Adv. Appl. Prob. **47,** 1088–1107 (2015)
Printed in Northern Ireland

COUNTABLE STATE MARKOV DECISION PROCESSES WITH UNBOUNDED JUMP RATES AND DISCOUNTED COST: OPTIMALITY EQUATION AND APPROXIMATIONS

H. BLOK [*] [**] AND
F. M. SPIEKSMA,[*] [***] *Leiden University*

Abstract

This paper considers Markov decision processes (MDPs) with unbounded rates, as a function of state. We are especially interested in studying structural properties of optimal policies and the value function. A common method to derive such properties is by value iteration applied to the uniformised MDP. However, due to the unboundedness of the rates, uniformisation is not possible, and so value iteration cannot be applied in the way we need. To circumvent this, one can perturb the MDP. Then we need two results for the perturbed sequence of MDPs: 1. there exists a unique solution to the discounted cost optimality equation for each perturbation as well as for the original MDP; 2. if the perturbed sequence of MDPs converges in a suitable manner then the associated optimal policies and the value function should converge as well. We can model both the MDP and perturbed MDPs as a collection of parametrised Markov processes. Then both of the results above are essentially implied by certain continuity properties of the process as a function of the parameter. In this paper we deduce tight verifiable conditions that imply the necessary continuity properties. The most important of these conditions are drift conditions that are strongly related to nonexplosiveness.

Keywords: Parametrised Markov processes; nonexplosiveness; discounted cost; drift conditions; perturbed MDPs

2010 Mathematics Subject Classification: Primary 90C40
Secondary 93E20; 60J27

1. Introduction

In this paper we study convergence and continuity properties of a collection of parametrised continuous-time Markov processes in countable state space with a discounted cost criterion. The parameter may represent a stationary or deterministic policy in a Markov decision process (MDP). It may also represent a perturbation of a Markov process. Or it can be a combination of both, i.e. control in a perturbed MDP.

The motivation for this paper is our interest in MDPs with unbounded transition rates. In order to study structural properties, the MDP has to be uniformisable. Structural properties of optimal policies and the value function follow from the propagation of these properties through a value iteration step. Note that often value iteration is applicable to the associated jump MDP. However, it is not clear that the desired structural properties propagate through the value iteration step in this case, since the expected sojourn times in the states may not be

Received 1 July 2014; revision received 3 November 2014.
[*] Postal address: Mathematisch Instituut, Leiden University, Postbus 9512, 2300 RA Leiden, The Netherlands.
[*] Email address: blokh1@math.leidenuniv.nl
[**] Email address: spieksma@math.leidenuniv.nl

equal and so they may affect the resulting immediate costs and transition probabilities in an undesirable manner.

Hence, we wish to perturb the MDP in such a manner that it allows uniformisation and the structural properties are preserved. Therefore, continuity in the parameter is necessary to infer properties of the original MDP from properties of the perturbed MDPs.

The conditions we impose on the Markov processes reduce down to the existence of a transformation of the process such that the transformed process is nonexplosive and, moreover, has a bounded cost function. These conditions should hold uniformly in the parameter and are expressed as drift conditions for the original Markov process as well as for the transformed process. Nonexplosiveness of the transformation guarantees continuity of the relevant performance measures as a function of the parameter, provided some standard continuity conditions hold.

The typical performance measure we have in mind is the discounted value function. If the parameter space has a product property then the parametrised process is an MDP. The continuity of the value function implies the existence of a solution of the discounted cost optimality equation (DCOE). We show that the solution provides a deterministic stationary optimal policy in the class of stationary policies. We do not study history dependent policies.

As an illustration we apply our results to the server farm with unbounded rates studied in [1]. In that paper it was shown that for bounded jump perturbations of the model a switching curve policy is optimal. However, the unbounded jump case remained open, since till recently no theory was available to justify taking the limit of the perturbation parameter going to 0—and the jumps becoming unbounded. In this paper we take the parameter space to be the product of the perturbation and control parameter. The obtained continuity results allow us to take the limit and show that a switching curve policy with the same structure is optimal.

The drift conditions that are used to show the existence of a solution of the DCOE are related to the conditions used in [4]–[7], [10], and [12]. These papers do not study convergence results, to the best of the authors' knowledge the only paper where convergence of perturbed MDPs is studied is [11]. We want to emphasise that our aim has been to give minimal conditions for the drift criteria. In the one-parameter case our drift conditions are proven to be necessary (cf. [17]). Furthermore, we have tried to highlight the role that the various conditions play in the derivations. The conditions we impose are weaker than those used in the above mentioned papers. A more detailed comparison with the other drift conditions is given later in the paper, in Section 4 and Remark 5.2.

The paper is organised as follows. In Section 2 we introduce a so-called V-transformation and provide a characterisation of nonexplosiveness in terms of drift conditions. In Section 3 we develop conditions implying the continuity properties we will need. In Section 4 the two main theorems regarding the solutions to the Poisson equation and optimality equation are stated. In Section 5 the translation to MDPs and perturbed MDPs is made. We provide an outline of the approach in order to obtain results for unbounded MDPs. Finally, in Section 6 we demonstrate this approach on the webfarm model studied by [1].

2. Basic settings

We will restrict our investigations to the following class of parametrised processes.

Assumption 2.1. *For each $a \in A$, $X(a)$ is a minimal, standard, stable Markov process, with right-continuous sample paths (with respect to the discrete topology), and with conservative q-matrix $Q(a) = (q_{xy}(a))_{x,y\in S}$, i.e. for all $x \in \boldsymbol{S}$, $a \in A$,*

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

(i) $0 \le q_x(a) = -q_{x,x}(a) < \infty$;

(ii) $\sum_y q_{xy}(a) = 0$.

With $P_t(a) = \{ p_{t,xy}(a)\}_{xy}$, $t \ge 0$, we denote the minimal transition function. A basic role in the discussion of relevant continuity properties of a parametrised Markov process is played by explosiveness properties. To this end we will first review the definition of explosiveness and a characterisation that is useful in this context. For the rest of this section we restrict ourselves to the one-parameter case.

We will define this properly. To this end, let X be a Markov process on $\boldsymbol{S}$ that satisfies Assumption 2.1 (for a parameter space consisting of one element). Let $\tau_0 = 0$ and $\tau_{n+1} = \inf\{t > \tau_n \mid X_t \neq X_{t-}\}$ if X_{τ_n} is nonabsorbing. Otherwise, put $\tau_k = \infty$ and $X_{\tau_k} = X_{\tau_n}$ for $k > n$. Let $J_\infty = \lim_{n\to\infty} \tau_k$. The Markov process X is said to be explosive if there exists a state $x \in \boldsymbol{S}$ with $\mathbb{P}\{J_\infty < \infty \mid X_0 = x\} > 0$. Nonexplosiveness is strongly related to the existence of a drift moment function, introduced below. First we need some notation. Let $f\colon \boldsymbol{S} \to \mathbb{R}$, then f can be viewed as a vector of dimension $|\boldsymbol{S}|$. By Qf and $P_t f$ we mean the matrix multiplied by vector products with elements $Qf(x) = \sum_{y\in S} q_{xy} f(y)$ and $P_t f(x) = \sum_{y\in S} p_{t,xy} f(y)$, $x \in \boldsymbol{S}$, respectively.

Definition 2.1. Let $\gamma \in \mathbb{R}$ and $V\colon \boldsymbol{S} \to \mathbb{R}_+ = (0, \infty)$, then

- V is said to be a γ-drift function for X if $QV \le \gamma V$, where we use componentwise ordering;
- V is said to be a moment function, if there exists an increasing sequence $\{K_n\}_n \subset \boldsymbol{S}$ of finite sets with $\lim_n K_n = \boldsymbol{S}$ such that $\inf_{x\notin K_n} V(x) \to \infty$ as $n \to \infty$.

Note that since Q is conservative, $V \equiv 1$ is always a 0-drift function. Furthermore, from [17, Theorem 2.1] we see that nonexplosiveness of X is equivalent to the existence of a γ-drift moment function for some constant $\gamma \in \mathbb{R}$.

Definition 2.2. Let $\gamma \in \mathbb{R}$ and V be a γ-drift function for X. Define the following associated transformation of X, denoted as X^V. Extend the state space with a coffin state Δ, i.e. $\boldsymbol{S}_\Delta = \boldsymbol{S} \cup \{\Delta\}$. Then define

$$q^V_{xy} = \begin{cases} \dfrac{q_{xy} V(y)}{V(x)}, & x \neq y, x, y \neq \Delta, \\ q_{xx} - \gamma, & x = y, x, y \neq \Delta, \\ \gamma - \dfrac{\sum_{y\in S} q_{xy} V(y)}{V(x)}, & x \neq \Delta, y = \Delta, \\ 0, & x = \Delta, y \in \boldsymbol{S}_\Delta. \end{cases}$$

This makes $Q^V = (q^V_{xy})_{x,y\in S_\Delta}$ a conservative q-matrix, with Δ an absorbing state. Denote by $\{ P^V_t\}_t$ again the (minimum) transition function on the enlarged state space $\boldsymbol{S}_\Delta$.

