_K f \, ds_i' = \int_K f \, ds_\infty, \qquad f \in Car_0(K). \tag{3.4}$$

Proof. Consider a sequence $\varepsilon_j \downarrow 0$ and sets X_j' and numbers B_j related to ε_j by (3.2)–(3.3). Replacing if necessary B_j by max $(B_1, B_2, \ldots, B_j)$, we may assume that $B_1 \le B_2 \le \cdots$. By replacing after that X_j' by $X_1' \cup \cdots \cup X_j'$, we may assume that also $X_1' \subseteq X_2' \subseteq \cdots$. Define

$$Z_1 = X_1', \quad Z_j = X_j' \backslash X_{j-1}' (j = 2, 3, \ldots), \quad Z_\infty = X \Big\backslash \bigcup_1^\infty X_j'.$$

Then $X = Z_\infty \cup Z_1 \cup Z_2 \cup \cdots$ is a decomposition of X into disjoint sets such that simultaneously for all $t = 1, 2, \ldots$

$$(\mathrm{Pr}_X s_t)(Z_{j+1} \cup Z_{j+2} \cup \cdots \cup Z_\infty) < \varepsilon_j, \tag{3.5}$$

$$u_t(x) \le B_j \quad \text{if } x \in Z_j. \tag{3.6}$$

Here $m(Z_j) < \infty$ for all finite j, and since $\varepsilon_j \downarrow 0$, (3.5) implies that

$$(\mathrm{Pr}_X s_t)(Z_\infty) = 0, \quad t \ge 1. \tag{3.7}$$

Consider now the sets $K_j = K \cap (Z_j \times A)$, $j = \infty, 1, 2, \ldots$ and the corresponding "restricted" measures s_{tj} defined by

$$s_{tj}(M) = s_t(M \cap K_j), \qquad M \in \mathscr{B}_K, \quad j = \infty, 1, 2, \ldots. \tag{3.8}$$

By (3.7) $s_t(K_\infty) = 0$ for all t. Therefore we may set $s_\infty(K_\infty) = 0$, and in proving (3.4) may neglect the set K_∞, i.e. consider the case when K_∞ and Z_∞ are empty sets, so that (3.4) reduces to

$$\lim_{i \to \infty} \sum_{j=1}^\infty \int_{K_j} f \, ds_i' = \sum_{j=1}^\infty \int_{K_j} f \, ds_\infty, \quad f \in Car_0(K). \tag{3.9}$$

The measures $s_{1j}, s_{2j}, \ldots$ are different from 0 only on K_j, and considered on this space they belong to $S_c(K_j, m)$ with $c = B_j$, as follows from (3.6) (cf. Definition 3.2(b)). By Lemma 3.1 the sequence $\{s_{1j}, s_{2j}, \ldots\}$, considered on K_j, has a subsequence converging to some measure $s_{\infty j} \in S_c(K_j, m)$; by setting $s_{\infty j}$ equal to zero on $K \backslash K_j$, we may treat $s_{\infty j}$ as a measure on K too. By an evident diagonal process applied to the set $\{s_{tj}, t \ge 1, j \ge 1\}$, we obtain a subsequence

$\{s'_i\} = \{s_{t_i}\}$ of the sequence $\{s_t\}$ such that

$$s_{ij} \xrightarrow{ws} s_{\infty j} \quad \text{on } K_j, \quad j = 1, 2, \ldots, \tag{3.10}$$

and we define the needed measure s_∞ on K by

$$s_\infty(E) = \sum_{j=1}^{\infty} s_{\infty j}(E) = \sum_{j=1}^{\infty} s_{\infty j}(E \cap K_j), \quad E \in \mathscr{B}_K. \tag{3.11}$$

We have to prove that s_∞ belongs to $S_{\text{fin}}(K, m)$ and that s_∞ satifies (3.9).

The first assertion means that the measure $\text{Pr}_X s_\infty$ is absolute continuous with respect to m, and that the measure s_∞ is finite. To prove the absolute continuity, we observe that, in general, if $s_n \xrightarrow{ws} s_\infty$ in the space $S_c(K, m)$, then $(\text{Pr}_X s_n)(E) \to (\text{Pr}_X s_\infty)(E)$ for each $E \in \mathscr{B}_X$; to see this, apply (3.1) to the indicator of the set $(E \times A) \cap K$. Applied to the convergence (3.10) in the space $S_c(K_j, m)$ with $c = B_j$, this observation together with (3.11) and (3.6) shows that

$$(\text{Pr}_X s_\infty)(E) = (\text{Pr}_X s_{\infty j})(E) = \lim_{i \to \infty} (\text{Pr}_X s'_{ij})(E) \le B_j m(E) \tag{3.12}$$

for every $E \in \mathscr{B}(Z_j)$. Since X is the union of the sets Z_j, (3.12) implies the absolute continuity of $\text{Pr}_X s_\infty$ with respect to m. By the same observation and by (3.11), we have

$$s_\infty(K) = (\text{Pr}_X s_\infty)(X) = \sum_{j=1}^{\infty} (\text{Pr}_X s_{\infty j})(Z_j) = \sum_{j=1}^{\infty} \lim_{i \to \infty} (\text{Pr}_X s'_{ij})(Z_j). \tag{3.13}$$

By (3.8), Fatou's lemma for positive series, (3.11) and (3.2)–(3.3), this implies

$$\begin{aligned} s_\infty(K) &= \sum_{j=1}^{\infty} \lim_{i \to \infty} (\text{Pr}_X s'_i)(Z_j) \le \varliminf_{i \to \infty} \sum_{j=1}^{\infty} (\text{Pr}_X s'_i)(Z_j) \\ &= \varliminf_{i \to \infty} (\text{Pr}_X s'_i)(X) \le \varepsilon + Bm(X') < \infty. \end{aligned} \tag{3.14}$$

It remains to prove (3.9). Since $f \in Car_0(K)$, we have $|f| \le C < \infty$, and therefore

$$\begin{aligned} \left| \int_K f \, ds'_i - \int_K f \, ds_\infty \right| \le \sum_{j=1}^{n} \left| \int_{K_j} f \, ds'_{ij} - \int_{K_j} f \, ds_{\infty j} \right| \\ + C(s'_i + s_\infty) \left(\bigcup_{n+1}^{\infty} K_j \right) \end{aligned} \tag{3.15}$$

for any $n = 1, 2, \ldots$. The convergence (3.10) implies that here the sum from 1 to n tends to 0 when $i \to \infty$. To get (3.9), we have to show that the factor at C

in (3.15) can be made less than ε uniformly in i by the selection of n. Since

$$s_i'\left(\bigcup_{n+1}^{\infty} K_j\right) = (\mathrm{Pr}_X s_i')\left(\bigcup_{n+1}^{\infty} Z_j\right), \quad i = 1, 2, \ldots,$$

this follows from (3.5) and from the relation

$$s_\infty\left(\bigcup_{n+1}^{\infty} K_j\right) \le \varliminf_{i\to\infty} (\mathrm{Pr}_X s_i')\left(\bigcup_{n+1}^{\infty} Z_j\right)$$

obtained similar to (3.13)–(3.14). □

4. Blackwell optimality in the space of all policies

In this section we prove that a stationary policy τ is Blackwell optimal in the space Π of all policies if τ is Blackwell optimal in the space Σ of stationary policies. Together with Theorem 2.1, this result proves Theorem 2.2.

The main idea of the proof was developed in Dekker and Hordijk (1988) for the case of a countable state space X and unbounded rewards. In Yushkevich (1994, 1997) the proof was extended to the case of a Borel state space but bounded rewards, a finite reference measure and a severe Doeblin condition. The present extension to the more general case of a Borel state space, unbounded rewards and less restrictive recurrence conditions is more technical and requires an extra Assumption 2.6. All assumptions of Section 2 are supposed to hold in this section.

