Structures of optimal policies in Markov Decision Processes with unbounded jumps: the State of our Art

H. Blok F.M. Spieksma

December 10, 2015

Contents

1 Introduction **1**

2 Discrete time Model **4**
2.1 Discounted cost 6
2.2 Approximations/Perturbations 10
2.3 Average cost 14

3 Continuous time parametrised Markov processes **19**
3.1 Uniformisation 20
3.2 Discounted cost 21
3.3 Average cost 23
3.4 Roadmap to structural properties 24

A Proof of Theorem 3.5 **28**

B Tauberian Theorem **33**

Acknowledgement **37**

1 Introduction

The question how to rigorously prove structural results for continuous time Markov decision problems (MDPs) with a countable state space and unbounded jump rates (as a function of state) seems to be an assiduous task. As a typical example one may consider the competing queues model with queue dependent cost rates per customer and per unit time, where the objective is to determine the server allocation that minimises the total expected discounted cost or expected average cost per unit time. Both discounted and average cost are known to be minimised by the $c\mu$-rule, which prescribes to allocate the server to the queue that yields the largest cost reduction per unit time. A possible method to tackle this problem is to apply Value Iteration (VI) to the uniformised discrete time MDP and show that optimality of the $c\mu$-rule propagates through the VI step.

If customer abandonment is allowed, the resulting MDP is a continuous time MDP with unbounded jumps, since customers may renege at a rate that is propertional to the number of customers present in each of the queues. In order to apply VI, one has to time discretise the MDP. One way to associate a discrete time MDP with this problem is by constructing the decision process embedded on the jumps of the continuous time MDP. However, it is not clear whether structural properties

propagate through the VI step (cf. Section 3.4). Another solution is to perturb or truncate the continuous time MDP, so that it becomes uniformisable and then apply VI. A suitable truncation or perturbation needs to be invariant with respect to structural properties of interest of the investigated MDP.

The first question is whether there exists generic truncation methods that possess such an invariance property. Clearly, this can never be systematically proved, since it depends on the properties that one wishes to prove. However, one might be able to formulate recommendations as to what kind of perturbation methods perform well, with regard to such an invariance requirement.

The paper [20] studies two competing queues with abandonments, and a problem-specific truncation is used. Later [7] has introduced a truncation method, called Smoothed Rate Truncation (SRT) that so far seems to work well for problems where a simple bounded rate truncation (as in Section 3.4) does not work. In addition, it can be used for numerical applications in bounded rate optimisation problems (cf. Section 2.2). The SRT method has been used in a companion paper [6] for identifying conditions under which a simple index policy is optimal in the competing queues problem with abandonments.

Consecutively, suppose that an application of the truncation method yields truncated MDPs with optimal policies and a value function that have the properties of interest. For instance, the above mentioned application of SRT to the competing queues example with abandoning customers yields optimality of an index policy for each truncated MDP. However, these properties still have to be shown to apply to the original nontruncated problem. Thus, convergence results are required that yield continuity of the optimal policy and value function in the truncation parameter, in order to deduce the desired results for the nontruncated MDP from the truncated ones.

A second question therefore is as to what kind of easily verifiable conditions on the input parameters of the perturbed MDP guarantees convergence of the value function and optimal policy to the corresponding ones of the original unperturbed MDP. In [20], the authors had to prove a separate convergence result apart from devising a suitable truncation and prove that VI propagates the properties of interest. Apparently, theory based on a set of generic conditions that can encorporate convergence within the optimisation framework was lacking. This lack is precisely what has hampered the analysis of the server farm model in [1], where the authors have restricted their analysis to showing threshold optimality of a bounded rate perturbed variant of the original model. Apparently no appropriate tools for deducing threshold optimality of the original unbounded problem from the results for the perturbed one were available to them.

A third major problem occurs in the context of the average cost criterion. In particular, VI is not always guaranteed to converge. This is true under weak assumptions in discounted cost problems, however, in average cost problems there are only limited convergence results (cf. Section 2.3). One of these requires strong drift conditions, that do not allow transience under any stationary deterministic policy. However, often this is not a convenient requirement. One may get around this difficulty by a vanishing discount approach, which analyses the expected average cost as a limit of expected α-discounted costs as the discount factor tends to 0 (or 1, depending on how the discount factor is modelled).

For a model like the competing queues model with abandonments, a multistep procedure to obtain structural results for the average cost problem then would be as follows. First, consider the α-discounted cost truncated problem. Structural results for the α-discounted cost non-truncated problem follow, by taking the limit for the truncation parameter to vanish. Finally, taking the limit of the discount factor to 0, hopefully yields the final structural results for the original continuous time average cost problem.

For some of these steps theoretical validation has been provided for in the literature, but not for all and not always under conditions that are easily checked. The main focus of this chapter is to fill

some gaps in the described procedure, whilst requiring conditions that are formulated in terms of the input parameters. Based on the obtained results, we aim to provide a systematic and feasible approach for attacking the validation of structural properties, in the spirit of the multistep procedure sketched above. We hope that this multistep procedure will also be beneficial to other researchers as a roadmap for tackling the problem of deriving structural results for problems modelled as MDPs.

We do not address the methods of propagating structures of optimal policies and value function through the VI step. Such methods belong to the domain of 'Event based dynamic programming', and they have been discussed thoroughly in [29], with extensions to SRT in [11]. Furthermore, we do not include an elaborate evaluation of close results from the literature. Some detailed comments has been included in the paper, whenever we thought it relevant.

Another omission in this work is the study of perturbed MDPs with the average cost criterion. However, the conditions required for achieving the desired continuity results as a function of a perturbation parameter are quite strong. Therefore a more recommendable approach would be the one we have developed in this paper, using the vanishing discount approach. As a last remark: we generally restrict to the class of stationary policies, and not history-dependent ones. Especially the results quoted for discrete time MDPs apply to the larger policy class. In continuous time MDPs allowing history-dependent policies causes extra technical complications that we do not want to address in this work.