Since we also need to take into account a cost or reward structure, the validity of the Kolmogorov forward integral equation is an important tool in guaranteeing the existence of solutions to DCOEs. The function $f\colon \boldsymbol{S} \to \mathbb{R}$ is said to satisfy the Kolmogorov forward equation if, for all $x \in \boldsymbol{S}$,

$$P_t f(x) = f(x) + \int_0^t P_s(Qf)(x)\,\mathrm{d}s, \qquad t \ge 0, \tag{2.1}$$

where $P_t f(x) = \sum_y p_{t,xy} f(y)$.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

The following result holds.

Theorem 2.1. (cf. [16, Theorem 3.2] and [17, Theorem 2.1].) *Suppose that Assumption 2.1 holds, and let V be a γ-drift function for X. The following are equivalent:*

(i) *V satisfies (2.1);*

(ii) *X^V is nonexplosive;*

(iii) *for some constant θ there exists a θ-drift V-moment function W for X.*

With W being a V-moment function we mean that W/V is a moment function. Then direct calculations yield that W being a θ-drift V-moment function for X is equivalent to W/V being a $(\theta-\gamma)$-drift moment function for X^V, where $(W/V)(x) = W(x)/V(x)$, $x \in \boldsymbol{S}$.

Under any of these three conditions, the functions bounded by V also satisfy (2.1), under suitable integrability conditions. A discounted version is needed later on, and so we make it precise in the theorem below. To do so, we need some further notation.

The Banach space of functions bounded by V (or V-bounded functions) on $\boldsymbol{S}$ is denoted by $\ell^\infty(\boldsymbol{S}, V)$. This means that $f \in \ell^\infty(\boldsymbol{S}, V)$ if $f\colon \boldsymbol{S} \to \mathbb{R}$ and

$$\|f\|_V := \sup_{x\in S} \frac{|f(x)|}{V(x)} < \infty.$$

If V is a γ-drift function then [2] implies that $P_t V \le \mathrm{e}^{\gamma t} V$ and [17] implies that $t \mapsto P_t V$ is continuous on $\mathbb{R}_+$. This implies that $t \mapsto P_t f$ is continuous for each $f \in \ell^\infty(\boldsymbol{S}, V)$ and, hence, integrable. Additionally, P_t is a V-bounded linear operator, mapping $\ell^\infty(\boldsymbol{S}, V)$ into itself, with induced norm

$$\|P_t\|_V = \sup_x \frac{P_t V(x)}{V(x)} \le \mathrm{e}^{\gamma t}. \tag{2.2}$$

Note that in general the q-matrix Q is not a V-bounded linear operator.

Theorem 2.2. (cf. [16, Theorem 3.4 and Lemma 3.1].) *Let Assumption 2.1 hold, and let V be a γ-drift function for X.*

(i) *If X^V is nonexplosive and, moreover, either $f \in \ell^\infty(\boldsymbol{S}, V)$ and $\int_0^t P_s|Qf|\,\mathrm{d}s < \infty$ or $f = V$ then for any $k \in \mathbb{R}$, f satisfies*

$$\mathrm{e}^{kt} P_t f(x) = f(x) + \int_0^t \mathrm{e}^{ks}[P_s(Qf)(x) + k P_s f(x)]\,\mathrm{d}s. \tag{2.3}$$

(ii) *Conversely, if V satisfies (2.3) for some $k \in \mathbb{R}$, then X^V is nonexplosive.*

Proof. The proof of Theorem 2.2(i) follows entirely from the proofs in the referenced theorem and lemma. The conditions in the referenced results are slightly different: f is assumed to be a γ'-drift function for some $\gamma' \in \mathbb{R}$. However, this is used only in the proofs to guarantee that $\int_0^t P_s|Qf|(x)\,\mathrm{d}s < \infty$. The latter is assumed explicitly here.

For the proof of Theorem 2.2(ii), we assume that (2.3) holds for V. By virtue of [2, Lemma 5.4.2], we have

$$p_{t,xy} = \frac{V(x)}{V(y)} \mathrm{e}^{\gamma t} p^V_{t,xy}, \qquad x, y \in \boldsymbol{S},\ t \ge 0. \tag{2.4}$$

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Hence, we can write (2.3) as

$$e^{kt}\sum_{y\in S} p_{t,xy}V(y) = V(x) + \int_0^t e^{ks}\left[\sum_{z\in S} p_{s,xz}\sum_{y\in S} q_{zy}V(y) + k\sum_{y\in S} p_{s,xy}V(y)\right]\mathrm{d}s.$$

Substituting (2.4) into the above expression, we have

$$e^{(k+\gamma)t}V(x)\sum_{y\in S} p^V_{t,xy} = V(x)\left(1+\int_0^t e^{(k+\gamma)s}\left[\sum_{z\in S} p^V_{s,xz}\sum_{y\in S}(q^V_{zy}+\delta_{zy}\gamma)+k\sum_{y\in S} p^V_{s,xy}\right]\mathrm{d}s\right).$$

Cancelling the $V(x)$ terms, we can express this as

$$e^{(k+\gamma)t}\sum_{y\in S} p^V_{t,xy} = 1 + \int_0^t e^{(k+\gamma)s}\left[\sum_{z\in S} p^V_{s,xz}\sum_{y\in S} q^V_{zy} + \sum_{y\in S}(k+\gamma)\, p^V_{s,xy}\right]\mathrm{d}s.$$

Now for $y=\Delta$, we directly have by the Kolmogorov forward equation:

$$e^{(k+\gamma)t}\, p^V_{t,x\Delta} = \int_0^t e^{(k+\gamma)s}\left[\, p^V_{s,xz}\sum_{z\in S} q^V_{z\Delta} + (k+\gamma)\, p^V_{s,x\Delta}\right]\mathrm{d}s.$$

Combining these, we obtain

$$\begin{aligned}
e^{(k+\gamma)t}\sum_{y\in S_\Delta} p^V_{t,xy} &= 1 + \int_0^t e^{(k+\gamma)s}\left[\sum_{z\in S_\Delta} p^V_{s,xz}\sum_{y\in S_\Delta} q^V_{zy} + \sum_{y\in S_\Delta}(k+\gamma)\, p^V_{s,xy}\right]\mathrm{d}s\\
&= 1 + \int_0^t e^{(k+\gamma)s}\sum_{y\in S_\Delta}(k+\gamma)\, p^V_{s,xy}\,\mathrm{d}s\\
&\geq 1 + \int_0^t e^{(k+\gamma)s}\sum_{y\in S_\Delta}(k+\gamma)\, p^V_{t,xy}\,\mathrm{d}s\\
&\geq 1 + e^{(k+\gamma)t}\sum_{y\in S_\Delta} p^V_{t,xy} - 1\\
&= e^{(k+\gamma)t}\sum_{y\in S_\Delta} p^V_{t,xy}.
\end{aligned}$$

The second equality is due to Q^V being conservative. The inequality is due to the non-increasingness of $s\mapsto\sum_{y\in S_\Delta} p^V_{s,xy}$ (cf. [2, Proposition 1.1.2(i)]). Since the first and last expressions are equal, the inequality is actually an equality. This yields that $\sum_{y\in S_\Delta} p^V_{s,xy}$ is constant on $(0,t)$. Because $\sum_{y\in S_\Delta} p^V_{s,xy}$ is continuous (cf. [2, Proposition 1.2.6]) it is also constant on $[0,t]$. Hence, $\sum_{y\in S_\Delta} p^V_{s,xy} = 1$ for $0\leq s\leq t$. From [2, Proposition 1.1.2(ii)] it follows that $\sum_{y\in S_\Delta} p^V_{s,xy} = 1$ for all $s\geq 0$. Hence, X^V is nonexplosive.

By virtue of the above theorem for the γ-drift function V, requiring the nonexplosiveness of X^V is necessary and sufficient for (2.3) to hold. Hence, (2.3) cannot hold for V under weaker conditions.

In the next section we develop, in our opinion, satisfactory conditions implying the continuity properties in the parameter set A that we will need.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

3. Continuity for the parametrised processes $X(a)$

In order to address continuity aspects, we have to assume some structure on the parameter set.

Assumption 3.1. *The set A is a locally compact topological space, in other words every point $a \in A$ has a compact neighbourhood.*

In what follows we will assume that the above condition holds.

Definition 3.1. We call $V : \boldsymbol{S} \to \mathbb{R}_+$ a (A, γ)-drift function if V is a γ-drift function for $X(a)$ for each $a \in A$. The notions (A, γ)-drift moment function and (A, θ)-drift V-moment function are defined accordingly. If the parameter space A consists of one element, we will drop the reference to A in the notation.