We recall the needed results from HY. Assumptions 2.1, 2.2 and 2.4 imply the existence of a constant $\beta_0 < 1$ depending on the numbers C and γ introduced there (or equivalently, a number $\rho_0 > 0$, $\rho_0 = \beta_0^{-1} - 1$) such that for every $\sigma \in \Sigma$ and $\beta_0 < \beta < 1$ (or $0 < \rho < \rho_0$)

$$v_\beta(x,\sigma) = (1+\rho)h_\sigma(x,\rho); \quad h_\sigma(x,\rho) := \sum_{n=-1}^{\infty} h_\sigma^{(n)}(x)\rho^n \tag{4.1}$$

where $\rho = \beta^{-1} - 1$, $h_\sigma^{(n)} \in V_\mu$ and

$$\|h_\sigma^{(n)}\|_\mu \le C_1 C_2^n, \quad n \ge -1 \tag{4.2}$$

(HY, Lemma 3.3). Here and afterwards, by $C_1, C_2, \ldots$ we denote positive constants depending on the parameters of the model; their exact values play no role in the reasoning.

In connection with (4.1)–(4.2) we introduce the space $\mathscr{H}$ of Laurent series in the variable ρ of the form

$$h := h(x,\rho) = \sum_{n=-1}^{\infty} h^{(n)}(x)\rho^n \quad \text{with } h^n \in V_\mu, \quad \varlimsup_{n\to\infty} \|h^{(n)}\|_\mu^{1/n} < \infty, \tag{4.3}$$

and a similar space $\mathscr{G}$ of the series

$$g(x,a,\rho) = \sum_{n=-1}^{\infty} g_n(x,a)\rho^n \quad \text{with} \quad \overline{\lim_{n\to\infty}} \|\hat{g}_n\|_\mu^{1/n} < \infty \tag{4.4}$$

where g_n are measurable functions on K (see (2.2) for the notation $\hat{g}_n$).

For series $\sum_{n=-1}^{\infty} b_n\rho^n$ with real coefficients b_n (converging for small $|\rho| > 0$) we consider the natural ordering equal to the lexicographical ordering of the sequences $\{b_{-1}, b_0, b_1, b_2, \ldots\}$ denoted by the symbols $\succ, \succeq, \prec, \preceq 0$, so that, for example, $\Sigma b_n\rho^n \succ 0$ means that $\sum b_n\rho^n > 0$ for all ρ in some sufficiently small interval $0 < \rho < \rho_1$. The same notation is used for the corresponding partial ordering in the case of Laurent series in ρ with coefficients being functions on X or K. Recall (HY, Proposition 4.1) that one may integrate inequalities $\succeq$ and $\preceq$: if, for example,

$$\sum_{n=-1}^{\infty} g_n(x,a)\rho^n \preceq 0, \quad (x,a) \in K \tag{4.5}$$

and if all functions g_n are integrable with respect to a σ-finite measure s on K, then also

$$\sum_{n=-1}^{\infty} \rho^n \int_K g_n\, ds \preceq 0. \tag{4.6}$$

If there is a collection of Laurent series $f_\alpha(\rho)$ in ρ depending on some parameter α, then by $\operatorname{Lexmax}_\alpha f_\alpha(\rho)$ we denote such of those series (if any) that $\operatorname{Lexmax}_\alpha f_\alpha(\rho) \succeq f_\alpha(\rho)$ for all values of α.

The following operator $L : \mathscr{H} \to \mathscr{G}$ (HY, Section 7) plays an important role: for $h \in \mathscr{H}$ as in (4.3),

$$Lh(x,a,\rho) = \sum_{n=-1}^{\infty} (Lh)_n(x,a)\rho^n \tag{4.7}$$

where

$$\begin{aligned}(Lh)_{-1}(x,a) &= Ph^{(-1)}(x,a) - h^{(-1)}(x),\\ (Lh)_0(x,a) &= r(x,a) + Ph^{(0)}(x,a) - h^{(0)}(x) - h^{(-1)}(x),\\ (Lh)_n(x,a) &= Ph^{(n)}(x,a) - h^{(n)}(x) - h^{(n-1)}(x), \quad n \geq 1.\end{aligned} \tag{4.8}$$

Assumptions 2.2, 2.3 and relations (4.3) show that the functions $(Lh)_n$ in (4.7)–(4.8) are continuous in a and satisfy bounds requested in (4.4), so that if $h \in \mathscr{H}$ then

$$\frac{1}{\mu}(Lh)_n \in Car_0(K), \quad n \geq -1. \tag{4.9}$$

Moreover, if $h = h_\sigma$ for some $\sigma \in \Sigma$ so that (4.2) holds, then

$$|Lh_\sigma^{(n)}(x,a)| \le C_3 C_2^n \mu(x), \quad n \ge -1, \quad (x,a) \in K. \tag{4.10}$$

With every $\sigma \in \Sigma$ we relate the operator $L_\sigma : \mathscr{H} \to \mathscr{H}$ ([HY], Section 4), which can be expressed in the form

$$L_\sigma h(x,\rho) = \int_{A_x} Lh(x,a,\rho)\sigma(x,da),$$

so that

$$L_\sigma h(x,\rho) = \sum_{n=-1}^{\infty} \rho^n (L_\sigma h)^{(n)}(x), \quad (L_\sigma h)^{(n)}(x) = \int_{A_x} (Lh)_n(x,a)\sigma(x,da). \tag{4.11}$$

In Sections 4 and 7 of [HY] (see, in particular, Theorems 4.2, 7.1 and 7.2) we have proved that in MDC satisfying Assumptions 2.1–2.5:

(i) the lexicographical Bellman operator $T : \mathscr{H} \to \mathscr{H}$ is well defined for each $h \in \mathscr{H}$ by either of the two expressions

$$Th(x,\rho) = \operatorname*{Lexmax}_{\sigma \in \Sigma} L_\sigma h(x,\rho) = \operatorname*{Lexmax}_{a \in A_x} Lh(x,a,\rho), \quad x \in X; \tag{4.12}$$

(ii) the Blackwell optimality equation

$$Th = 0, \quad h \in \mathscr{H} \tag{4.13}$$

has a unique solution h^*, and a policy $\tau \in \Sigma$ is Blackwell optimal within Σ if and only if $h_\tau = h^*$;

(iii) there exists a policy $\varphi \in \Phi$ with $h_\varphi = h^*$.

After recalling those results we can move forward. The identities given in the next two lemmas go back to Sladký (1974) and Hordijk and Sladký (1977).

Lemma 4.1. *In MDP satisfying Assumptions 2.1, 2.2 and 2.4 there exists for every Markov policy* $\pi = \{\sigma_1, \sigma_2, \ldots\}$ *and every* $h \in \mathscr{H}$ *a positive number* $\rho_1 = \rho_1(h)$ *such that in the notations*

$$P_t = P_{\sigma_t} \quad (t \ge 1), \qquad Q_t = Q_\pi^{(t)} \quad (t \ge 0)$$

(see (2.4), (2.9)) we have

$$\sum_{n=-1}^{\infty} \rho^{(n)} \left[Q_0 h^{(n)} + \sum_{t=0}^{\infty} \beta^{t+1} (Q_{t+1} h^{(n)} - Q_t h^{(n)} - Q_t h^{(n-1)}) \right] = 0 \tag{4.14}$$

for $0 < \rho < \rho_1$, $\beta = (1+\rho)^{-1}$, *where the double series in (4.14) converges in the* μ*-norm, and where for the uniformity of all terms we set* $h^{(-2)} = 0$.