A short overview of the paper content is provided next. In Section 2 we discuss discrete time, countable state MDPs, with compact action sets. First, the α-discount optimality criterion is discussed, cf. Section 2.1. This will be the base case model, to which the MDP problems might have to be reduced in order to investigate its structural properties. We therefore describe it quite elaborately. In addition, we have put it into a framework that incorporates truncations or perturbations. We call this a parametrised Markov process. Interestingly enough, 'standard' but quite weak drift conditions introduced for α-discounted cost MDPs in discrete time, allowed this extension to parametrised Markov processes, with no extra effort and restriction. It incorporates the finite state space case, elaborated on in the seminal book [19].

In Section 2.2 we provide a discussion of SRT, as a method for numerical investigation of structural properties of a countable state MDP. The conditions that we use are a weak drift condition on the parametrised process, plus reasonable continuity conditions. This has been based on the work in [30, 49] for MDPs.

In Section 2.3 we study the expected average cost criterion, whilst restricting to non-negative cost, i.e. negative dynamic programming. This restriction allows transience, and the analysis follows [15], in the form presented by [40]. Basically, the conditions imposed require the existence of one 'well-behaved' policy, and a variant of inf-compact costs. The latter ensures that optimal policies have a guaranteed drift towards a finite set of low cost states. The contribution of these works is that they validate the vanishing discount approach, thus allowing to analyse the discrete time average cost problem via the discrete time α-discounted cost problem.

Then we turn to studying continuous time MDPs in Section 3. First the α-discounted cost problem is considered. The drift conditions on parametrised discrete time Markov processes have a natural extension to continuous time. The results listed are based on [13], but the literature contains quite some work in the same spirit within the framework of MDPs with more general state spaces, cf. e.g. [25, 34], and references therein. A closely related perturbation approach has been studied in [33]. Since perturbations are incorporated in the parametrised framework, the approach allows to study bounded jump perturbations. Indeed, optimal policies and value functions are continuous as a function of the perturbation parameter. In this way, [13] obtains threshold optimality of the original unbounded α-discounted cost variant of the server farm model studied in [1].

Finally, for the expected average cost criterion, we use the natural generalisation of the discrete

time conditions. Although closely related to analyses in [25, 34] and references therein, as far as we know this form has not appeared yet in the literature. The vanishing discount approach is validated in the same way as was done for the discrete time MDP. This reduces the problem of studying structural properties for average cost MDPs, satisfying the proposed conditions, to analysing a continuous time α-discounted cost MDP, for which the solution method has already been described. As a consequence, also average cost threshold optimality for the above mentioned server farm model from [1] follows from α-disount optimality of a threshold policy, cf. [14].

Dispersed through the paper are roadmaps for attacking the validation of structural properties. These are summarised in Section 3.4.

2 Discrete time Model

In this section we will set up a framework of parametrised Markov processes in discrete time. With an extra assumption – the product property – a parametrised Markov process reduces to a discrete time MDP.. However, treating this in the parametrised framework allows for results on perturbations or approximations of MDPs as well. Notice that instead of the usual nomenclature 'Markov chain' for a Markov process in discrete time, we will consistently use 'Markov process', whether it be a process in discrete or continuous time.

Let Φ be a parameter space. Let $\boldsymbol{S}$ denote a countable space. Each parameter ϕ is mapped to an $\boldsymbol{S} \times \boldsymbol{S}$ stochastic matrix $P(\phi)$, and a cost vector $c(\phi) : \boldsymbol{S} \to \mathbf{R}$. We denote the corresponding elements by $p_{xy}(\phi)$, $x, y \in \boldsymbol{S}$ and $c_x(\phi)$, $x \in \boldsymbol{S}$. If $f : \boldsymbol{S} \to \mathbf{R}$, then $P(\phi)f$ is the function with value

$$P(\phi)f_x = \sum_y p_{xy}(\phi) f_y$$

at point $x \in \boldsymbol{S}$, provided the integral is well-defined.

To transition matrix $P(\phi)$ one can associate a Markov process on the path space $\Omega = \boldsymbol{S}^\infty$. Given a distribution ν on $\boldsymbol{S}$, the Kolmogorov consistency theorem (see e.g. [10]) provides the existence of a probability measure P_ν^ϕ on Ω, such that the canonical process $\{X_n\}_n$ on Ω, defined by

$$X_n(\omega) = \omega_n$$

is a Markov process with transition matrix $P(\phi)$, and probability distribution P_ν^ϕ. The corresponding expectation operator is denoted by E_ν^ϕ. To avoid overburdened notation, we have put the dependence on the parameter ϕ in the probability and expectation operators, and not in the notation for the Markov process, which we did do in the earlier paper [13]. We further denote $P^{(n)}(\phi)$ for the n-th iterate of $P(\phi)$, where $P^{(0)}(\phi) = \mathbf{I}$ equals the $\boldsymbol{S} \times \boldsymbol{S}$ identity matrix.

We assume the following basic assumption.

Assumption 2.1 The following conditions hold:

i) the parameter space Φ is locally compact;

ii) $\phi \mapsto p_{xy}(\phi)$ continuous on Φ for each $x, y \in \boldsymbol{S}$;

iii) $\phi \mapsto c_x(\phi)$ is continuous on Φ for each $x \in \boldsymbol{S}$.

To incorporate MDPs in this set up, we use the following concept.

Definition 2.1 Let $\Phi' \subset \Phi$, inheriting the topology on Φ. We say that $\{P(\phi), c(\phi)\}_{\phi \in \Phi'}$ has the product property with respect to Φ' if

i) there exist compact sets Φ'_x, $x \in \boldsymbol{S}$, such that $\Phi' = \prod_{x\in S} \Phi_x$; then Φ' is compact in the product topology;

ii) for any $\phi, \phi' \in \Phi'$, $x \in \boldsymbol{S}$ with $\phi_x = \phi'_x$, it holds that

- $(P(\phi))_{x\cdot} = (P(\phi'))_{x\cdot}$, where $(P(\phi))_{x\cdot}$ stands for the x-row of $P(\phi)$;
- $c_x(\phi) = c_x(\phi')$.