Recall the construction of the minimal transition function. Define

$$f_{t,xy}^{(n)}(a) = \begin{cases} \delta_{xy} \mathrm{e}^{-q_x(a)t}, & n = 0, \\ f_{t,xy}^{(0)}(a) + \displaystyle\int_0^t \mathrm{e}^{-q_x(a)} \sum_{k \neq x} q_{xk}(a) f_{t-s,ky}^{(n-1)}(a)\, \mathrm{d}s, & n \geq 1. \end{cases}$$

By minimality of $X(a)$ [2, Theorem 2.2.2], we have

$$f_{t,xy}^{(n)}(a) \uparrow \; p_{t,xy}(a), \qquad x, y \in \boldsymbol{S},\ t \geq 0,\ a \in A.$$

The interpretation is that $f_{t,xy}^{(n)}(a)$ is the probability that the process $X(a)$ reaches y within t time units with at most n jumps when starting from state x.

Theorem 3.1. *Suppose that Assumptions 2.1 and 3.1 hold and that*

(i) *$a \mapsto q_{xy}(a)$ is continuous on A for $x, y \in \boldsymbol{S}$;*

(ii) *there exists an (A, γ)-drift function V;*

(iii) *$(a, t) \mapsto P_t(a)V(x)$ is continuous on $A \times [0, \infty)$ for each $x \in \boldsymbol{S}$.*

Then $(a, t) \mapsto p_{t,xy}^V(a)$ continuous on $A \times [0, \infty)$ for each $x, y \in \boldsymbol{S}$. Hence, $(a, t) \mapsto p_{t,xy}(a)$ is continuous on $A \times [0, \infty)$ for each $x, y \in \boldsymbol{S}$.

Proof. Let $f_{t,xy}^{V,(n)}(a)$ be the above probabilities for the V-transformed process $X^V(a)$. Thus,

$$f_{t,xy}^{V,(n)}(a) = \begin{cases} \delta_{xy} \mathrm{e}^{-q_x^V(a)t}, & n = 0, \\ f_{t,xy}^{V,(0)}(a) + \displaystyle\int_0^t \mathrm{e}^{-q_x^V(a)} \sum_{k \neq x} q_{xk}^V(a) f_{t-s,ky}^{V,(n-1)}(a)\, \mathrm{d}s, & n \geq 1. \end{cases}$$

We will inductively show that $(a, t) \mapsto \sum_{y \in K} f_{t,xy}^{V,(n)}(a)$ is continuous for each $n \geq 1$, $x \in \boldsymbol{S}_\Delta$, $K \subset \boldsymbol{S}_\Delta$.

First we will show this statement for $K = \{y\}$. Note that $(a, t) \mapsto f_{t,xy}^{V,(0)}(a) = \mathrm{e}^{-q_x^V(a)t}\delta_{xy}$ is continuous for $x, y \in \boldsymbol{S}_\Delta$.

Assume that $(a, t) \mapsto f_{t,xy}^{V,(n-1)}(a)$ is continuous for each $x, y \in \boldsymbol{S}_\Delta$. Because $f_{t,xy}^{V,(n-1)}(a) \leq 1$, the generalised dominated convergence theorem [14, Proposition 11.18] implies that $(a, t) \mapsto \sum_{k \neq x} q_{xk}^V(a) f_{t,ky}^{V,(n-1)}(a)$ is continuous for each $x, y \in \boldsymbol{S}_\Delta$. For each $(a, t) \in A \times [0, \infty)$ this expression is bounded by $q_x^V(a)$. Applying the generalised dominated convergence theorem

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

once more yields that the integral $\int_0^t \mathrm{e}^{-q_x^V(a)} \sum_{k\neq x} q_{xk}^V(a) f_{t-s,ky}^{V,(n-1)}(a)\,\mathrm{d}s$ is a continuous function of $(a,t) \in A \times [0,\infty)$. This gives continuity of $(a,t) \mapsto f_{t,xy}^{V,(n)}(a)$ for $x, y \in \boldsymbol{S}_\Delta$.

An analogous argument shows continuity of $(a,t) \mapsto \sum_{y\in K} f_{t,xy}^{V,(n)}(a)$ for any subset $K \subset \boldsymbol{S}_\Delta$, $x \in \boldsymbol{S}_\Delta$. By virtue of (2.4), Theorem 3.1(iii) is equivalent to requiring continuity of $(a,t) \mapsto \sum_{y\in S} p_{t,xy}^V(a)$.

Next let $x, y \in \boldsymbol{S}$. We wish to show that $(a,t) \mapsto p_{t,xy}^V(a)$ is continuous at some arbitrary point $(a_0,t_0) \in A \times [0,\infty)$. Let $B_0 \subset A \times [0,\infty)$ be a compact neighbourhood of (a_0,t_0). Hence, $(a,t) \mapsto \sum_{y\in S} f_{t,xy}^{V,(n)}(a)$ is a nondecreasing sequence of continuous functions on a compact set, converging to the (assumed) continuous function $(a,t) \mapsto \sum_{y\in S} p_{t,xy}^V(a)$. By Dini's theorem on uniform convergence [15, Theorem 7.13], the convergence is uniform. In other words, for each $\varepsilon > 0$, there exists N_ε such that

$$\varepsilon \geq \left| \sum_{y\in S} p_{t,xy}^V(a) - \sum_{y\in S} f_{t,xy}^{V,(n)}(a) \right| = \sum_{y\in S} (p_{t,xy}^V(a) - f_{t,xy}^{V,(n)}(a)), \qquad (a,t) \in B_0,\ n \geq N_\varepsilon.$$

As a consequence, $f_{t,xy}^{V,(n)}(a)$ converges uniformly in $(a,t) \in B_0$ to $p_{t,xy}^V(a)$ for $x, y \in \boldsymbol{S}$, $t \geq 0$.

By virtue of the uniform limit theorem (cf. [9, Theorem 21.6, p. 132] and [14, Exercise 2.42]) $(a,t) \mapsto p_{t,xy}^V(a)$ is continuous in (a_0,t_0) for $x, y \in \boldsymbol{S}$. Continuity of $(a,t) \mapsto p_{t,xy}(a)$ then follows by (2.4).

Corollary 3.1. *Suppose that Assumptions 2.1 and 3.1 hold and that $X(a)$ is nonexplosive for all $a \in A$. Furthermore, assume that $a \mapsto q_{xy}(a)$ is continuous for each $x, y \in \boldsymbol{S}$. Then $(a,t) \mapsto p_{t,xy}(a)$ is continuous for $x, y \in \boldsymbol{S}$.*

Proof. It holds that $V \equiv 1$ is always a 0-drift function. Furthermore, $P_t(a)V(x) = 1$, $x \in \boldsymbol{S}$; hence, $(a,t) \mapsto P_t(a)V$ is continuous on $\boldsymbol{S} \times [0,\infty)$. The result follows from the previous theorem.

Clearly, Theorem 3.1(iii) is not easily verified for general drift functions. The next theorem provides verifiable conditions in order for the conditions of Theorem 3.1 to hold.

Simultaneously with the preparation of this work, this question has been addressed in [12, Proposition 2.20]. Due to the equivalence result Theorem 2.1, the result of [12] is close to ours. The book [12] restricts the problem to a product set parameter space, and requires compactness of the parameter space. We will provide an alternative proof. The conditions required are the existence of a γ-drift function V and θ-drift V-moment function W, uniform in the parameter $a \in A$.

Assumption 3.2. (i) *It holds that $a \mapsto q_{xy}(a)$ is continuous on A for $x, y \in \boldsymbol{S}$.*

(ii) *There exists an (A,γ)-drift function V.*

(iii) *There exists an (A,θ)-drift V-moment function W.*

Theorem 3.2. *Suppose that Assumptions 2.1, 3.1, and 3.2 hold. Then for each $x \in \boldsymbol{S}$, $(a,t) \mapsto P_t(a)V(x)$ is continuous on $A \times [0,\infty)$ and $a \mapsto Q(a)V(x)$ is continuous on A.*

Proof. Denote by $P_t^V(a)$, $t \geq 0$, the transition function of $X^V(a)$. Since $X^V(a)$ is nonexplosive by virtue of Theorem 2.1,

$$\sum_{y\in S_\Delta} p_{t,xy}^V(a) = 1 \quad \text{for all } (a,t) \in A \times [0,\infty),\ x \in \boldsymbol{S}_\Delta.$$

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Moreover, Corollary 3.1 yields that $(a, t) \mapsto p^V_{t,xy}(a)$ is continuous for $x, y \in \boldsymbol{S}_\Delta$. Combining this gives the continuity of

$$(a, t) \mapsto \sum_{y \in S} p^V_{t,xy}(a) \tag{3.1}$$

on $A \times [0, \infty)$ for $x \in \boldsymbol{S}$. Substituting (2.4) into (3.1) yields that $a \mapsto \sum_{y \in S} p_{t,xy}(a)V(y)$ is continuous for each $x \in \boldsymbol{S}$.