Proof. The convergence in norm follows from (4.3) and the bound

$$\sum_{t=0}^{\infty} \beta^t Q_t \mu(x) \leq \frac{C_4}{1-\beta} \tag{4.15}$$

obtained in HY, Lemma 2.1. The total sum = 0, because the coefficient at each term $Q_t h^{(n)}$ is equal to

$$\rho^n \beta^t - \rho^n \beta^{t+1} - \rho^{n+1}\beta^{t+1} = \rho^n \beta^t (1 - \beta - \beta\rho) = 0. \qquad \square$$

Lemma 4.2. *Assume the conditions of Lemma 4.1, and denote*

$$r_t = r_{\sigma_t}, \quad L_t = L_{\sigma_t} \qquad (t \geq 1). \tag{4.16}$$

For every $\tau \in \Sigma$ *and* $0 < \rho < \min(\rho_0, \rho_1(h_\tau))$

$$v_\beta(\pi) - v_\beta(\tau) = \sum_{t=0}^{\infty} \beta^t Q_t L_{t+1} h_\tau. \tag{4.17}$$

Proof. By (4.7)–(4.8) and (4.11), (2.4)

$$L_{t+1} h_\tau = r_{t+1} + \sum_{n=-1}^{\infty} \rho^n [P_{t+1} h_\tau^{(n)} - h_\tau^{(n)} - h_\tau^{(n-1)}]$$

(as in Lemma 4.1, $h_\tau^{(-2)} = 0$). Therefore, since $Q_{t+1} = Q_t P_{t+1}$ (cf. (2.4)), we have

$$Q_t L_{t+1} h_\tau = Q_t r_{t+1} + \sum_{n=-1}^{\infty} \rho^n [Q_{t+1} h_\tau^{(n)} - Q_t (h_\tau^{(n)} - h_\tau^{(n-1)})].$$

Hence

$$\sum_{t=0}^{\infty} \beta^t Q_t L_{t+1} h_\tau = \sum_{t=0}^{\infty} \beta^t Q_t r_{t+1} + \sum_{n=-1}^{\infty} \rho^n \sum_{t=0}^{\infty} \beta^t [Q_{t+1} h_\tau^{(n)} - Q_t (h\tau^{(n)} - h_\tau^{(n-1)})].$$

This is equivalent to (4.17) because the first sum on the right side is equal to $v_\beta(\pi)$ (cf. (2.8)), while the second sum according to (4.14) and (4.1) simplifies to

$$-\beta^{-1} \sum_{n=-1}^{\infty} \rho^n Q_0 h_\tau^{(n)} = -\beta^{-1} \sum_{n=-1}^{\infty} h_\tau^{(n)} \rho^n = -\beta^{-1}(1+\rho) v_\beta(\tau) = -v_\beta(\tau). \qquad \square$$

Theorem 4.1. *In MDC satisfying Assumptions 2.1–2.6, a policy* τ, *Blackwell optimal within* Σ, *is Blackwell optimal in the class* Π *as well.*

Proof. Many details of the proof coincide with the corresponding arguments in the preceding papers by Dekker and Hordijk (1988) and Yushkevich (1994, 1997), but for the completeness of the paper we present the proof in full.

As well-known (Strauch (1969)), in a Borelian MDC to every initial state x_0 and policy $\pi \in \Pi$ there corresponds a Markov policy π' such that for every $t \geq 0$ the joint distribution of the pair (x_t, a_{t+1}) is the same for policies π and π' and the given x_0. It follows that in MDC we consider $v_\beta(x_0, \pi) = v_\beta(x_0, \pi')$ for all $\beta \in (0,1)$, and therefore it is sufficient to prove that τ is Blackwell optimal within the class M of Markov policies. To do this, we fix x_0 and a policy $\pi = \{\sigma_1, \sigma_2, \dots\} \in M$, and prove the existence of a number $\beta^* < 1$ $\left(\text{or } \rho^* = \dfrac{1}{\beta^*} - 1 > 0\right)$ such that

$$v_\beta(x_0, \pi) - v_\beta(x_0, \tau) \leq 0 \quad \text{if } \beta^* < \beta < 1 \quad (\text{or } 0 < \rho < \rho^*). \tag{4.18}$$

We use the notations of Lemmas 4.1 and 4.2, and write h in place of $h_\tau = h^*$. By (4.12)–(4.13) $L_{t+1} h_\tau \preceq T h_\tau = 0$, and therefore (4.17) makes the assertion (4.18) rather plausible. However, the proof is very technical. By Lemmas 4.1 and 4.2 there exists $\beta^* < 1$ such that (with $\beta^{-1}(1+\rho) = 1$)

$$v_\beta(x_0, \pi) - v_\beta(x_0, \tau) = \sum_{n=-1}^{\infty} \sum_{t=0}^{\infty} a_{tn} \beta^t \rho^n, \quad \beta^* < \beta < 1 \tag{4.19}$$

where the double series converges absolutely, and according to (4.17), (4.16), (4.11), for every $n \geq 1$

$$\begin{aligned} a_{0n} &= Q_0 (L_1 h)^n (x_0) = \int_A (Lh)^n (x_0, a) \sigma(x_0, da) \\ a_{tn} &= Q_t (L_{t+1} h)^{(n)} (x_0) \\ &= \int_K (Lh)^{(n)} (x, a) \sigma(x, da) q_\pi^{(t)} (x_0, x)\, dx, \quad \text{if } t \geq 1 \end{aligned} \tag{4.20}$$

(we use the densities of operators $Q_t = Q_\pi^{(t)}$ introduced in (2.10) and utilized in Assumption 2.6). By the second form of the operator T in (4.12) and by the equation (4.13), we have

$$\sum_{n=-1}^{\infty} (Lh)^{(n)} (x, a) \rho^n \preceq 0, \quad (x, a) \in K. \tag{4.21}$$

It is convenient to introduce separate series

$$A_n(\beta) = \sum_{t=0}^{\infty} a_{tn} \beta^t, \qquad a_t(\rho) = \sum_{n=-1}^{\infty} a_{tn} \rho^n \qquad (n \geq -1, t \geq 0). \tag{4.22}$$

Then (4.19) takes on the form

$$v_\beta(x_0, \pi) - v_\beta(x_0, \tau) = \sum_{n=-1}^{\infty} A_n(\beta) \rho^n, \tag{4.23}$$

while relations (4.5)–(4.6), applied to (4.21) and to the measure $s(dx\,da)$ present in the integral (4.20), show that

$$a_t(\rho) \preceq 0, \quad t \geq 0. \tag{4.24}$$

There is a trivial case when all the coefficients a_{tn} in (4.19) are zeros: in that case (4.18) holds with the equality sign. Otherwise there exists such $N \geq -1$ that all a_{tn} with $n < N$ are zeros, but at least one of the numbers a_{tN} is different from 0. In that case (4.23) becomes

$$v_\beta(x_0, \pi) - v_\beta(x_0, \tau) = A_N(\beta)\rho^N + A_{N+1}(\beta)\rho^{N+1} + R \tag{4.25}$$

with

$$R = \sum_{n=N+2}^{\infty} A_n(\beta)\rho^n. \tag{4.26}$$

On the other hand, the formula for $a_t(\rho)$ in (4.22) turns into

$$a_t(\rho) = \sum_{n=N}^{\infty} a_{tn}\rho^n$$

where because of (4.24) and the choice of N

$$a_{tN} \leq 0 \quad (t = 0, 1, \ldots), \quad b := \sum_{t=0}^{\infty} a_{tN} < 0. \tag{4.27}$$