For notational convenience we will simply say that Φ' has the product property. Under the product property, with a slight abuse of notation we may write $c_x(\phi_x)$ and $p_{xy}(\phi_x)$ instead of $c_x(\phi)$ and $p_{xy}(\phi)$. In case the dependence on ϕ is expressed in the probability or expectation operators, we write c_{X_n} instead $c_{X_n}(\phi)$.

Remark 2.1 If Φ has the product property, then the parametrised Markov process is an MDP. The set Φ may represent the collection of deterministic stationary policies, and we will denote it by $\mathcal{D}$. In this case $\mathcal{D}_x$ is the action set in state $x \in \boldsymbol{S}$.

For any $x \in \boldsymbol{S}$, let π_x by a probability distribution on $\mathcal{D}_x$. Then $\pi = (\pi_x)_x$ is a stationary, randomised policy. The collection Π of all stationary randomised policies can be viewed as a parameter set having the product property as well. We will not consider this explicitly, but all discussed results cover this case as well.

Next we define the various performance measures and optimality criteria that we will study. Lateron we will provide conditions under which these are well-defined, and optimal polices exist.

For $0 < \alpha < 1$, define the *expected total α-discounted cost value function* $v^\alpha(\phi)$ under parameter $\phi \in \Phi$ by

$$v_x^\alpha(\phi) = \mathsf{E}_x^\phi\Big[\sum_{n=0}^{\infty}(1-\alpha)^n c_{X_n}\Big], \quad x \in \boldsymbol{S}. \tag{2.1}$$

Notice that the discount factor is taken to be $1-\alpha$. Usually the discount factor is taken to equal α instead. Our choice here allows a more direct analogy with the continuous time case.

Next let $\Phi' \subset \Phi$ have the product property. Define the *minimum expected total α-discounted cost* v^α w.r.t. Φ' by

$$v_x^\alpha = \inf_{\phi\in\Phi'}\{v_x^\alpha(\phi)\}, \quad x \in \boldsymbol{S}.$$

If for some $\phi \in \Phi'$ it holds that $v^\alpha = v^\alpha(\phi)$, then ϕ is said to be α-discount optimal (in Φ').

The *expected average cost* $\mathsf{g}(\phi)$ under parameter $\phi \in \Phi$ is given by

$$\mathsf{g}_x(\phi) = \limsup_{N\to\infty}\frac{1}{N+1}\mathsf{E}_x^\phi\Big[\sum_{n=0}^{N} c_{X_n}\Big], \quad x \in \boldsymbol{S}.$$

If $\Phi' \subset \Phi$ has the product property, the *minimum expected average cost* w.r.t. Φ' is defined as

$$\mathsf{g}_x = \inf_{\phi}\{\mathsf{g}_x(\phi)\}, \quad x \in \boldsymbol{S}.$$

If for $\phi \in \Phi'$ it holds that $\mathsf{g}(\phi) = \mathsf{g}$, then ϕ is called average optimal (in Φ').

A stronger notion of optimality, called Blackwell optimality, applies more often than is generally noted. We define it next (see also [16]).

Let $\Phi' \subset \Phi$ have the product property. The policy $\phi^* \in \Phi'$ is *Blackwell optimal* w.r.t. Φ', if for any $x \in \boldsymbol{S}$, $\phi \in \Phi'$, there exists $\alpha(x,\phi) > 0$, such that $v_x^\alpha(\phi^*) \le v_x^\alpha(\phi)$ for $\alpha < \alpha(x,\phi)$. Additionally, ϕ^* is *strongly Blackwell optimal* if $\inf_{x\in S,\phi\in\Phi'}\alpha(x,\phi) > 0$.

2.1 Discounted cost

To determine the discounted cost v^α an important instrument is the *α-(discrete time) discount optimality equation* (α-DDOE)

$$u_x = \inf_{\phi_x \in \Phi_x} \Big\{ c_x(\phi_x) + (1-\alpha) \sum_{y \in S} p_{xy}(\phi_x) u_x \Big\}, \quad x \in \boldsymbol{S}, \tag{2.2}$$

for $\Phi' = \prod_{x \in S} \Phi'_x$ having the product property. In this subsection we show that mild conditions guarantee the existence of a unique solution to this equation in a certain space of functions. Moreover, the inf is a min, and a minimising policy in (2.2) is optimal in Φ' (and even optimal within the larger set of randomised and nonstationary policies generated by Φ').

The condition used here has been taken from [30, 49].

Definition 2.2 The function $V : \boldsymbol{S} \to (0, \infty)$ is called a (γ, Φ)-drift function if $P(\phi)V \leq \gamma V$ for all $\phi \in \Phi$. Note that '$\leq$' stands for componentwise ordering.

Definition 2.3 The Banach space of 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 $f : \boldsymbol{S} \to \mathbf{R}$ and

$$\|f\|_V = \sup_{x \in S} \frac{|f_x|}{V_x} < \infty.$$

Assumption 2.2 ($\boldsymbol{\alpha}$) i) There exist a constant $\gamma < 1/(1-\alpha)$ and a function $V : \boldsymbol{S} \to (0, \infty)$ such that V is (γ, Φ)-drift function and that $\phi \mapsto P(\phi)V$ is component-wise continuous;

ii) $c_V := \sup_\phi \|c(\phi)\|_V < \infty$.

The above assumption allows to rewrite (2.1) as

$$v^\alpha(\phi) = \sum_{n=0}^{\infty} (1-\alpha)^n P^{(n)}(\phi) c(\phi). \tag{2.3}$$

The following lemma is quite straightforward to prove. For completeness we give the details.