The only thing left to prove is the continuity of $a \mapsto Q(a)V(x)$. To this end we use a nice argument from [12, Proposition 2.20]. Let $x \in \boldsymbol{S}$ be given. Let $\{K_n\}_n \subset \boldsymbol{S}$ be an increasing sequence of finite sets with $x \in K_n$ for all n, $\lim_n K_n = \boldsymbol{S}$, and $\inf_{y \notin K_n} W(y)/V(y) \to \infty$ as $n \to \infty$. Then, for all $a \in A$,

$$\begin{aligned}\sum_{y \notin K_n} q_{xy}(a)V(y) &= \sum_{y \notin K_n} q_{xy}(a)W(y)\frac{V(y)}{W(y)} \\ &\le \frac{1}{\inf_{z \notin K_n} W(z)/V(z)} \sum_{y \notin K_n} q_{xy}(a)W(y) \\ &\le \frac{1}{\inf_{z \notin K_n} W(z)/V(z)}(\theta + q_x(a))W(x).\end{aligned}$$

Let $a_0 \in A$. We wish to show that $a \mapsto \sum_y q_{xy}(a)V(y)$ is continuous in a_0. Let A_0 be a compact neighbourhood of a_0. Then $b := \sup_{a \in A_0}(\theta + q_x(a))W(x) < \infty$. For any $\varepsilon > 0$ there exists N_ε such that

$$\frac{b}{\inf_{z \notin K_n} W(z)/V(z)} \le \varepsilon, \qquad n \ge N_\varepsilon.$$

It follows that $\sum_{y \in K_n} q_{xy}(a)V(y)$ converges to $Q(a)V(x)$ uniformly in $a \in A_0$. Since $a \mapsto \sum_{y \in K_n} q_{xy}(a)V(y)$ is continuous by assumption, we may apply the uniform limit theorem (cf. [9, Theorem 21.6, p. 132]) to obtain $a \mapsto Q(a)V(x)$ is continuous.

Theorem 3.2 connects the continuity properties of the integral $(a, t) \mapsto P_t(a)f(x)$ for $f \in \ell^\infty(\boldsymbol{S}, V)$, to the continuity of the measures of compact sets and nonexplosiveness properties of $X(a)$. The next example illustrates that if Assumptions 3.2(i) and 3.2(ii) hold but $X^V(a)$ is explosive for some $a \in A$, then $a \mapsto P_t(a)V(x)$ need not be continuous on A. This is the basic example from [16, Section 4].

Example 3.1. Let $\boldsymbol{S} = \mathbb{Z}_+$. Consider the q-matrix Q given by

$$q_{xy} = \begin{cases} p2^x, & y = x + 1,\ x \ne 0, \\ -2^x, & y = x,\ x \ne 0, \\ (1 - p)2^x, & y = x - 1,\ x \ne 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $p < \frac{1}{2}$ and 0 is an absorbing state. This is the q-matrix of a nonexplosive Markov process.

Let $V(x) = \alpha^x$ for $\alpha = (1 - p)/p$. Then $QV = 0 \le 0 \cdot V$. The q-matrix Q^V of the associated V-transformation, however, defines an explosive Markov process X (cf. [16]).

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

We define the following parametrised collection of Markov processes. Let $A = \{1, 2, \ldots, \infty\}$. This is a compact set. Let $X(0)$ be the Markov process with q-matrix $Q(\infty) = Q$. For each $a \in A$, we define the perturbation $X(a)$ with q-matrix $Q(a)$ given by

$$q_{xy}(a) = \begin{cases} q_{xy}, & x \le a, \\ -2^a, & y = x > a, \\ 2^a, & y = x - 1,\ x > a. \end{cases}$$

Then $Q(a)V \le 0 \cdot V$ for every $a \in A$. Also $a \mapsto q_{xy}(a)$ is trivially continuous on A. Hence, Assumptions 3.2(i) and 3.2(ii) are satisfied. Note that due to the boundedness of jumps, $X^V(a)$, $a < \infty$, is nonexplosive.

Since $X^V = X^V(\infty)$ is explosive, there exists a state $x \in \boldsymbol{S}_\Delta$ such that

$$1 = \lim_{a\to\infty} \sum_y p^V_{t,x,y}(a) > \sum_y p^V_{t,x,y}(\infty).$$

By virtue of (2.4), $\sum_y p_{t,xy}(a)V(y) \not\to \sum_y p_{t,xy}(\infty)V(y)$ as $a \to \infty$ for $t > 0$. Hence, $a \mapsto P_t(a)V(x)$ is not continuous on A.

4. The Poisson equation and optimality equation for the α-discounted cost criterion

Suppose next that Assumptions 2.1 and 3.2(ii) hold, in other words there exists a γ-drift function V. Assume that a cost $c_x(a)$ per unit time is incurred when the process $X(a)$ resides in state x under parameter $a \in A$. Denote by $c(a) = (c_x(a))_{x\in S}$ the associated cost vector.

Assumption 4.1. (i) *$a \mapsto c_x(a)$ is continuous on A;*

(ii) *there is a finite constant c_V such that* $\sup_{x,a} |c_x(a)|/V(x) \le c_V$*;*

(iii) *for the discount factor α it holds that $\alpha > \gamma$.*

Define the expected α-discount total cost associated with parameter $a \in A$ by

$$v^\alpha(a) = \int_0^\infty \mathrm{e}^{-\alpha t}\, P_t(a)c(a)\,\mathrm{d}t,$$

and the xth component by $v^\alpha(x, a)$. Note that Assumptions 2.1, 3.2(ii), and 4.1(ii) imply that $t \mapsto P_t(a)V$ is continuous. By (2.4), $P_tV(x) \le \mathrm{e}^{\gamma t}V(x)$. Hence, $\alpha > \gamma$ guarantees that $v^\alpha(a)$ is well-defined and finite.

If we further require the nonexplosiveness of X^V then it can be shown that $v^\alpha(a)$ is the unique solution of the Poisson equation (4.1) in $\ell^\infty(\boldsymbol{S}, V)$.

Theorem 4.1. *Let Assumption 2.1 hold.*

(i) *If Assumptions 3.2(ii) and 3.2(iii), and Assumptions 4.1(ii) and 4.1(iii) hold, then $v^\alpha(a)$ is the unique solution in $\ell^\infty(\boldsymbol{S}, V)$ to the α-discounted equation*

$$\alpha f = c(a) + Q(a) f. \tag{4.1}$$

(ii) *If, additionally, Assumptions 3.1, 3.2(i), and 4.1(i) hold, then $a \mapsto v^\alpha(a)$ is componentwise continuous on A.*

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Proof. Let $a \in A$. We first prove that $v^\alpha(a)$ is a solution to (4.1) in the space $\ell^\infty(\boldsymbol{S}, V)$. Note that $\|v^\alpha(a)\|_V \le c_v/(\alpha - \gamma)$, so that $v^\alpha(a) \in \ell^\infty(\boldsymbol{S}, V)$. Moreover, $Q(a)|v^\alpha(a)|$ is well defined and finite. We obtain

$$\begin{aligned} Q(a)v^\alpha(x,a) &= \sum_y q_{xy}(a) \int_0^\infty \mathrm{e}^{-\alpha t} \sum_z p_{t,yz}(a)\, c_z(a)\, \mathrm{d}t \\ &= \sum_y q_{xy}(a) \sum_z \int_0^\infty \mathrm{e}^{-\alpha t}\, p_{t,yz}(a)\, \mathrm{d}t\, c_z(a) \\ &= \sum_z \int_0^\infty \mathrm{e}^{-\alpha t}\, p'_{t,xz}(a)\, \mathrm{d}t\, c_z(a) \\ &= \sum_z \left(-\delta_{xz} + \alpha \int_0^\infty \mathrm{e}^{-\alpha t}\, p_{t,xz}(a)\, \mathrm{d}t \right) c_z(a) \\ &= -c_x(a) + \alpha v^\alpha(x,a). \end{aligned}$$

The interchange of summation and integration in the second equality is justified by Fubini's theorem; in the third equality, by the additional fact that $Q(a)$ has at most one negative element per row. The fourth equality is due to partial integration. As a consequence, $v^\alpha(a)$ is a solution of (4.1) in $\ell^\infty(\boldsymbol{S}, V)$.

Suppose that $f \in \ell^\infty(\boldsymbol{S}, V)$ is another solution. Then $\alpha(v^\alpha(a) - f) = Q(a)(v^\alpha(a) - f)$, and so $v^\alpha(a) - f \in \ell^\infty(\boldsymbol{S}, V)$ is an eigenvector of $Q(a)$ to eigenvalue $\alpha > 0$. Direct calculations show that $g\colon \boldsymbol{S}_\Delta \to \mathbb{R}$, given by $g = (v^\alpha(a) - f)/V$ on $\boldsymbol{S}$ and $g(\Delta) = 0$, is a bounded eigenvector of $Q^V(a)$ to eigenvalue $\alpha - \gamma > 0$. Nonexplosiveness of a Markov process can be characterised by the nonexistence of a bounded (nonzero) eigenvector of the corresponding q-matrix to positive eigenvalues (cf. [13, Theorem 7] and [17, Theorem 2.1]). By Assumption 3.2, $X^V(a)$ is nonexplosive. Hence, a nonzero eigenvector cannot exist and so we conclude that $f = v^\alpha(a)$.