Also, from (4.22),

$$\lim_{\beta \uparrow 1} A_N(\beta) = b < 0. \tag{4.28}$$

The cases $b = -\infty$ and $b > -\infty$ are treated in different ways. In the first of them we proceed as in the preceding papers: the two last terms in (4.25) are negligible in comparison with the main term $A_N(\beta)\rho^N$, which, due to (4.28), becomes negative as $\beta \uparrow 1$ (so that $\rho \downarrow 0$). Namely, from the bound (4.10) applied to $\sigma = \tau \in \Sigma$ and from (4.20) we get

$$|a_{tn}| \leq C_3 C_2^n Q_t \mu(x_0), \quad n \geq -1, \quad t \geq 0. \tag{4.29}$$

Therefore, by (4.15) and (4.22),

$$|A_n(\beta)| \leq \frac{C_5 C_2^n}{1 - \beta} = C_5 C_2^n \frac{1 + \rho}{\rho}, \tag{4.30}$$

and by (4.26)

$$|R| \leq C_5(1 + \rho) \sum_{n=N+2}^{\infty} C_2^n \rho^{n-1} \leq C_6 \rho^{N+1}$$

for positive ρ sufficiently close to 0. Utilizing the last bound, we can rewrite (4.25) in the form

$$v_\beta(x_0,\pi) - v_\beta(x_0,\tau) = \rho^N[A_N(\beta) + \rho A_{N+1}(\beta) + o(1)] \quad \text{as } \rho \downarrow 0. \tag{4.31}$$

By (4.30) the term $\rho A_{N+1}(\beta)$ remains bounded as $\rho \downarrow 0$, and the condition $b = -\infty$ together with (4.28) shows that the difference $v_\beta(x_0,\pi) - v_\beta(x_0,\tau)$ becomes negative as $\rho \downarrow 0$.

The case of a finite negative b is the most complicated one. Here we use Assumption 2.6 and the related Lemma 3.4[1]. Formula (4.31) in this case simplifies to

$$v_\beta(x_0,\pi) - v_\beta(x_0,\tau) = \rho^N[b + \rho A_{N+1}(\beta) + o(1)] \quad \text{as } \rho \downarrow 0.$$

To get the needed negativity of the left side, it is sufficient to show that $\lim_{\rho\downarrow 0} \rho A_{N+1}(\beta) \leq 0$. We prove this inequality by contradiction.

Suppose the contrary. Then, by a well-known property of the Abel summation, and since $\rho = (1-\beta)\beta^{-1}$,

$$0 < \overline{\lim_{\beta\uparrow 1}} \frac{1-\beta}{\beta} A_{N+1}(\beta) = \overline{\lim_{\beta\uparrow 1}} (1-\beta) \sum_{t=0}^{\infty} a_{t,N+1}\beta^t \leq \overline{\lim_{t\to\infty}}\, a_{t,N+1}.$$

Hence there are a number ε and a subsequence $\{t_i\}$ of the sequence $\{0,1,2,\ldots\}$ such that

$$a_{t_i,N+1} \geq \varepsilon > 0, \quad i = 1,2,\ldots. \tag{4.32}$$

The definition (4.20) of the coefficients a_{tn} with $t \geq 1$ can be rewritten in the form

$$a_{tn} = \int_K g_n(x,a) s_t(dx\,da), \quad n \geq -1, \quad t \geq 1$$

where

$$g_n(x,a) = \frac{(Lh)^{(n)}(x,a)}{\mu(x)}, \quad s_t(dx\,da) = g_\pi^{(t)}(x_0,x)\mu(x)\sigma(x,da)\,dx \tag{4.33}$$

(the coefficients a_{0n} play no role in the forthcoming reasoning). By (4.21) and since $\mu(x) \geq 1 > 0$

$$g(x,a,\rho) := \sum_{n=-1}^{\infty} g_n(x,a)\rho^n \preceq 0, \quad (x,a) \in K. \tag{4.34}$$

[1] In the proof of the corresponding Theorem 5.4 in Dekker and Hordijk (1988), in the case of a finite b, the Fatou lemma is used without an explanation why it is applicable. Indeed at this point the uniform integrability as in Assumption 2.6 is needed, which in the countable case can be deduced from the Assumptions 2.2, 2.3 and 2.4 (see Lemmas 4.6 and 4.7 in Dekker, Hordijk and Spieksma (1994)). In Yushkevich (1994), (1997) a similar problem does not arise because of severe boundedness assumptions.

By (4.9) $g_n \in Car_0(K)$. Also $\Pr_X s_t(dx) = q_\pi^{(t)}(x_0, x)\mu(x)\,dx$, so that the functions $u_t(x) := q_\pi^{(t)}(x_0, x)\mu(x)$ can be taken for Radon-Nikodym derivatives $d\Pr_X s_t/dm$. By Assumption 2.6 there exist a number $B > 0$ and a subset $X' \subset X$ of finite measure m such that (uniformly in $t = 1, 2, \ldots$)

$$s_t(K) = \Pr_X s_t(X) \le Bm(X') + \varepsilon < \infty,$$

so that $s_t \in S_{\text{fin}}(K, m)$ (cf. Definition 3.4). By Assumption 2.6, all the conditions of Lemma 3.4 are satisfied. According to this lemma, there exists a subsequence of the sequence $\{t_i\}$, which we denote $\{T_j\}$, and a finite measure s_∞ on K such that for each $n = -1, 0, 1, 2, \ldots$

$$a_{\infty n} := \int_K g_n\,ds_\infty = \lim_{j\to\infty} \int_K g_n\,ds_{T_j} = \lim_{j\to\infty} a_{T_j n}. \tag{4.35}$$

The Laurent series

$$a_\infty(\rho) = \sum_{n=-1}^{\infty} a_{\infty n}\rho^n$$

converges for small $|\rho| > 0$, as follows from (4.35), (4.33) and bounds (4.10): $|a_{\infty n}| \le C_3 C_2^n s_\infty(K)$. From (4.34) and (4.35) we have

$$a_\infty(\rho) = \int_K g(x, a, \rho)\,ds_\infty \preceq 0 \tag{4.36}$$

(cf. (4.5)–(4.6)). On the other hand, since a_{tn} with $n \le N$ are zeros, by (4.35) $a_{\infty n}$ with $n < N$ also are zeros. The same is true for $a_{\infty N}$, because the series defining b in (4.27) converges, so that $a_{tN} \to 0$ as $t \to \infty$. Thus $a_\infty(\rho) = a_{\infty,N+1}\rho^{N+1} + a_{\infty,N+2}\rho^{N+2} + \cdots$ where according to (4.32) and (4.35) $a_{\infty,N+1} \ge \varepsilon > 0$, and this contradicts with (4.36). □

Finally, the existence of a deterministic stationary Blackwell optimal policy in the space Π of all policies (Theorem 2.2) follows immediately from Theorems 2.1 and 4.1.

5. Recurrence conditions for Blackwell optimality

The uniform ergodicity and integrability conditions used in [HY] and Section 4 for Blackwell optimality (Assumptions 2.4 and 2.6) are to complicated for a straightforward verification in specific models with a noncompact state space X and an unbounded reward function. In this section we consider recurrence-type conditions more suitable for applications, which imply Assumptions 2.4 and 2.6 (if other assumptions of Section 2 hold). We refer the reader to the paper by Hernández-Lerma et al. (1991) for a comprehensive survey of recurrence and ergodicity conditions in a more general context of Markov and controlled Markov chains. Assumptions 2.1–2.3 and 2.5 are supposed to be satisfied in this section. Main results of this section are summarized in Theorems 5.1 and 5.2 and Corollary 5.1.