Lemma 2.4. *Suppose that the Assumptions 2.1 and 2.2 (α) hold, then $\phi \mapsto v^\alpha(\phi)$ is componentwise continuous and $v^\alpha(\phi)$ is the unique solution in $\ell^\infty(\boldsymbol{S}, V)$ to*

$$u = c(\phi) + (1-\alpha) P(\phi) u. \tag{2.4}$$

Proof. First notice that $v^\alpha(\phi) \in \ell^\infty(\boldsymbol{S}, V)$, since

$$\begin{aligned} |v^\alpha(\phi)| &= |\sum_{n=0}^{\infty} (1-\alpha)^n P^{(n)}(\phi) c(\phi)| \leq \sum_{n=0}^{\infty} (1-\alpha)^n P^{(n)}(\phi) c_V \cdot V \\ &\leq (1-\alpha)^n \gamma^n c_V \cdot V = \frac{c_V}{1-(1-\alpha)\gamma} V. \end{aligned}$$

Next, $v^\alpha(\phi)$ is a solution to (2.4), since

$$\begin{aligned}(1-\alpha)P(\phi)v^\alpha(\phi) &= (1-\alpha)P(\phi)\sum_{n=0}^{\infty}(1-\alpha)^n P^{(n)}(\phi)c(\phi)\\ &= \sum_{n=1}^{\infty}(1-\alpha)^n P^{(n)}(\phi)c(\phi) = v^\alpha(\phi) - c(\phi).\end{aligned}$$

Let $f = (f_x)_x \in \ell^\infty(\boldsymbol{S}, V)$ be any solution to (2.4), then

$$\begin{aligned}v_x^\alpha(\phi) - f_x &= (1-\alpha)\sum_y p_{xy}(\phi)(v_y^\alpha(\phi) - f_y)\\ &= (1-\alpha)^n \sum_y p_{xy}^{(n)}(\phi)(v_y^\alpha(\phi) - f_y).\end{aligned}$$

Hence,

$$\begin{aligned}|v_x^\alpha(\phi) - f_x| &\le (1-\alpha)^n \sum_y p_{xy}^{(n)}(\phi)|v_x^\alpha(\phi) - f_x|\\ &\le (1-\alpha)^n P^{(n)}(\phi)V_x \cdot (c_V + \|f\|_V)\\ &\le (1-\alpha)^n \gamma^n V_x \cdot (c_V + \|f\|_V) \to 0, \quad n \to \infty.\end{aligned}$$

This implies $f = v^\alpha$, hence v^α is the unique solution to (2.4) in $\ell^\infty(\boldsymbol{S}, V)$.

Finally, to show $\phi \mapsto v_x^\alpha(\phi)$, $x \in \boldsymbol{S}$, is continuous, notice that by assumption $\phi \mapsto P(\phi)V$ is component-wise continuous. It follows that $\phi \mapsto P^{(n)}(\phi)V$ component-wise continuous. Since $P^{(n)}(\phi)V \le \gamma^n V$, the dominated convergence theorem yields that $\phi \mapsto \sum_{n=0}^\infty (1-\alpha)^n P^{(n)}(\phi)V < \infty$ component-wise continuous. Further, since $\phi \mapsto c(\phi)$ is componentwise continuous and $|c(\phi)| \le c_V \cdot V$, an application of the generalised dominated convergence theorem ([37, Proposition 11.18]) implies componentwise continuity of $\phi \mapsto \sum_{n=0}^\infty (1-\alpha)^n P^{(n)}(\phi)c(\phi) = v^\alpha(\phi)$. □

The following theorem is a well-known result by Wessels [49].

Theorem 2.5 (cf. Wessels [49]). *Suppose that $\Phi' = \prod_x \Phi'_x$ has the product property and that Assumptions 2.1 and 2.2(α) hold. Then v^α is the unique solution in $\ell^\infty(\boldsymbol{S}, V)$ to the α-discounted optimality equation (α DDOE) (2.2).*

Moreover, the infimum is attained as a mininum. For any $\phi^* = (\phi_x^*)_x \in \Phi'$, for which ϕ_x^* achieves the minimum in (2.2) for all $x \in \boldsymbol{S}$, it holds that $v^\alpha(\phi^*) = v^\alpha$ and ϕ^* is (α-discounted) optimal in Φ'.

The versatile applicability of (γ, Φ)-drift functions is illustrated in [6, 14] and the example below.

Example 2.1 First note for the bounded cost case that the function $V_x \equiv 1$ is an appropriate function satisfying Assumption 2.2(α).

As a simple example, consider a discrete time single server queue, where the probability of an arrival in the next time slot is $\lambda \in (0,1)$. The system state represents the number of customers in the system, hence, the state space is $\boldsymbol{S} = \{0, 1, \ldots\}$. The probability of a service completion in the next time slot depends on a parameter $\phi = \{\phi_x\}_{x=1}^\infty$, where ϕ_x stands for the probability of a service completion in the next time-slot, indepent of arrivals, when the system state is x. The parameter ϕ stands for an a priori determined service control. The parameter space Φ maybe any compact subset of $\{0\} \times [0,1]^\infty$.

To any fixed parameter ϕ, one may associate a Markov process, representing the system state at any time t, with transition probabilities given by

$$p_{xy}(\phi) = \begin{cases} p(1-\phi_x), & y = x+1 \\ (1-p)\phi_x, & y = (x-1)\mathbf{1}_{\{x\neq 0\}} \\ 1-p-\phi_x+2p\phi_x, & y = x. \end{cases}$$

As an appropriate (γ, Φ)-drift function, we may choose $V_x = e^{\epsilon x}$, with $\epsilon > 0$ to be determined below:

$$\begin{aligned} \sum_y p_{xy}(\phi)e^{\epsilon y} &= (1-p)\phi_x e^{\epsilon(x-1)} + (1-(1-p)\phi_x - p(1-\phi_x))e^{\epsilon x} + p(1-\phi_x)e^{\epsilon(x-1)} \\ &= e^{\epsilon x}\Big(1 + p(1-\phi_x)(e^{\epsilon}-1) + (1-p)\phi_x(1-e^{-\epsilon})\Big) \\ &\leq e^{\epsilon x}\big(1 + p(e^{\epsilon}-1)\big). \end{aligned}$$

For $\epsilon = 0$, the coefficient of $e^{\epsilon x}$ in the above equals 1. Since $1/(1-\alpha) > 1$, one can always choose ϵ small enough so that

$$\gamma := e^{\epsilon x}\big(1 + p(e^{\epsilon}-1)\big) < \frac{1}{1-\alpha}.$$

As the example shows, the existence of a (γ, Φ)-drift function does not impose any restrictions on the class structure of the associated Markov processes and transience is allowed as well. Moreover, it is often a good and simply checked choice to take V exponential. Since generally cost structures are linear or quadratic as a function of state, they are dominated by exponential functions. Thus, they fit in the framework discussed here.