We finally turn to proving the componentwise continuity of $v^\alpha(a)$. By virtue of Theorem 3.2, $a \mapsto P_t(a)V$ is componentwise continuous. Equation (2.2) yields that $P_t(a)V \le \mathrm{e}^{\gamma t}V$. Hence, the dominated convergence theorem implies that $a \mapsto \int_0^\infty \mathrm{e}^{-\alpha t}\, P_t(a)V\, \mathrm{d}t$ is componentwise continuous. Another application of the dominated convergence theorem implies that $a \mapsto v^\alpha(a) = \int_0^\infty \mathrm{e}^{-\alpha t}\, P_t(a)c(a)\, \mathrm{d}t$ is componentwise continuous.

We will next consider the special case that the collection $\{Q(a)\}_{a\in A}$ and $\{c(a)\}_{a\in A}$ have the product property (cf. [8]) in the following sense.

Assumption 4.2. *There exist compact metric sets A_x, $x \in \boldsymbol{S}$, such that the following conditions hold:*

(i) $A = \prod_{x\in S} A_x$, *and A is equipped with the product topology;*

(ii) $\{Q(a)\}_{a\in A}$ *and* $\{c(a)\}_{a\in A}$ *have the product property. In other words, for any $a, a' \in A$, $x \in \boldsymbol{S}$ such that $a_x = a'_x$, it holds that $(Q(a))_{x\cdot} = (Q(a'))_{x\cdot}$, and $c_x(a) = c_x(a')$. Here $(Q(a))_{x\cdot}$ stands for the x-row of $Q(a)$.*

Note that A is compact and metrisable, and the product topology is the topology of componentwise convergence. Hence, A is sequentially compact.

Under Assumption 4.2, the xth row and xth component of $Q(a)$ and $c(a)$ depend on the value a_x only. Therefore, with a slight abuse of notation, we may write $q_{xy}(a_x)$ and $c_x(a_x)$.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Then $\inf_{a\in A}(c(a) + Q(a)f)$ is well defined and may also be written as $\inf_{a_x\in A_x}\{c_x(a_x) + \sum_y q_{xy}(a_x)f(y)\}$ for all $x \in \boldsymbol{S}$. As an application, the set A may represent the collection of stationary policies in an MDP, or the set of deterministic policies.

We say that *parameter a^* is optimal in A if $v^\alpha(a^*) \le v^\alpha(a)$ for all $a \in A$*. In this case we have the following result.

Theorem 4.2. *Suppose that Assumption 2.1 holds.*

(i) *Suppose that Assumptions 3.2(ii), 4.1(ii), 4.1(iii), and 4.2 hold. Moreover, suppose that there exists a function m such that $m(x) \ge \sup_a q_x(a)$. Then the equation*

$$\alpha f(x) = \inf_{a_x\in A_x}\left\{c_x(a_x) + \sum_y q_{xy}(a_x)f(y)\right\}, \qquad x \in \boldsymbol{S}, \tag{4.2}$$

has a solution v^α in $\ell^\infty(\boldsymbol{S}, V)$.

(ii) *If, moreover, Assumptions 3.2(i), 3.2(iii), and 4.1(i) hold, this solution is unique in $\ell^\infty(\boldsymbol{S}, V)$ and the infimum is a minimum. For any $a^* = (a_x^*)_x \in A$ for which a_x^* achieves the minimum in (4.2), $x \in \boldsymbol{S}$, we have $v^\alpha(a^*) = v^\alpha$ and a^* is optimal in A.*

Proof of Theorem 4.2(i). We use the same line of reasoning as in the proof of [12, Theorem 3.7]. Suppose that Assumptions 3.2(ii), 4.1(ii), and 4.2 hold. Let $m\colon \boldsymbol{S} \to \mathbb{R}_+$ be such that $m(x) \ge \sup_{a_x\in A_x} q_x(a_x)$ for $x \in \boldsymbol{S}$. Then define $p_{xy}(a_x) = q_{xy}(a_x)/m(x) + \delta_{xy}$ for $x, y \in \boldsymbol{S}, a_x \in A_x$, which is a probability measure for each state action pair (x, a_x). Furthermore, define the operator T for $f \in \ell^\infty(\boldsymbol{S}, V)$ by

$$(Tf)(x) = \inf_{a_x\in A_x}\left\{\frac{c_x(a_x)}{\alpha + m(x)} + \frac{m(x)}{\alpha + m(x)}\sum_{y\in S} p_{xy}(a_x)f(y)\right\}, \qquad x \in \boldsymbol{S}.$$

Define the sequence $\{f_n\}_n$ in $\ell^\infty(\boldsymbol{S}, V)$ by $f_0(x) = (c_V/(\alpha - \gamma))V(x)$, and $f_n = Tf_{n-1}$ for $n \ge 1$. First, nonnegativity of the coefficients in the second term between brackets implies that T is monotone (i.e. $f \ge g$ implies that $Tf \ge Tg$). Secondly, direct calculations show that $f_0 \ge f_1$. This implies that $\{f_n\}_n$ is a monotone decreasing sequence. Furthermore, it is easy to show that

$$\|f_n\|_V \le \frac{c_V}{\alpha - \gamma}.$$

Thus, $\{f_n\}_n$ has a pointwise limit $f^* \in \ell^\infty(\boldsymbol{S}, V)$ with $f^* \le f_n$ for all n. Hence, $Tf^* \le Tf_n = f_{n+1}$ for all n and, thus, $Tf^* \le \lim_{n\to\infty} f_n = f^*$.

Next we prove that $f^* \le Tf^*$. First note that

$$f^* \le f_{n+1} = Tf_n, \qquad n = 1, \ldots. \tag{4.3}$$

For notational convenience, denote

$$(T_{a_x} f)(x) = \frac{c_x(a_x)}{\alpha + m(x)} + \frac{m(x)}{\alpha + m(x)}\sum_{y\in S} p_{xy}(a_x)f(y), \qquad x \in \boldsymbol{S},$$

so that $Tf(x) = \inf_{a_x} T_{a_x} f(x)$. By monotone convergence $T_{a_x} f_n(x) \downarrow T_{a_x} f^*(x)$, $n \to \infty$, $a_x \in A_x$. Let $\varepsilon > 0$, $x \in \boldsymbol{S}$, and $a_x \in A_x$. Then there exists N_{ε,x,a_x} such that

$$T_{a_x} f_n(x) \le T_{a_x} f^*(x) + \varepsilon, \qquad n \ge N_{\varepsilon,x,a_x}. \tag{4.4}$$

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Combining (4.4) with (4.3) yields

$$f^*(x) \le T_{a_x} f^*(x) + \varepsilon, \qquad a_x \in A_x.$$

Taking the infimum on both sides, we obtain

$$f^*(x) \le Tf^*(x) + \varepsilon.$$

Since $\varepsilon > 0$ and $x \in \boldsymbol{S}$ were arbitrary, we obtain the desired inequality $f^* \le Tf^*$. We conclude that $Tf^* = f^*$.

By direct calculations it is seen that this last equality is equivalent to (4.2); thus, we have proven that there is a solution and we call this v^α.

Proof of Theorem 4.2(ii). Suppose now that Assumptions 3.2(i), 3.2(iii), and 4.1(i) hold as well. By Theorem 4.1, $a \mapsto c(a) + Q(a)v^\alpha$ is componentwise continuous on A. Since A is compact, this implies that the infimum is attained. So there is an $a^* \in A$ such that

$$\begin{aligned}\alpha v^\alpha(x) &= \inf_{a_x \in A_x} \left\{ c_x(a_x) + \sum_y q_{xy}(a_x) v^\alpha(y) \right\} \\ &= \min_{a_x \in A_x} \left\{ c_x(a_x) + \sum_y q_{xy}(a_x) v^\alpha(y) \right\} \\ &= c_x(a_x^*) + \sum_y q_{xy}(a_x^*) v^\alpha(y).\end{aligned}$$

Then $v^\alpha = v^\alpha(a^*)$ by Theorem 4.1. Next we will show that $v^\alpha = v^\alpha(a^*) \le v^\alpha(a)$ for any $a \in A$, in other words a^* is optimal in A. To this end, let $\hat{a} \in A$. Enumerate $\boldsymbol{S} = \{s_1, s_2, \ldots\}$. Define $a^n \in A$ by

$$a_x^n = \begin{cases} \hat{a}_x, & x \in \{s_1, \ldots, s_n\}, \\ a_x^*, & x \in \{s_{n+1}, \ldots\}. \end{cases}$$

Then $a^n \to \hat{a}$, $n \to \infty$, in the product topology and, in particular,

$$\alpha v^\alpha \le c(a^n) + Q(a^n)v^\alpha.$$

Define

$$d^n = c(a^n) + Q(a^n)v^\alpha - \alpha v^\alpha,$$

then d^n has at most n nonzero components and so $d^n \in \ell^\infty(\boldsymbol{S}, V)$. It follows that $|Q(a^n)v^\alpha| \in \ell^\infty(\boldsymbol{S}, V)$. Hence, $t \mapsto P_t(a^n)(Q(a^n)v^\alpha)$ is finite and continuous, and

$$\alpha P_t(a^n)v^\alpha \le P_t(a^n)c(a^n) + P_t(a^n)(Q(a^n)v^\alpha).$$

Multiplying both sides by $\mathrm{e}^{-\alpha t}$, integrating over $(0, T)$, and rearranging terms, we obtain, for any $T > 0$,

$$\int_0^T \mathrm{e}^{-\alpha t}[\alpha P_t(a^n)v^\alpha - P_t(a^n)(Q(a^n)v^\alpha)]\,\mathrm{d}t \le \int_0^T \mathrm{e}^{-\alpha t} P_t(a^n)c(a^n)\,\mathrm{d}t.$$

By virtue of Theorem 2.2, (2.3) is applicable with $k = -\alpha$, thus yielding

$$v^\alpha - \mathrm{e}^{-\alpha T} P_T(a^n)v^\alpha \le \int_0^T \mathrm{e}^{-\alpha t} P_t(a^n)c(a^n)\,\mathrm{d}t, \qquad T > 0.$$

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Note that $\| P_T(a^n)v^\alpha \|_V \leq e^{\gamma T}\|v^\alpha\|_V \leq e^{\gamma T}c_V/(\alpha-\gamma)$. Taking the limit $T \to \infty$, we obtain the desired result that $v^\alpha \leq v^\alpha(a^n)$. Since $a \mapsto v^\alpha(a)$ is componentwise continuous, we can finally take the limit $n \to \infty$ and obtain $v^\alpha \leq v^\alpha(\hat{a})$. Uniqueness now follows immediately.