The following *uniform minorant condition* is stronger than the existence of a uniformly small set as defined in Hordijk, Spieksma and Tweedie (1995) (a $\mathscr{T}_1$-set in their terminology).

Assumption 5.1. *There exist two sets* $D, X' \in \mathscr{B}_X$ *with* $m(D) > 0$, $m(X') > 0$ *and a number* δ *such that the transition density*

$$p(x, a, y) \geq \delta > 0 \quad \text{for } x \in D, \quad a \in A_x, \quad y \in X'.$$

We mention that in the particular case $D = X$ this assumption is precisely the simultaneous Doeblin-Doob condition used in Yushkevich (1997)(cf. HY, (2.31)–(2.32)), which is much more restrictive and often fails to hold in models with an unbounded set X.

The next assumption is a generalization to the case of a Borel space X of the *μ-uniform* (geometric) *recurrence condition* from Dekker and Hordijk (1992).

Assumption 5.2. *There exist: 1) a set* $D \in \mathscr{B}_X$ *with* $m(D) > 0$ *and*

$$\sup_{x \in D} \mu(x) < \infty, \tag{5.1}$$

and 2) a number $0 < \alpha < 1$, *such that*

$$\int_{X \setminus D} p(x, a, y)\mu(y)\, dy \leq \alpha\mu(x) \quad \text{for all } (x, a) \in K. \tag{5.2}$$

The following *uniform drift condition* is closely related to Assumption 5.2, as stated in Lemma 5.1 below.

Assumption 5.3. *There exist: 1) a set D as in Assumption 5.2, and 2) numbers* $0 < \gamma < 1$, $b > 0$, *such that*

$$P\mu(x, a) \leq \gamma\mu(x) + b \cdot 1_D(x) \quad \text{for all } (x, a) \in K. \tag{5.3}$$

Lemma 5.1. *Assumption 5.3 implies Assumption 5.2 with the same set D and with* μ *replaced by* μ^*, *where*

$$\mu^*(x) = \mu(x) + b1_D(x). \tag{5.4}$$

Moreover, with μ *replaced by* μ^*, *Assumptions 2.1–2.3, 2.5 remain valid, only with maybe a larger constant C.*

Proof. It is the same as in the case of a countable X; see Hordijk and Spieksma (1992), pp. 350–351. We only mention for more clarity that $\mu \leq \mu^* \leq (1 + b)\mu$ (as follows from (5.4) and the condition $\mu \geq 1$), so that $\|f\|_{\mu^*} \leq \|f\|_{\mu} \leq (1 + b)\|f\|_{\mu^*}$ for every function f on X. Therefore relations (5.3)–(5.4) imply (5.2) with

$$\alpha = \max\left(\gamma, \frac{\gamma + b}{1 + b}\right). \qquad \square$$

The next *uniform accessibility condition* is introduced in Hordijk, Spieksma and Tweedie (1995).

Assumption 5.4. *There exists a set* $D \in \mathscr{B}_X$ *such that for every sublevel set*

$$M_c = \{x \in X : \mu(x) \leq c\} \tag{5.5}$$

there are an integer $N \geq 1$ *and a number* η *such that uniformly in* $x \in M_c$ *and* $\sigma \in \Sigma$

$$P^N_\sigma(x, D) := \mathbf{P}^\sigma_x\{x_N \in D\} \geq \eta > 0. \tag{5.6}$$

Utilizing results from Meyn and Tweedie (1993), Hordijk, Spieksma and Tweedie (1995) proved equivalence results for various conditons on a collection of Markov chains. The Key Theorem in their paper implies, in particular, the following result stating uniform ergodicity of Markov chains generated by stationary policies, needed for applications of our theorems on Blackwell optimality.

Theorem 5.1. *Assumptions 5.1, 5.3 and 5.4 with the same set D, together with Borel measurability of the model and Assumptions 2.2(b) and 2.5(a), imply Assumption 2.4.*

Proof. See the above reference. □

We now turn to conditions guaranteeing Assumption 2.6. A more standard form of the uniform integrability than in (2.11)–(2.12) is given in the following assumption.

Assumption 5.5. (a) *For every* $c \geq 1$ *we have (cf. (5.5))*

$$m(M_c) < \infty. \tag{5.7}$$

(b) *For every* $x_0 \in X$

$$\lim_{c \to \infty} \sup_{t \geq 1} \sup_{\pi \in \mathrm{M}} \int_{X \setminus M_c} \mu(x) q^{(t)}_\pi(x_0, x)\, dx = 0. \tag{5.8}$$

Lemma 5.2. *If for every* $x_0 \in X$ *and* $\pi \in \mathrm{M}$ *the densities* $q^{(t)}_\pi(x_0, x)$, $x \in X$, $t \geq 1$ *are uniformly bounded (in particular, if the transition density* $p(x, a, y)$ *is bounded), then Assumption 5.5 implies Assumption 2.6.*

Proof. If $p(x, a, y)$ is bounded by a constant C_1, then $q^{(t)}_\pi(x, y)$ is also bounded by the same C_1; this is a direct consequence of equations (2.10), proved by induction in t. To get (2.11)–(2.12) from (5.7)–(5.8), it is sufficient to take $X' = M_c$ for a sufficiently large c, and to set $B = C_1 c$. □

Finally, in Theorem 5.2 we show that the μ-uniform recurrence condition (Assumption 5.2) together with the following *dominance-integrability condition* for the transition density $p(x, a, y)$ imply the uniform integrability as stated in Assumption 5.5. The proof of this theorem follows the pattern of the proof of Lemma 4.7 in Dekker, Hordijk and Spieksma (1994), with adjustments due to a Borel space X.

Assumption 5.6. *Assumption 5.5(a) holds, and there exists a measurable function* $\ell \geq 0$ *such that*

$$\int_X \ell(x)\mu(x)\,dx < \infty \quad \text{and} \quad \hat{p}(x,y) \leq \ell(y), \quad x \in D, y \in X \tag{5.9}$$

(D is the set from Assumption 5.2).

Theorem 5.2. *Assumptions 2.1, 2.5, 5.2 and 5.6 imply Assumption 5.5.*

Proof. We need to show that the $\sup_{t\geq 1}\ldots$ in (5.8) becomes less than $\varepsilon > 0$ when $c \to \infty$. From (5.1) and (5.5) we have

$$D \subseteq M_c \quad \text{if } c \geq c_0 = \sup_{x \in D} \mu(x). \tag{5.10}$$

To evaluate the integral in (5.8), we apply the last exit decomposition with respect to the set D. If $\pi = \{\sigma_1, \sigma_2, \ldots\}$ and $c \geq c_0$, then, due to (5.10),

$$\begin{aligned}\int_{X \setminus M_c} &\mu(x) q_\pi^{(t)}(x_0, x)\,dx \\ &= \int_{X_1} \cdots \int_{X_{t-1}} \int_{\overline{M}_t} p_{\sigma_1}(x_0, x_1) \ldots p_{\sigma_t}(x_{t-1}, x_t)\mu(x_t)\,dx_1 \ldots dx_t \\ &= \sum_{k=1}^{t} I_k \end{aligned} \tag{5.11}$$

where

$$I_1 = \int_{\bar{D}_1} \cdots \int_{\bar{D}_{t-1}} \int_{\overline{M}_t} F(z)\,dz,$$

$$I_2 = \int_{D_1} \int_{\bar{D}_2} \cdots \int_{\bar{D}_{t-1}} \int_{\overline{M}_t} F(z)\,dz,$$

$$I_k = \int_{X_1} \cdots \int_{X_{k-2}} \int_{D_{k-1}} \int_{\bar{D}_k} \cdots \int_{\bar{D}_{t-1}} \int_{\overline{M}_t} F(z)\,dz, \quad 3 \leq k \leq t-1,$$

$$I_t = \int_{X_1} \cdots \int_{X_{t-2}} \int_{D_{t-1}} \int_{\overline{M}_t} F(z)\,dz,$$

and where: (a) for clarity, X_k, D_k, M_k denote identical copies of the sets X, D, M_c corresponding to the integration with respect to dx_k, and $\overline{E}$ is the complement of $E \subseteq X$; (b) for the compactness of formulas

$$F(z)\,dz := p_{\sigma_1}(x_0, x_1) \dots p_{\sigma_t}(x_{t-1}, x_t)\mu(x_t)\,dx_1 \dots dx_t \tag{5.12}$$

(if $t < 4$, some of the terms are absent; we leave this case to the reader, and perform the calculations for $t \geq 4$).