Value Iteration A very important algorithm to calculate v^α is the value iteration algorithm (VI), originally due to Bellman [5].

Algorithm 1 VI for an α-discounted cost ϵ-optimal policy

1. Select $v^{\alpha,0} \in \ell^\infty(\boldsymbol{S}, V)$, specify $\epsilon > 0$, set $n = 0$.

2. For each $x \in \boldsymbol{S}$, compute $v_x^{\alpha,n+1}$ by

$$v_x^{\alpha,n+1} = \min_{\phi_x \in \Phi'_x} \Big\{ c_x(\phi_x) + (1-\alpha)\sum_{y\in S} p_{xy}(\phi_x) v_x^{\alpha,n} \Big\}, \tag{2.5}$$

   and let

$$\phi^{n+1} \in \arg\min_{\phi\in\Phi'} \{c(\phi) + (1-\alpha)P(\phi)v^{\alpha,n}\}. \tag{2.6}$$

3. If

$$\|v^{\alpha,n+1} - v^{\alpha,n}\|_V \leq \frac{1-(1-\alpha)\gamma}{2(1-\alpha)\gamma}\epsilon, \tag{2.7}$$

   then put $v^\epsilon := v^{\alpha,n+1}$, $\phi^\epsilon := \phi^{n+1}$, stop. Otherwise increment n by 1 and return to step 2.

Theorem 2.6 (cf. [49], [35, Theorem 6.3.1]). *Suppose that $\Phi' = \prod_x \Phi'_x \subset \Phi$ has the product property and that Assumptions 2.1 and 2.2(α) hold. Let let $v^{\alpha,0} \in \ell^\infty(\boldsymbol{S}, V)$ and $\epsilon > 0$. Let $\{v^{\alpha,n}\}_{n\in\mathbb{N}}$ satisfy (2.5) for $n \geq 1$. Then the following hold.*

i) $\lim_{n\to\infty} \|v^\alpha - v^{\alpha,n}\|_V = 0$, *in particular,*

$$\|v^\alpha - v^{\alpha,n}\|_V \leq \frac{1}{1-(1-\alpha)\gamma}\|v^{\alpha,n+1} - v^{\alpha,n}\|_V \leq \frac{((1-\alpha)\gamma)^n}{1-(1-\alpha)\gamma}\|v^{\alpha,1} - v^{\alpha,0}\|_V.$$

Any limit point of the sequence $\{\phi^n\}_n$ *is an* α*-discount optimal policy.*

ii) v^ϵ *is an* $\epsilon/2$*-approximation of* v^α*, in other words,* $\|v^\alpha - v^{\alpha,n+1}\|_V \leq \frac{\epsilon}{2}$.

iii) ϕ^ϵ *is an* ϵ*-optimal policy, in other words,* $\|v^\alpha - v^\alpha(\phi^\epsilon)\|_V \leq \epsilon$.

Proof. The proof of Theorem 2.6 (i) is straightforward using that

$$v^\alpha - v^{\alpha,n} = \lim_{N\to\infty} \sum_{k=n}^{N} (v^{\alpha,k+1} - v^{\alpha,k}).$$

The bounds are somewhat implicit in [49]. They are completely analogously to the bounds of e.g. [35, Theorem 6.3.1] for the bounded reward case, with λ replaced by $(1-\alpha)\gamma$. The derivation is similar. □

Remark 2.2 The reader may wish to point out that a solution to the α-DDOE yielding an α-discount deterministic policy exists without any further conditions in the case of non-negative cost (negative dynamic programming, cf. [45]) and a finite action space per state. Also VI converges provided $v_0 \equiv 0$, although no convergence bounds can be provided. If the action space is compact, additional continuity and inf-compactness (cf. [21, Corollary 5.7]) properties are necessary for the existence of a stationary deterministic policy attaining the minimum in the α-DDOE. It is not clear to us how these conditions could be extended in order to include parametrised Markov processes.

Notice further, that unfortunately in general there is no unique solution to the α-DDOE (cf. [21], [40, Section 4.2]). Using norm conditions as in this paper, allows to identify the value function as the unique one in the Banach space of functions bounded by V (cf. Theorem 2.5). In the non-negative cost case, the value function is the minimum solution to the α-DDOE (see [40, Theorem 4.1.4]).

In case of a finite state space, VI can be numerically implemented. In the case of a countable space, its use is restricted to the derivation of structural properties of the value function and α-discount optimal policy. Structural properties such as non-decreasingness, convexity, etc. can be used to show for instance that a threshold policy or an index policy is optimal.

To prove properties via VI, first select a function v_0 possessing the properties of interest. Then show by induction that $v^{\alpha,n}$ has this property for all n. Under the assumptions of Theorem 2.6 one has $v^{\alpha,n} \to v^\alpha$, for $n \to \infty$, and so we may conclude v^α has this property as well. The existence of an optimal policy with desired properties can be directly derived from the structure of the value function v^α in combination with the α-DDOE. Alternatively, this can be deduced from the fact that since each ϕ^n has these properties, any limit point has.

The main reference on the propagation of structural properties through the VI induction step (2.5) is [29]. The technique discussed in this monograph is called 'Event Based Dynamic Programming', and it presents a systematic framework using 'event operators' for the propagation of the desired structural properties. New operators have been developed in [7, 11] for special perturbations or truncations of non-uniformisable MDPs, as described below. In Section 3.4 we present an example.

2.2 Approximations/Perturbations

Next we focus our attention to parameters capturing a perturbation of the MDP. This parameter set should capture the collection of deterministic policies $\mathcal{D}$, as well as a perturbation set $\mathcal{N}$. This perturbation can have multiple interpretations, depending on the context. It can be a finite state approximation, or it can represent some uncertainty in the input parameters. Put $\Phi = \mathcal{N} \times \mathcal{D}$. Notice, that the set $\{N\} \times \mathcal{D} \subset \Phi$ need not automatically have the product property, $N \in \mathcal{N}$.