Remark 4.1. The question arises whether an optimal policy in A is optimal in the class of *Markov policies*, as defined in [12], or even in more general classes of policies. Note that a Markov policy generates a nonhomogeneous Markov process. Following the proof that a solution to the α-discount optimality equation dominates the expected α-discounted cost under a Markov policy in [12, Lemma 3.5], one needs the result of Theorem 2.2 to hold for a nonhomogeneous Markov process. To the best of the authors' knowledge such a result has not yet been formally proved.

Discussion on related conditions in the literature. In [4], [7], and [12] the parametrised process $X(a)$ as well as $X^V(a)$ are supposed to be nonexplosive for all $a \in A$. We require only $X^V(a)$ to be nonexplosive uniformly in A. This relaxation might be useful if the cost function goes to 0 'fast enough' as the state grows large. See the example below, it is a variation on Example 3.1. In [10], $X(a)$ is not required to be nonexplosive either; however, the extra condition that $q_x(a)V(x) \leq W(x)$ for $x \in \boldsymbol{S}$, $a \in A$, is required there. In [17] a detailed discussion on the relation between the various drift conditions used in this context is presented.

Example 4.1. Let $\boldsymbol{S} = \mathbb{Z}_+$. Define the following q-matrices $Q(a)$ by

$$q_{xy}(a) = \begin{cases} a_x 2^x, & y = x+1,\ x \neq 0, \\ -2^x, & y = x,\ x \neq 0, \\ (1-a_x)2^x, & y = x-1,\ x \neq 0, \\ 0 & \text{otherwise}, \end{cases}$$

for any $a_x \in A_x = [p_0, p_1]$ with $\frac{1}{2} < p_0 \leq p_1 \leq 1$. Hence, $A = \prod_{x\in S} A_x$ is a compact product set. Note that, clearly, $a \mapsto q_{xy}(a)$ is continuous on A. Note also that since $a_x \geq p_0 > \frac{1}{2}$ for all a_x, this is an explosive Markov process for every $a \in A$.

Next define the reward structure $r(a)$ (note that nowhere in the theory above is it essential whether to maximise or minimise). We let the reward rate consist of two parts: a fixed reward rate B for staying in the finite set $\{x \leq U\}$, and a bonus depending on the current state for taking actions that move the system to a higher state with larger probability. Therefore, put $r_x(a) = b_x(a) + c_x(a)$ with $b_x(a) = B\mathbf{1}_{\{x\leq U\}}$ and $c_x(a) = Ca_x(1-\varepsilon)^x$, where $\mathbf{1}$ is the indicator function.

We will make a transformation that makes the transformed process nonexplosive. Take $V(x) = \beta^x$ with $\max((1-p_0)/p_0, 1-\varepsilon) \leq \beta < 1$. Note that since $\beta < 1$, V is not a moment function. Then for all $a \in A$, $x \in \boldsymbol{S}$,

$$\begin{aligned} Q(a)V(x) &= (a_x\beta^{x+1} - \beta^x + (1-a_x)\beta^{x-1})2^x \\ &= (a_x\beta^2 - \beta + (1-a_x))\beta^{x-1}2^x \\ &= \left(a_x\left(\beta - \frac{1-a_x}{a_x}\right)(\beta-1)\right)\beta^{x-1}2^x \\ &\leq 0 \cdot V, \end{aligned}$$

where the inequality holds because $(1-a_x)/a_x < \beta \leq 1$. Hence, V is a $(A, 0)$-drift function.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Moreover, $r(a)$ is uniformly V-bounded, since

$$\sup_{x\in\mathbb{Z}_+,a\in A}\frac{|b_x(a)|}{V(x)}=\max_{x\le U}B\beta^{-x}=B\beta^{-U}$$

and

$$\sup_{x\in\mathbb{Z}_+,a\in A}\frac{|c_x(a)|}{V(x)}=\max_{x\in\mathbb{Z}_+}Cp_1\left(\frac{1-\varepsilon}{\beta}\right)^x=Cp_1.$$

Next take $W\equiv 1$, then $QW=0\le 0\cdot W$ and $\lim_{x\to\infty}W(x)/V(x)=\lim_{x\to\infty}\beta^{-x}=\infty$. Hence, W is a $(A,0)$-drift V-moment function. Then Theorem 3.2 yields that the transformed process $X^V(a)$ is nonexplosive for all $a\in A$.

Now all assumptions of Theorem 4.2 hold; hence, (4.2) (with the infimum replaced by a supremum) has a unique solution $v^\alpha\in\ell^\infty(\boldsymbol{S},V)$ for any $\alpha>0$ and there is a parameter $a^*\in A$ that achieves this supremum.

5. MDPs and perturbations

In this section we show how Theorem 4.2 can be applied to MDPs. In order to do so, we take the parameter set $A:=\mathcal{D}=\prod_x\mathcal{D}_x$, where $\mathcal{D}$ is the set of all deterministic (stationary) policies, and $\mathcal{D}_x=\{\text{set of actions available in state } x\}$, $x\in\boldsymbol{S}$. Then $A=\mathcal{D}$ has the product property described in Assumption 4.2. We use the notation $\delta\in\mathcal{D}$ for a deterministic (stationary) policy and by $\delta(x)\in\mathcal{D}_x$ the corresponding action prescribed in state x by δ. If we assume that $\mathcal{D}_x$ is compact, metric for each $x\in\boldsymbol{S}$ then $\mathcal{D}$ is a compact, metric space as well. Consequently, an MDP with compact action space and deterministic policies $\mathcal{D}$ can be identified with a parametrised collection of Markov processes satisfying Assumption 4.2.

Remark 5.1. If Assumptions 2.1, 3.2, 4.1, and 4.2 hold for $A=\mathcal{D}$, it is a standard construction to show that these assumptions apply as well for the parameter set equal to the set $\mathcal{S}$ of stationary, randomised policies. For an example of this construction, see [3]. Hence, the assertion of Theorem 4.2 then also applies for this larger parameter set. Furthermore, it is a simple consequence that if $\mathcal{A}=\mathcal{S}$ in (4.2) then there exists a minimiser $\delta^*\in\mathcal{D}$ for which $v^\alpha(\delta^*)=v^\alpha$. As a consequence, we may (and we will) restrict our analysis to $\mathcal{D}$.

Perturbation of MDPs. In this paragraph we will discuss how Theorems 4.1 and 4.2 can be applied to analyse MDPs by adding a perturbation. The application we have in mind is the analysis of the structural properties of an MDP with unbounded transition rates (i.e. $\sup_{x\in S,\delta\in\mathcal{D}}q_x(\delta)=\infty$) and, thus, the uniformisation technique is not applicable. In particular, we are interested in the structure of optimal strategies and of the value function. To this end we perturb the MDP to obtain bounded rates so that it can be studied using the discrete-time equivalent MDP. This perturbation is indexed by an extra parameter N, typically $N\in\mathcal{N}$, where $\mathcal{N}:=\{1,2,\dots,\infty\}$, a compact set. Thus, we obtain a collection of extended parametrised processes, $\{X(N,\delta)\}_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$. For fixed N the parametrised process $\{X(N,\delta)\}_{\delta\in\mathcal{D}}$ is an MDP and for $N=\infty$ this coincides with the original MDP. The theorems in the previous section provide the framework that guarantees continuity in the perturbation parameter. This induces convergence of the results for the perturbed models to the original model if the perturbation vanishes, i.e. the parameter goes to ∞.

Theorem 5.1. *Consider an MDP, in other words a parametrised collection of processes* $\{X(\delta)\}_{\delta\in\mathcal{D}}$ *with cost function* $c(\delta)_{\delta\in\mathcal{D}}$. *Furthermore, consider an extended parametrised*

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

collection of processes $\{X(N,\delta)\}_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$ *with cost function* $c(N,\delta)_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$ *such that* $X(\infty,\delta)=X(\delta)$ *and* $c(\infty,\delta)=c(\delta)$.