Let

$$b_1 = \max(c_0, \mu(x_0)) \tag{5.13}$$

and choose $N = N(\varepsilon, x_0, c_0)$ so large that

$$\frac{\alpha^N}{1-\alpha} b_1 < \frac{\varepsilon}{2} \tag{5.14}$$

where α is the constant in (5.2). Since $\overline{M}_t \subseteq \bar{D}$ by (5.10), we have

$$I_1 \leq \int_{\bar{D}_1} \dots \int_{\bar{D}_t} F(z)\,dz,$$

and from (5.2) and (5.12) we obtain by induction

$$I_1 \leq \alpha^t \mu(x_0). \tag{5.15}$$

By a similar reasoning we obtain from (5.2) and (5.10) for all $k > 1$

$$I_k \leq \int_{X_1} \dots \int_{X_{k-2}} \int_{D_{k-1}} \alpha^{t-k+1} p_{\sigma_1}(x_0, x_1) \dots p_{\sigma_{k-1}}(x_{k-2}, x_{k-1})\mu(x_{k-1})\,dx_1 \dots dx_{k-1}$$

(in the case $k = 2$, integrals over X should be skipped). By (5.10) $\mu(x_{k-1}) \leq c_0$ on the set D_{k-1}, so that we have

$$I_k \leq c_0 \alpha^{t-k+1}, \quad k = 2, 3, \dots, t. \tag{5.16}$$

From (5.13) and (5.15)–(5.16), and then (5.14) we get the bound

$$I_1 + I_2 + \dots + I_{t-N+1} \leq b_1(\alpha^t + \alpha^{t-1} + \dots + \alpha^N) < \frac{\varepsilon}{2} \quad \text{if } c \geq b_1 \tag{5.17}$$

(if $t < N$, (5.17) is also true, with the left side equal 0).

The terms in (5.11) with $k \geq k_0 = \max(t - N + 2, 2)$ we evaluate by means of Assumption 5.6. From the definition of those I_k and (5.9) we have

$$\begin{aligned} I_k &\leq \max_{x \in D_{k-1}} \int_{\bar{D}_k} \dots \int_{\bar{D}_{t-1}} \int_{\overline{M}_t} p_{\sigma_k}(x, x_k) \dots p_{\sigma_t}(x_{t-1}, x_t)\mu(x_t)\,dx_k \dots dx_t \\ &\leq \int_{\bar{D}_k} \dots \int_{\bar{D}_{t-1}} \int_{\overline{M}_t} \ell(x_k) p_{\sigma_{k+1}}(x_k, x_{k+1}) \\ &\qquad \dots p_{\sigma_t}(x_{t-1}, x_t)\mu(x_t)\,dx_k \dots dx_t. \end{aligned} \tag{5.18}$$

In (5.18) we do the following: (1) change the domains of integration to $X_k, \dots, X_t$ by multiplying the integrand by the indicators of the sets $\overline{D}_k, \dots, \overline{M}_t$: (2) consider copies $A_{k+1}, \dots, A_{t+1}$ of the space A, and using (2.10) represent the factors p_σ as

$$p_{\sigma_{i+1}}(x_i, x_{i+1}) = \int_{A_{i+1}} p(x_i, a_{i+1}, x_{i+1})\sigma_{i+1}(x_i, da_{i+1}),$$

$i = k, \dots, t-1$; (3) fix an arbitrary $\sigma_{t+1} \in \Sigma$, and insert an additional factor

$$1 = \int_{A_{t+1}} \sigma_{t+1}(x_t, da_{t+1})$$

into the integrand; (4) regroup the integrations in the alternating order: over A_{t+1} with respect to $\sigma_{t+1}(x_t, da_{t+1})$, then over X_t with respect to dx_t, then over A_t with respect to $\sigma_t(x_{t-1}, da_t), \dots,$ finally over X_1 with respect to dx_1; (5) combine, using Lemma 3.2, iterated integrations over A_{i+1} and then X_i into one integral over a copy K_i of the space K with respect to the measure $s_i = j(\sigma_{i+1})$, starting from $i = t$ and up to $i = k$. Then (5.18) becomes

$$I_k \le J_c(s_k, \dots, s_t) = \int_{\overline{K}} f_c \, d\bar{s}, \quad \overline{K} = K_k \times \cdots K_t, \tag{5.19}$$

where

$$\begin{aligned} f_c(x_k, a_{k+1}, \dots, x_t) &= 1_{\overline{D}}(x_k)\ell(x_k)p(x_k, a_{k+1}, x_{k+1}) \\ &\qquad \dots 1_{\overline{D}}(x_{t-1})p(x_{t-1}, a_t, x_t)1_{\overline{M}_c}(x_t)\mu(x_t), d\bar{s} \\ &= \bar{s}(dx_k \, da_{k+1} \dots dx_t \, da_{t+1}) \\ &= s_k(dx_k \, da_{k+1}) \dots s_t(dx_t, da_{t+1}). \end{aligned} \tag{5.20}$$

Our goal now is the continuity of $J_c(\bar{s})$ in $\bar{s}$. As obtained from (5.18), formula (5.20) defines $J_c(\bar{s})$ on the product space $\tilde{S} = S_k \times \cdots \times S_t$ (here S_j are copies of S). However, the integrand in (5.19) is a nonnegative measurable function on $\overline{K}$, so that $J_c(\bar{s})$ has sense on the whole space $\overline{S} = S(\overline{K}, \overline{m})$ where $\overline{m} = m_k \times \cdots \times m_t$, m_j being copies of m. Recall that a measure $\bar{s}$ on $\overline{K}$ belongs to $\overline{S}$ if $\mathrm{Pr}_{\overline{X}}\bar{s} = \overline{m}$ (where $\overline{X} = X_1 \times \cdots \times X_t$), so that evidently $\tilde{S} \subset \overline{S}$.