The following continuity result follows directly from Lemma 2.4.

Corollary 2.7 (to Lemma 2.4 and Theorem 2.5)**.** *Suppose that Assumptions 2.1, and 2.2*(α) *hold. Further assume that* $\{N\} \times \mathcal{D}$ *has the product property, for* $N \in \mathcal{N}$*. Then,*

i) $\lim_{N \to N_0} v^\alpha(N) = v^\alpha(N_0)$*;*

ii) any limit point of $\{\delta_N^*\}_{N \to N_0}$ *is optimal in* $\{N_0\} \times \mathcal{D}$*.*

Without the existence of a (γ, Φ)-drift function bounding the one-step cost uniformly in the parameter, the above convergence result may fail to hold.

Example 2.1. (cf. [40, Example 4.6.1]) Let the parameter set $\mathcal{N} = \mathbb{N} \cup \{\infty\}$, $\boldsymbol{S} = \{0, 1, \ldots\}$ for $N \in \mathcal{N}$. The transition probabilities are as follows.

$$p_{xy}(\infty) = \frac{1}{2} \quad y \in \{0, x+1\},$$

and for $N < \infty$

$$p_{xy}(N) = \begin{cases} \frac{1}{2}, & x \neq N-1, N, y = 0 \\ \frac{1}{2} - \frac{1}{N}, & x \neq N-1, N, y = x+1 \\ \frac{1}{N}, & x \neq N-1, N, y = N \\ 1 - \frac{1}{N}, & x = N-1, y = 0 \\ \frac{1}{N}, & x = N-1, y = N \\ 1, & x = y = N. \end{cases}$$

Further let $\alpha < \frac{1}{2}$ and define $c_x(N) = x^2$ for $N \leq \infty$. The calculations in [40, Example 4.6.1] show that $v_0^\alpha(\infty) < \lim_{N \to \infty} v_0^\alpha(N) = \infty$. It is simply checked that any (γ, Φ)-drift function can at most be linear. Indeed for $1 < N < \infty$, it must hold that

$$\frac{1}{2}V(0) + (\frac{1}{2} - \frac{1}{N})V(2) + \frac{1}{N}V(N) \leq \gamma V(1),$$

leading to the requirement that $V(N) \leq N \cdot V(1)$. This implies that $\sup_{N,x} \frac{c_x(N)}{V_x} = \infty$.

Literature related to Corollary 2.7 can be found in [35, Section 6.10.2] and [40, Section 4.6]. Both consider finite space truncations. The first reference further assumes the existence of a (γ, Φ)-drift function without continuity, but with an additional tail condition and a prescribed truncation method. These conditions are implied by ours.

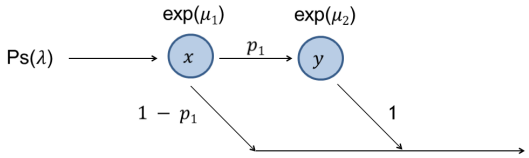
The second reference considers minimisation of non-negative costs, as well as a finite action space per state. No assumption on the truncation method needs to be made if there is a uniform bound on the costs. If the costs are allowed to be unbounded as a function of state, then conditions on the truncation method have to be made. A related setup to the one presented in the present paper is discussed in [33, Theorem 3.1], but the conditions imposed are (slightly) more restrictive (cf. [13, Remark 5.2]).

Example 2.2. The above example is a case where neither the conditions from [40] are met, nor the ones presented in this paper.

However, the conditions in the approximating sequence method of [40] are even not fulfilled, if we change the cost function to $c_x(N) = x$ for all $N \leq \infty$, $x \in \boldsymbol{S}$. On the other hand, the function V defined by $V_x = x$, $x \in \boldsymbol{S}$, is a $(\gamma, \mathcal{N})$-drift function, for which the conditions of Corollary 2.7 are trivially met. Hence, the set-up in this paper can be applied and we find that $\lim_{N\to\infty} v_0^\alpha(N) = v_0^\alpha(\infty) < \infty$.

Type of perturbations Although any perturbation satisfying the conditions of Corollary 2.7 yields (componentwise) continuity of the value function as a function the perturbation parameter, not any perturbation is desirable in terms of structural properties.

To explain this, consider the following server allocation problem.



Customers arrive at a service unit according to a Poisson(λ) process. Their service time at unit 1 takes an exponentially distributed amount of time with parameter μ_1. After finishing service in unit 1, with probability p_1 an additional exp(μ_2) amount of service is requested in unit 2, and with probability $1 - p_1$ the customer leaves the network, $p_1 \in (0, 1)$. There is only one server who has to be allocated to one of the units. We assume that idling is not allowed. The goal is to determine an allocation policy that minimises the α-discounted holding cost. The holding cost per unit time is given by the number of customers in the system.

Clearly this problem can be modelled as a continuous time MDP. However, it is equivalent to study the associated uniformised discrete time system (cf. Section 3.1). The data of this discrete time MDP are as follows. Denote by X_n the number of customers in unit 1 and unit 2 respectively, at time n, $n = 0, 1, \ldots$. Then $\boldsymbol{S} = \mathbf{Z}_+^2$, where state (x, y) represents that x customers are present in unit 1 and y in unit 2. By uniformisation we may assume that $\lambda + \mu_1 + \mu_2 = 1$, and so the rates represent probabilities. Independently of the allocation decision, a cost $c_{x,y}(\phi) = x + y$ is incurred.