Suppose that Assumptions 2.1, 3.1, 3.2, and 4.1 hold for $\{X(N,\delta)\}_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$. *Suppose that, additionally, Assumption 4.2 holds for* $\{X(N,\delta)\}_{\delta\in\mathcal{D}}$ *for all* $N\in\mathcal{N}$. *Let* v_N^α *be the value function for the MDP* $\{X(N,\delta)\}_{\delta\in\mathcal{D}}$ *and* δ_N^* *an optimal policy,* $N\in\mathcal{N}$. *Then the following hold:*

(i) $\lim_{N\to\infty} v_N^\alpha = v^\alpha$;

(ii) *any limit point of* $(\delta_N^*)_{N\in\mathcal{N}}$ *is optimal for* $\{X(\delta)\}_{\delta\in\mathcal{D}}$.

Proof. The assertions of Theorem 4.2 hold for $\{X(N,\delta)\}_{\delta\in\mathcal{D}}$ for fixed $N\in\mathcal{N}$. This yields the existence of a pair (v_N^α,δ_N^*) satisfying (4.2), so that $v_N^\alpha=v^\alpha(N,\delta_N^*)$ for fixed $N\in\mathcal{N}$.

The sequence $\{v_N^\alpha\}_{N<\infty}$ is a bounded sequence in $\ell^\infty(\boldsymbol{S},V)$. Consider any limit point of this sequence, say it is achieved along the subsequence $\{v_{N_k}^\alpha\}_{k=1,\dots}$. By sequential compactness of $\mathcal{N}\times\mathcal{D}$, we have that $(\delta_{N_k}^*)_k$ has a convergent subsequence that we denote by $(\delta_{N_k}^*)_{N_k\in\mathcal{N}}$, again with limit δ^* say.

Since the assertions of Theorem 4.1 hold for $\{X(N,\delta)\}_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$, this implies that $(N,\delta)\mapsto v^\alpha(N,\delta)$ is continuous on $\mathcal{N}\times\mathcal{D}$. In particular, we have

$$\lim_{k\to\infty} v_{N_k}^\alpha = \lim_{k\to\infty} v^\alpha(N_k,\delta_{N_k}^*) = v^\alpha(\infty,\delta^*) = v_\infty^\alpha(\delta^*).$$

Continuity of the map $(N,\delta)\mapsto v^\alpha(N,\delta)$, the fact that $v^\alpha(N_k,\delta_{N_k}^*)$ solves the optimality equation for the N_k-perturbation by Theorem 4.2, and the continuity result of Theorem 3.2 together imply that $v^\alpha(\delta^*)$ solves the optimality equation for the ∞-perturbation, in other words for the original MDP. Hence, $v^\alpha(\delta^*)=v^\alpha$ and δ^* is optimal. This holds for any limit point of $\{v_N^\alpha\}_N$. Since the solution of the optimality equation is unique, any limit point is equal to v^α and corresponding limit points of $\{\delta_N\}_N$ are optimal. This proves (i). For the proof of (ii), we consider a limit point of the sequence of policies $\{\delta_N\}_N$ (for any sequence of optimal policies for the N-perturbation, $N=1,2,\dots$). Then choose a subsequence along which $\{v_N^\alpha\}_N$ converges and we apply the same argument as in the above.

The approach of extended parametrisation. 1. Start with a parametrised process $\{X(\delta)\}_{\delta\in\mathcal{D}}$, the original MDP. Our interest is in the structural properties of v^α and δ^*. The assumptions of Theorem 4.2 must hold for this parametrised process.

2. Add a perturbation, parametrised by $N\in\mathcal{N}$. In this way we obtain an extended parametrised process $\{X(N,\delta)\}_{(N,\delta)\in\mathcal{N}\times\mathcal{D}}$. The extended parametrised process does not need to satisfy the product property of Assumption 4.2. However, all other assumptions from Theorem 4.2 are assumed to be satisfied for the extended parametrised process.

3. Fixing $N\in\mathcal{N}$, we obtain the parametrised processes $\{X(N,\delta)\}_{\delta\in\mathcal{D}}$, satisfying the product property of Assumption 4.2, and so all assumptions of Theorem 4.2 hold. Hence, there exists a unique solution v_N^α satisfying (4.2) and any maximiser δ_N^* is optimal, $N\in\mathcal{N}$.

4. If the parametrised processes are uniformisable for $N<\infty$, we can determine structural properties of (v_N^α,δ_N^*) for all $N<\infty$ by e.g. value iteration.

5. Now Theorem 5.1 implies that $\lim_{N\to\infty} v_N^\alpha = v^\alpha$ and that any limit point of $(\delta_N^*)_N$ is optimal for the original model. As a conclusion, both the optimal policy and the minimum expected α-discounted cost of the original model can be approximated by the corresponding quantities for the perturbed model for large perturbation parameters.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Remark 5.2. Theorem 5.1 is strongly related to [11, Theorem 3.1]. The paper gives conditions for convergence of finite state MDPs to infinite state processes. However, the drift conditions imposed are more restrictive (cf. [16, Example 5.4]). In particular, the authors impose three extra conditions on the rate matrix. First, that V is a moment function. Secondly, that

$$\sup_{\delta} q_x(N, \delta) \leq V(x) \quad \text{for all } N \in \mathcal{N}.$$

Thirdly, they require a particular V-moment function W, namely $W = V^2$.

The last part of the paper is an illustration of the application of the approach to a server farm model.

6. Optimal control of a server farm

Consider the server farm model studied by [1]. This model has an infinite server pool, implying that the transition rates are not bounded. To derive structural properties of the optimal policy, the authors bound the departure rate. After uniformisation, analysis of the equivalent discrete-time chain shows that a specific switching curve is optimal for the bounded rate model. However, this paper does not give any results on the original unbounded model.

We will demonstrate here that the same structural results apply for the unbounded model by using the approach of extended parametrisation.

The mathematical setup is as follows. There is a Poisson stream of arrivals with rate λ. Each customer requires an exponential service time with parameter μ. There is an infinite server pool, where servers can be in three states. They can be either active (on), turned off (off), or in standby modus (idle). After service completion the controller has two options, either turn the server off, or leave the server idle. A server in the idle state costs c per unit time, due to energy consumption. Upon customer arrival, there are two possibilities.

(i) There is an available idle server. Then a customer is assigned one of these, and the server changes from idle to on.

(ii) There are no idle servers. Then an off-server is turned on, and instantaneous startup costs K have to be paid.

The goal is to minimise the total expected discounted cost over all stationary policies.

We will model this as follows. Let i be the number of idle servers and j the number of busy servers. The state space $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \{(i, j) \mid i,\ j \in \mathbf{Z}_+\}.$$

Possible actions at service completion are either to turn the server off (0) or leave the server idle (1). The action space is

$$\mathcal{D}_{(i,j)} = \{0, 1\} \quad \text{for } (i, j) \in \boldsymbol{S}.$$

Hence, the set stationary deterministic policies is $\mathcal{D} = \{0, 1\}^S$. Then the rate matrix $Q(\delta)$ is

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

TABLE 1: If it is optimal to turn the server off (respectively leave the server idle) in state (i, j) then it is also optimal in the following states.

	leave idle	turn off	structural property of v_N^α
(i)	$\downarrow: (i, j-1)$	$\uparrow: (i, j+1)$	SuperM(1, 2)
(ii)	$\nwarrow: (i-1, j+1)$	$\searrow: (i+1, j-1)$	SuperCx(1)

given by

$$q_{(i,j),(i',j')}(\delta) = \begin{cases} j\mu, & (i', j') = (i, j-1),\ \delta(i, j) = 0, \\ & \text{or } (i', j') = (i+1, j-1),\ \delta(i, j) = 1, \\ \lambda, & (i', j') = (i-1, j+1),\ i > 0, \\ & \text{or } (i', j') = (i, j+1),\ i = 0, \\ -(j\mu + \lambda), & (i', j') = (i, j). \end{cases}$$

The associated cost function $c(\delta)$ is given by

$$c_{(i,j)}(\delta) = ci + \lambda K \mathbf{1}_{\{i=0\}}, \qquad (i, j) \in \boldsymbol{S}.$$

Note that we have remodelled the instantaneous costs as a cost rate. This can be done without loss of generality.

As pointed out in the above, the rates $q_{(i,j)}(\delta) = j\mu + \lambda$ are not uniformly bounded. To analyse this system, [1] assumes that the service rates are a concave, nondecreasing, bounded function $\mu(j)$ of the number of busy servers j and thereby they make it uniformisable.