By Lemma 3.1 applied to $\overline{K}$, the space $\overline{S}$ is compact in the ws-topology. By Lemma 3.3, the function $J_c(\bar{s})$ is continuous in $\bar{s} \in \overline{S}$ if only the integrand f_c in (5.19) is contiuous in $\bar{a} = (a_{k+1}, \dots, a_{t+1})$ and satisfies the integrability condition $\int_{\overline{X}} \hat{f}_c \, d\overline{m} < \infty$. The continuity of f_c in $\bar{a}$ follows from the continuity of $p(x, a, y)$ in a (Assumption 2.5). Next,

$$\hat{f}_c(x_k, \dots, x_t) \le \ell(x_k)\hat{p}(x_k, x_{k+1}) \dots \hat{p}(x_{t-1}, x_t)\mu(x_t)$$

so that by applying recursively (2.6) (Assumption 2.5) and then (5.9) we have

$$\int_{\bar{X}} \hat{f}_c d\bar{m} = \int_{X_k} \dots \int_{X_t} \ell(x_k)\hat{p}(x_k, x_{k+1}) \dots \hat{p}(x_{t-1}, x_t)\mu(x_t)\, dx_k \dots dx_t$$

$$\leq C^{t-k} \int_{X_k} \ell(x_k)\mu(x_k)\, dx_k < \infty.$$

Thus, for each $c \geq 1$, the function $J_c(\bar{s})$ is continuous on the compact space $\bar{S}$. Also, since the integral in (5.20) defining $J_c(\bar{s})$ converges absolutely, and since the factor $1_{\overline{M}_c}(x_t)$ of the integrand monotonically decreases to 0 as $c \to \infty$ (cf. (5.5)), we have

$$\lim_{c\to\infty} J_c(\bar{s}) = 0, \quad \bar{s} \in \bar{S}.$$

Hence, by Dini's theorem,

$$\lim_{c\to\infty} \max_{\bar{s} \in \bar{S}} J_c(\bar{s}) = 0.$$

This, together with the uniform in k bound (5.19), implies the existence of a number b_2 such that

$$I_k < \frac{\varepsilon}{2N}, \quad k = k_0, k_0 + 1, \dots, t \quad \text{if } c \geq b_2.$$

Here $k_0 = \max(t - N + 2, 2)$, so that the number of those terms is less than N, and therefore

$$\sum_{k=k_0}^{t} I_k < \frac{\varepsilon}{2} \quad \text{if } c \geq b_2.$$

This, together with (5.17) shows that the integral in (5.11) and (5.8) is less than ε if $c \geq \max(b_1, b_2)$, uniformly in t and π, so that (5.8) holds. □

Corollary 5.1. *In MDC with a bounded transition density $p(x, a, y)$ Assumptions 2.1–2.3, 2.5 and 5.1, 5.3, 5.4, 5.6 imply the whole set of Assumptions 2.1–2.6.*

Proof. Follows from Lemma 5.2 and Theorems 5.1, 5.2. □

6. Cash-balance model

In this section we consider a discrete-time cash-balance model in which the rate of return is controlled, whereas the risk parameter is fixed.

The evolution of the proces $\{x_t\}$ is governed by the equation

$$x_t = x_{t-1} + a_t + W_t \tag{6.1}$$

where x_t is the state at time t, a_t is the control parameter and W_t are independent standard normal random variables. Here the state x has the meaning of the current cash balance, while the action a corresponds to a withdrawal of size $-a$ (if $a < 0$) of the money in cash, or to a supply in the amount a (if $a > 0$).

Linear systems of the type (6.1) and their multi-dimensional generalizations are well known in MDC, especially in the case of the Gaussian noise W_t and the quadratic cost criterion; see, for example, Kushner (1971), or for the one-dimensional case, Dynkin and Yushkevich (1979). Our model is a special case of the controlled linear system studied in Meyn (1997). For a continuous-time cash-balance model see van Dijk and Hordijk (1996) and references there.

We now describe the elements of the model and introduce conditions that guarantee all the assumptions of the preceding sections, so that there exists a Blackwell optimal policy.

The state space is evidently $X = R = (-\infty, \infty)$. The action sets A_x are closed intervals in R

$$A_x = \{a : a_e(x) \le a \le a_u(x)\} \tag{6.2}$$

where a_e and a_u are Borel-measurable functions from R to R with $a_e(x) \le a_u(x)$ for every x. To satisfy recurrence conditions of Section 5, we need to suppose that these two functions are bounded, say

$$-M \le a_e(x) \le a_u(x) \le M \tag{6.3}$$

for some constant $M > 0$, and that

$$\overline{\lim_{x\to+\infty}}\, a_u(x) < -\frac{1}{2}, \quad \underline{\lim_{x\to-\infty}}\, a_e(x) > \frac{1}{2}. \tag{6.4}$$

The last condition together with (6.1) assures a sufficient drift towards the origin from the remote states x[2]. For the action space A according to (6.3) we may take the interval $A = [-M, M]$ (or any larger interval, for instance the whole R). In X and A we consider the usual Euclidian metrics and the corresponding topology and Borel σ-algebras of measurable sets.

As reference measure on X we take the Lebesgue measure. Then in accordance with (6.1) the transition density is

$$p(x, a, y) = \varphi(y - x - a), \quad a \in A_x, \quad x \in R \tag{6.5}$$

where

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in R \tag{6.6}$$

is the standard normal density. As bounding function we take the even function

$$\mu(x) = e^x + e^{-x}, \quad x \in R \tag{6.7}$$

[2] Note that in the approach of Meyn (1997) this uniformly drift condition is relaxed for average optimality. It is an interesting question whether the Meyn's conditions also imply the convergence of the Howard-Blackwell-Veinott policy improvement method for sensitive optimality criteria.

The reward function $r(x, a)$ may be any Borel function on the set $K = \{(x, a) : a \in A_x, x \in R\}$ continuous in a with $|r(x, a)| \leq C\mu(x)$, $x \in R$ for some constant $C > 0$.

To check the assumptions of Sections 2 and 5, it is convenient to prepare some elementary formulas.

Lemma 6.1. *For every $b \in R$*

$$\int_R e^y \varphi(y + b)\, dy = \int_R e^{-y} \varphi(y - b)\, dy = e^{(1/2)-b}. \tag{6.8}$$

Proof. Use the substitutions $x \pm b = z$ and the fact that (6.6) is a probability density. □

Lemma 6.2. *Consider the functions*

$$F_h(z) = \begin{cases} \varphi(z + h), & \text{if } z \leq -h \\ \varphi(0), & \text{if } -h \leq z \leq h \\ \varphi(z - h), & \text{if } z \geq h \end{cases} \tag{6.9}$$

with $h \geq 0$. Then

$$\hat{p}(x, y) := \max_{a \in A_x} p(x, a, y) \leq F_M(y - x), \quad x \in R, \quad y \in R. \tag{6.10}$$

Proof. By (6.5) and (6.2)–(6.3)

$$\hat{p}(x, y) \leq \max_{|a| \leq M} \varphi(y - x - a),$$

and (6.6) follows from the fact that $\varphi(z)$ is increasing on the negative half-axis and is decreasing on the positive one. □

Lemma 6.3. *For the functions F_h defined in (6.9) and any $B \geq 0$*

$$\max_{|x| \leq B} F_h(y - x) = F_{B+h}(y), \quad y \in R. \tag{6.11}$$

Proof. Compare the interval $-B \leq x \leq B$ with the intervals $-\infty < x \leq y - h$, $y - h \leq x \leq y + h$, and $y + h \leq x < \infty$ where the function $f(x) = F_h(y - x)$ is respectively, increasing, constant$= \varphi(0)$, and decreasing. □

Theorem 6.1. *In the cash-balance model, the transition density is bounded and all the Assumptions 2.1–2.3, 2.5 and 5.1, 5.3–5.4, 5.6 hold, so that by Corollary 5.1 and Theorem 2.2 there exits a deterministic stationary Blackwell optimal policy.*

Proof. Evidently, $p(x, a, y)$ is bounded. We first verify the assumptions of Section 2. Assumption 2.1 (measurability and compactness), 2.2(a) (μ-boundedness of r), 2.3(a) (continuity of r in a), 2.5(a,b) (existence of a transition density continuous in a) hold trivially. Assumptions 2.2(b) (μ-boundedness of the operator P) and 2.3(b) follow from 2.5(a,b,c). Thus it remains to verify only Assumption 2.5(c) stating in our case that

$$\int_R \hat{p}(x,y)\mu(y)\,dy \le C\mu(x), \quad x \in R \tag{6.12}$$

for some number $C > 0$.