Suppose that the state equals (x, y). If both units are non-empty, then either unit 1 is served (decision $\delta_{x,y} = 1$) or unit 2 (decision $\delta_{x,y} = 2$). If one of the units is empty, but not both, the server will be allocated to the non-empty unit during the next time-slot. This leads to the following transition probabilities:

$$p_{(x,y)(x',y')}(\delta_{x,y}) = \begin{cases} \lambda, & x' = x+1, y' = y \\ p_1\mu_1, & x > 0, x' = x-1, y' = y+1, \delta_{x,y} = 1 \\ (1-p_1)\mu_1 & x > 0, x' = x-1, y' = y, \delta_{x,y} = 1 \\ \mu_2 & y > 0, x' = x, y' = y-1, \delta_{x,y} = 2 \\ 1 - \sum_{(x'',y'')\neq(x,y)} p_{(x,y)(x'',y'')}(\delta_{x,y}), & x' = x, y' = y. \end{cases}$$

Let the discount factor α be given. It is easily verified that there exists a $(\gamma, \mathcal{D})$-drift function $V : \boldsymbol{S} \to \mathbf{R}_+$ of the form

$$V(x, y) = e^{\epsilon_1 x + \epsilon_2 y}, \tag{2.8}$$

with $\gamma < 1/(1-\alpha)$ and $\epsilon_1, \epsilon_2 > 0$. Assumptions 2.1 and 2.2 (α) are satisfied for V and $\Phi = \mathcal{D}$ and so the results of Theorem 2.5 apply.

Assume that $(1-p_1)\mu_1 > \mu_2$. By using VI, event based dynamic programming yields that allocating the server to unit 1, when non-empty, is α-discount optimal. Indeed, this gives a larger cost reduction per unit time due to customer service completions than allocating to unit 2. Thus, noticing that the cost rate per unit time and per server unit are equal to 1, this allocation policy is a generalised $c\mu$-rule. Let us refer to this policy as AP1 (allocation to unit 1 policy).

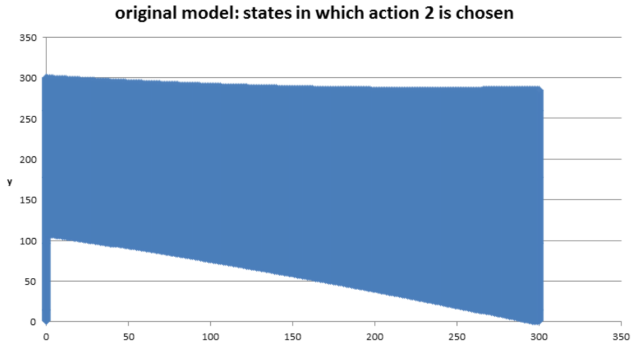
Since this is true for any $0 < \alpha < 1$, AP1 is *strongly Blackwell optimal*! We therefore expect to see this structure in numerical experiments. To perform such an experiment, one needs truncate the state space.

Straightforward perturbation A straightforward truncation is as follows. Choose $M, N \in \mathbf{Z}_+$, $N, M > 0$. At the truncation boundary $\{(x,y) \,|\, x = N, \text{ and/or } y = M\}$, transitions leading out of the rectangle $\{(x,y) \,|\, x \leq N, y \leq M\}$ are directed back to the initial state. The perturbation set is $\mathcal{N} = \big\{(N,M) \,|\, N, M \in \{1,2,\ldots\}\big\}$. Also in this case, one can easily check that there exist ϵ_1, $\epsilon_2 > 0$ and $\gamma \in \mathbf{R}$, such that V from (2.8) is a $(\gamma, \mathcal{N} \times \mathcal{D})$-drift function for ϵ_1, ϵ_2 small enough. Moreover, Assumptions 2.1 and 2.2(α) are satisfied for this V and for $\Phi = \mathcal{N} \times \mathcal{D}$. Note that the states outside the rectangle $\{(x,y) \,|\, x \leq N, y \leq M\}$ have become transient, and so the choice of actions in those states does not affect the optimal actions within the rectangle.

Let δ^* be α-discount optimal for $\mathcal{D}$, and $\delta^{N,M}$ for $\{(N,M)\} \times \mathcal{D}$. Then by virtue of Lemma 2.4

$$v^\alpha(\delta^{N,M}) \to v^\alpha(\delta^*), \quad (N,M) \to \infty,$$

and any limit policy is optimal. However, choosing the following parameters: $\lambda = 1$, $\mu_1 = 8.56$, $\mu_2 = 0,28$, $p_1 = 0.22$, $N = M = 300$ leads to the optimal policy in the rectangle shown in the picture below. The blue color stands for server allocation to unit 2.



This optimal policy is very far from being the index policy AP1, although the truncation size seems large enough to exhibit an optimal policy that is 'closer' to AP1. One starts to wonder what the effect of such a straightforward truncation has been on numerical approximations of other models studied in the literature.

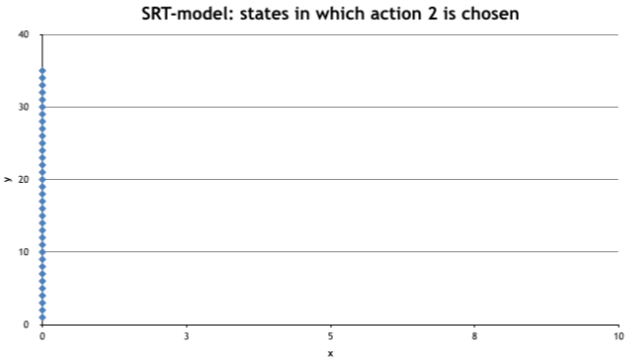
Smoothed Rate Truncation (SRT) SRT is a perturbation introduced in [7], where 'outward bound' probabilities are linearly decreased as a function of state. This creates a perturbed MDP with a finite closed class under any policy. It is not meaningful to try and give a complete definition, but the idea is best illustrated by specifying possible SRT's for the above example. The experience with SRT so far is that it leaves the structure of an optimal policy intact (cf. [7], [6]). On the other hand, since it perturbs transition probabilities from *all* states in the finite closed class, the value function itself is better approximated by a straightforward cut-off, such as the one as described above.