We will use this to define a suitable perturbation of the model, i.e. a uniformisable MDP, with the service rates a concave, nondecreasing, and bounded function of the number of busy servers. In other words, denoting our original MDP by $\{X(\delta)\}_{\delta \in \mathcal{D}}$, we define a collection of perturbed MDPs $\{X(N, \delta\}_{N \in \mathcal{N}, \delta, \in \mathcal{D}}$ with $\mathcal{N} = \{1, \ldots, \infty\}$. Let the rate matrix $Q(N, \delta)$ be given by

$$q_{(i,j),(i',j')}(N, \delta) = \begin{cases} \min\{j, N\}\mu, & (i', j') = (i, j-1),\ \delta(i, j) = 0, \\ & \text{or } (i', j') = (i+1, j-1),\ \delta(i, j) = 1, \\ \lambda, & (i', j') = (i-1, j+1),\ i > 0, \\ & \text{or } (i', j') = (i, j+1),\ i = 0, \\ -(\min\{j, N\}\mu + \lambda), & (i', j') = (i, j). \end{cases}$$

The cost function remains unchanged.

Note that $X(\infty, \delta)$ coincides with the original unbounded model. On the other hand, for each $N < \infty$, the N-perturbation is uniformisable and satisfies the service rate conditions of [1]. Hence, the structural properties of the value function v_N^α can be derived by value iteration. By virtue of the results in [1], it follows that the optimal policy for the N-perturbation, $N < \infty$, has the switching curve structure shown in Table 1.

With the approach of 'extended parametrisation', we are able to extend this result to the original unbounded model. The only thing remaining is to check that the assumptions of Theorem 5.1 hold. If the conditions hold, by virtue of the theorem we may conclude that a switching curve policy with the structure given in Table 1 is optimal for the original unbounded MDP. This yields the following result.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

Theorem 6.1. *For the webfarm model $\{X(\delta)\}_\delta$ there exists a deterministic policy with the threshold structure described in Table 1, that is α-discount optimal within the class $\mathcal{S}$ of stationary policies.*

Proof. Note that the assumptions are of such nature that if they are satisfied by the extended parametrised process they are also satisfied by the parametrised process. As has been pointed out, we have to verify the assumptions of Theorems 4.1 and 4.2. We will do so in a systematic way.

- It is clear that Assumption 2.1 holds for both the parametrised as the extended parametrised process, since there are no instantaneous jumps and the rate matrix is conservative.
- For Assumption 3.2, there are three properties to check.

  (i) Continuity of $\delta \mapsto q_{(i,j),(i',j')}(N,\delta)$ for fixed $N \in \mathcal{N}$ is clear. Also, we have $\lim_{N\to\infty} q_{(i,j),(i',j')}(N,\delta) = q_{(i,j),(i',j')}(\infty,\delta)$ (for large N these values are equal for any fixed pair of states). As a consequence, it follows that $(N,\delta) \mapsto q_{(i,j),(i',j')}(N,\delta)$ is continuous on $\mathcal{N}\times\mathcal{D}$.

  (ii) Let $0 < \gamma < \alpha$. Take $V(i,j) = \exp\{\varepsilon(i+j)\}$ with $\varepsilon = \frac{1}{2}\log(\gamma/\lambda + 1) > 0$. Then V clearly is a moment function. Moreover, it is a $(\mathcal{N}\times\mathcal{D},\gamma)$-drift function, since

$$\begin{aligned}
&\sum_{(i',j')} q_{(i,j),(i',j')}(N,\delta)V(i,j)\\
&\quad= \mathrm{e}^{\varepsilon(i+j)}\begin{cases}\min\{j,N\}\mu(\mathrm{e}^{-\varepsilon}-1)+\lambda(\mathrm{e}^{\varepsilon}-1), & \delta(i,j)=0,\ i=0,\\ \lambda(\mathrm{e}^{\varepsilon}-1), & \delta(i,j)=1,\ i=0,\\ \min\{j,N\}\mu(\mathrm{e}^{-\varepsilon}-1), & \delta(i,j)=0,\ i>0,\\ 0, & \delta(i,j)=1,\ i>0,\end{cases}\\
&\quad\le \lambda(\mathrm{e}^{\varepsilon}-1)\mathrm{e}^{\varepsilon(i+j)}\\
&\quad\le \lambda(\mathrm{e}^{2\varepsilon}-1)\mathrm{e}^{\varepsilon(i+j)}\\
&\quad= \gamma\mathrm{e}^{\varepsilon(i+j)}\\
&\quad= \gamma V(i,j).
\end{aligned}$$

  So V is a $(\mathcal{N}\times\mathcal{D},\gamma)$-drift function for $X(N,\delta)$.

  (iii) Take $W(i,j) = \exp\{2\varepsilon(i+j)\}$, then $W/V = V$ is a moment function. Hence, W is a V-moment function, in particular, W is a $(\mathcal{N}\times\mathcal{D},\gamma)$-drift V-moment function, since

$$\sum_{(i',j')} q_{(i,j),(i',j')}(N,\delta)W(i,j) \le \lambda(\mathrm{e}^{2\varepsilon}-1)\mathrm{e}^{2\varepsilon(i+j)} = \gamma\mathrm{e}^{2\varepsilon(i+j)} = \gamma W(i,j).$$

- Consider Assumption 4.1.

  (i) $(N,\delta) \mapsto c_{(i,j)}(N,\delta)$ is clearly continuous on $\mathcal{N}\times\mathcal{D}$ for any $(i,j) \in S$.

https://doi.org/10.1239/aap/1449859801 Published online by Cambridge University Press

(ii) Take $c_V = c/\lambda\varepsilon + \lambda K$. Then for any (i, j), (N, δ),

$$\begin{aligned}\frac{|c_{(i,j)}(n,\delta)|}{V(i,j)} &= \frac{ci/(\lambda + j\mu) + \lambda K \mathbf{1}_{\{i=0\}}}{\exp[\varepsilon(i+j)]} \\ &\le \frac{(c/\lambda)i + \lambda K}{1 + \varepsilon i} \\ &\le \frac{c}{\lambda\varepsilon} + \lambda K \\ &= c_V.\end{aligned}$$

Hence, the supremum over all (i, j), (N, δ) is also bounded by c_V.

- Condition (i) of Assumption 4.2 holds for both the parametrised process and the extended parametrised process.

  (i) The parameter set is a product space $\mathcal{D} = \prod_{(i,j)\in S} \mathcal{D}_{(i,j)}$ with $D_{(i,j)}$ a finite set; hence, compact and metric for each state $(i, j) \in \boldsymbol{S}$. The set $\mathcal{N}$ is compact; hence, $\mathcal{N} \times \mathcal{D}$ is compact.

  Condition (ii) of Assumption 4.2 only holds for the parametrised process $\{X(N, \delta)\}_\delta$, $N \in \mathcal{N}$, and not for the extended parametrised process.

  (ii) $\{Q(\delta)\}_{\delta\in\mathcal{D}}$ and $\{c(\delta)\}_{\delta\in\mathcal{D}}$ both have the product property. In other words, the transition rates and the cost rates in state (i, j) depend only on the action in state (i, j).

References

[1] Adan, I. J. B. F., Kulkarni, V. G. and van Wijk, A. C. C. (2013). Optimal control of a server farm. *Inf. Syst. Operat. Res.* **51,** 241–252.

[2] Anderson, W. J. (1991). *Continuous-Time Markov Chains*. Springer, New York.

[3] Federgruen, A. (1978). On N-person stochastic games with denumerable state space. *Adv. Appl. Prob.* **10,** 452–471.

[4] Guo, X. and Hernández-Lerma, O. (2003). Continuous-time controlled Markov chains with discounted rewards. *Acta Appl. Math.* **79,** 195–216.

[5] Guo, X. and Piunovskiy, A. (2011). Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Operat. Res.* **36,** 105–132.

[6] Guo, X. and Zhu, W. (2002). Denumerable-state continuous-time Markov decision processes with unbounded transition and reward rates under the discounted criterion. *J. Appl. Prob.* **39,** 233–250.

[7] Guo, X., Hernández-Lerma, O and Prieto-Rumeau, T. (2006). A survey of recent results on continuous-time Markov decision processes. *Top* **14,** 177–261.

[8] Hordijk, A. (1974). *Dynamic Programming and Markov Potential Theory* (Math. Centre Tracts **51**). Mathematisch Centrum, Amsterdam.

[9] Munkres, J. R. (2000). *Topology*, 2nd edn. Prentice Hall, Upper Saddle River, NJ.

[10] Piunovskiy, A. and Zhang, Y. (2014). Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach. *4OR-Q J. Operat. Res.* **12**, 49–75.

[11] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). Discounted continuous-time controlled Markov chains: convergence of control models. *J. Appl. Prob.* **49,** 1072–1090.

[12] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games* (ICP Adv. Texts Math. **5**). Imperial College Press, London.

[13] Reuter, G. E. H. (1957). Denumerable Markov processes and the associated contraction semigroups on l. *Acta Math.* **97,** 1–46.

[14] Royden, H. L. (1988). *Real Analysis*, 2nd edn. Macmillan, New York.
[15] Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd edn. McGraw-Hill, New York.
[16] Spieksma, F. M. (2012). Kolmogorov forward equation and explosiveness in countable state Markov processes. *Ann. Operat. Res.* 10.1007/s10479-012-1262-7.
[17] Spieksma, F. M. (2013). Countable state Markov processes: non-explosiveness and moment function. *Prob. Eng. Inf. Sci.* **29,** 623–637.