From (6.9)–(6.10) we have

$$\begin{aligned}\int_R \hat{p}(x,y)\mu(y)\,dy \\ &\le \int_R \mu(y)F_M(y-x)\,dy \\ &= \int_{-\infty}^{x-M} \mu(y)\varphi(y-x+M)\,dy + \int_{x-M}^{x+M} \mu(y)\varphi(0)\,dy \\ &\quad + \int_{x+M}^{+\infty} \mu(y)\varphi(y-x-M)\,dy. \end{aligned}\tag{6.13}$$

The second integral at the right side is, by (6.7),

$$\varphi(0)\int_{x-M}^{x+M} (e^y + e^{-y})\,dy = (e^x + e^{-x})(e^M - e^{-M}) \le e^M \mu(x).$$

The first of the integrals in (6.13) we treat by Lemma 6.1. It is less than

$$\begin{aligned}\int_R (e^y + e^{-y})\varphi(y-x+M)\,dy &= e^{(1/2)+x-M} + e^{(1/2)-x+M} \\ &\le e^{(1/2)+M}(e^x + e^{-x}) = e^{(1/2)+M}\mu(x).\end{aligned}$$

For the third of the integrals in a similar way we have the bound

$$\int_R (e^y + e^{-y})\varphi(y-x-M)\,dy = e^{(1/2)+x+M} + e^{(1/2)-x-M} \le e^{(1/2)+M}\mu(x).$$

Thus, by (6.13), (6.12) holds with $C = (1 + 2e^{1/2})e^M$.

Now we turn to the assumptions of Section 5. Assumption 5.1, stating that $p(x,a,y) \ge \delta > 0$ for $x \in D$, $a \in A_x$, $y \in X'$ where $m(D) > 0$ and $m(X') > 0$, trivially holds for any interval $D = [-B,B]$ and $X' = D$ with $B > 0$, since on the compact $-B \le x, y \le B$, $-M \le a \le M$ the normal density $\varphi(y-x-a)$ is bounded from 0 (cf. (6.5) and (6.3)). We will select the set $D = [-B,B]$ in the next paragraph, where we consider Assumption 5.3 stating that

$$P\mu(x,a) \le \gamma\mu(x) + b \cdot 1_D(x), \quad (x,a) \in K \tag{6.14}$$

for some $\gamma \in (0,1)$ and $b > 0$.

By (6.8) and (6.2)–(6.3) we have

$$\begin{aligned} P\mu(x,a) &= \int_R (e^y + e^{-y})\varphi(y-x-a)\,dy \\ &= e^{(1/2)+x+a} + e^{(1/2)-x-a} \\ &\le e^{(1/2)+x+a_u(x)} + e^{(1/2)-x+M}, \quad a \in A_x, \quad x \in R \end{aligned} \tag{6.15}$$

and therefore, by (6.4),

$$\begin{aligned} \overline{\lim_{x\to+\infty}} \frac{\max_a P\mu(x,a)}{\mu(x)} &\le \overline{\lim_{x\to+\infty}} \frac{e^{(1/2)+x+a_u(x)} + e^{(1/2)-x+M}}{e^x + e^{-x}} \\ &= \overline{\lim_{x\to+\infty}} e^{(1/2)+a_u(x)} < 1. \end{aligned}$$

A similar reasoning with the bound (6.4) for $a_e(x)$ shows that

$$\overline{\lim_{x\to-\infty}} \frac{\max_a P\mu(x,a)}{\mu(x)} < 1.$$

Hence there exist numbers $\gamma \in (0,1)$ and $B > 0$ such that

$$P\mu(x,a) \le \gamma\mu(x) \quad \text{if } |x| \ge B. \tag{6.16}$$

To obtain (6.14) from (6.16), it remains to set

$$D = [-B, B], \quad b = \sup_{\substack{a\in A_x \\ x\in D}} P\mu(x,a) \tag{6.17}$$

(by (6.15) and (6.3), $P\mu(x,a)$ is bounded if x is bounded).

Assumption 5.4 states that for every $c > 0$ there exist such N and η that

$$P_\sigma^N(x, D) \ge \eta > 0 \quad \text{if } x \in M_c,\ \sigma \in \Sigma$$

(here $M_c = \{x : \mu(x) \le c\}$). It holds trivially with $N = 1$ and

$$\eta = \inf \int_D p(x,a,y)\,dy \qquad \text{over} \quad a \in A_x,\ x \in M_c$$

because $m(D) > 0$ and $p(x,a,y) = \varphi(y-x-a)$ is bounded from 0 when x, a, y run over the bounded set $y \in D$, $x \in M_c$, $a \in A_x$.

The last Assumption 5.6 requires the sets M_c to be of finite measure (what trivially holds in the case of $\mu(x) = e^x + e^{-x}$ and the Lebesgue measure on R), and the existence of a Borel function $\ell(y)$ such that

$$\hat{p}(x,y) \le \ell(y), \quad x \in D,\ y \in R \quad \text{and} \quad \int_R \ell(y)\mu(y)\,dy < \infty. \tag{6.18}$$

By (6.10)–(6.11) and (6.17) we have the first of the relations (6.18) for the function $\ell(y) = F_{B+M}(y)$ (cf. (6.9)). Finally, the integral in (6.18) converges for $\ell = F_{B+M}$ because $F_{B+M}(y)$ is decaying as $e^{-y^2/2}$ as $|y| \to \infty$, while $\mu(y)$ is growing as $e^{|y|}$. $\square$

References

[1] Dekker R, Hordijk A (1988) Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. Math Oper Res 13:395–420

[2] Dekker R, Hordijk A (1992) Recurrence conditions for average and Blackwell optimality in denumerable state Markov decision chains. Math Oper Res 17:271–289

[3] Dekker R, Hordijk A, Spieksma FM (1994) On the relation between recurrence and ergodicity properties in denumerable Markov decision chains. Math Oper Res 19:1–21

[4] Van Dijk NM, Hordijk A (1996) Time discretization for controlled Markov processes Part II: A jump and diffusion application. Kybernetika (Prague) 32:139–158

[5] Dynkin EB, Yushkevich AA (1979) Controlled Markov processes. Springer-Verlag

[6] Hernandéz-Lerma O, Montes-de-Oca R, Cavazos-Cadena R (1991) Recurrence conditions for Markov decision processes with Borel state space: a survey. Ann Oper Res 28:29–46

[7] Hordijk A, Sladký K (1977) Sensitive optimality criteria in countable state dynamic programming. Math Oper Res 2:1–14

[8] Hordijk A, Spieksma FM (1992) On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. Adv Appl Prob 24:343–376

[9] Hordijk A, Spieksma FM, Tweedie RL (1995) Uniform stability conditions for general space Markov decision processes. Technical report, Leiden University and Colorado State University

[10] Hordijk A, Yushkevich AA (1999) Blackwell optimality in the class of stationary policies in Markov decision chains with a Borel state space and unbounded rewards. Math Meth Oper Res 49:1–39

[11] Kushner H (1971) Introduction to stochastic control. Holt, New York

[12] Meyn SP (1997) The policy improvement algorithm for Markov decision processes with general state space. Transactions on Automatic Control AC-42:191–196

[13] Meyn SP, Tweedie RL (1993) Markov chains and stochastic stability. Springer-Verlag

[14] Sladký K (1974) On the set of optimal controls for Markov chains with rewards. Kybernetika (Prague) 10:350–367.

[15] Strauch RE (1969) Negative dynamic programming. Ann Math Stat 37:871–890

[16] Yushkevich AA (1994) Blackwell optimal policies in a Markov decision process with a Borel state space. Z Oper Res 40:253–288

[17] Yushkevich AA (1997) Blackwell optimality in Borelian continuous in action Markov decision processes. SIAM J Control 35:2157–2182