One can apply SRT as follows. Fix N, M. Then we put

$$p_{(x,y)(x',y')}((N,M),\delta_{x,y}) = \begin{cases} \lambda(1-\frac{x}{N})^+, & x'=x+1, y'=y \\ p_1\mu_1\big(1-\frac{y}{M}\big)^+ & x>0, x'=x-1, y'=y+1, \delta_{x,y}=1 \\ (1-p_1)\mu_1 & x>0, x'=x-1, y'=y, \delta_{x,y}=1 \\ \mu_2 & y>0, x'=x, y'=y-1, \delta_{x,y}=2, \end{cases}$$

and $p_{(x,y)(x,y)}((N,M),\delta_{x,y})$ is chosen such as to make the transition probabilities to add up to 1. Again, the function V from (2.8) is a $(\gamma, \mathcal{N}\times\mathcal{D})$-drift function satisfying Assumptions 2.1 and 2.2(α).

The following picture illustrates the numerical results with $N = M = 35$. This confirms the results in [7] and [6], suggesting that SRT allows to obtain information on the structure of an optimal policy for an infinite state MDP.



Here, in both the truncated MDP and the non-truncated one AP1 is optimal. We have not proven this result, but we did so for another SRT (cf. [12]) that is not based on a rectangle but on a triangle. Using a triangular truncation requires less events to be smoothly truncated. This also shows that smooth truncations are not uniquely defined, but different choices are possible.

Fix N. Then we put

$$p_{(x,y)(x',y')}(N,\delta_{x,y}) = \begin{cases} \lambda(1-\frac{x+y}{N})^+, & x'=x+1, y'=y \\ p_1\mu_1 & x>0, x'=x-1, y'=y+1, \delta_{x,y}=1 \\ (1-p_1)\mu_1 & x>0, x'=x-1, y'=y, \delta_{x,y}=1 \\ \mu_2 & y>0, x'=x, y'=y-1, \delta_{x,y}=2, \end{cases}$$

with $p_{(x,y)(x,y)}(N,\delta_{x,y})$ equal to the residual probability mass. Using event based dynamic programming (using special SRT operators), [12] shows that the $c\mu$-rule is optimal, for each N.

Our intuition why this works is as follows. Consider the second SRT. Then, with $u = (x+y)$, $u \mapsto \lambda(1-\frac{u}{N})^+$ is linearly decreasing on the closed class of the Markov chains associated with this SRT for N fixed. How it behaves outside the closed class is not of interest.

In order to show structural properties of policies, we need structural properties of value functions. Often we need a property like convexity (in each variable). The function $u \mapsto \lambda(1-\frac{u}{N})^+$ is convex (and concave) on the non-transient states. However, if we would truncate at N, or linearly decrease the arrival probability starting from $u = N_0 < N$ on, then convexity is lost. The arrival probability as a function of $u = x$ in the first case, and $u = x+y$ in the second, would be given by the functions $u \mapsto \lambda \mathbf{1}_{\{u<N\}}$ and $u \mapsto \lambda(1-\frac{(u-N_0)^+}{N})^+$ respectively. Both functions are not convex!

2.3 Average cost

Establishing a framework for the average cost optimality criterion is more difficult than for the discounted cost case. There are several cautionary examples in the literature highlighting the complications. In our opinion, the most intricate one is the Fisher-Ross example [22]. In this example, all action spaces are finite, the Markov process associated with any stationary deterministic policy is irreducible and has a stationary distribution. However, there is an optimal nonstationary policy, but no stationary deterministic one is optimal.

In this chapter we provide conditions under which there exists a stationary average optimal policy satisfying the average cost optimality equation. The proof requires the vanishing discount approach. Our focus in this paper are non-negative cost functions, the analysis of which does not require heavy drift conditions, imposing positive recurrence of the Markov process associated with any parameter. However, below, we do touch upon benefits of using the heavier conditions.

The conditions that we will focus on, are based on the work of Borkar [15], in the form discussed in [40]. They resemble the conditions from the earlier work in [48]. They imply conditions developed in [38], requiring (i) lower-bounded direct cost; (ii) the existence of a finite α-discounted value function, (iii) the existence of a constant L and a function $M : \boldsymbol{S} \to \mathbf{R}$, such that $-L \leq v_x^\alpha - v_z^\alpha \leq M_x$ for some state z, and all α sufficiently small, and (iv) for each $x \in \boldsymbol{S}$, there exists ϕ_x, such that $\sum_y p_{xy}(\phi_x) M_y < \infty$. The paper [38] proves the statement in Theorem 2.8 below, with equality in (2.9) replaced by inequality, and without property (3). It is appropriate to point out that the approach initiated in [38] was based on a bounded cost average optimality result in [36].

The following definitions are useful. Define $\tau_z := \min_{n \geq 1} \mathbf{1}_{\{z\}}(X_n)$ to denote the hitting time of $z \in \boldsymbol{S}$. Let $m_{xz}(\phi) = \mathsf{E}_x^\phi[\tau_z]$ and $c_{xz}(\phi) = \mathsf{E}_x^\phi\big[\sum_{n=0}^{\tau_z} c_{X_n}(\phi)\big]$.

Assumption 2.3 Let $\Phi' = \prod_x \Phi'_x \subset \Phi$ have the product property. The following holds.

i) Non-negative cost rates: $c_x(\phi) \geq 0$ for all $x \in \boldsymbol{S}, \phi \in \Phi'$.

ii) There exist $z \in \boldsymbol{S}$ and $\phi_0 \in \Phi'$ such that $m_{xz}(\phi_0) < \infty$, $c_{xz}(\phi_0) < \infty$ for all $x \in \boldsymbol{S}$, with the potential exception of z, that is allowed to be absorbing. Note that this implies that $\mathsf{g}_x(\phi_0) =: \mathsf{g}(\phi_0)$ is independent of $x \in \boldsymbol{S}$.

iii) There exists $\epsilon > 0$ such that $D = \{x \in \boldsymbol{S} \,|\, c_x(\phi) \leq \mathsf{g}(\phi_0) + \epsilon \text{ for some } \phi \in \Phi'\}$ is a finite set.

iv) For all $x \in D$ there exists $\phi^x \in \Phi'$ such that $m_{zx}(\phi^x) < \infty$ and $c_{zx}(\phi^x) < \infty$.

Theorem 2.8. *Suppose that Assumptions 2.1, 2.2(α), $\alpha \in (0,1)$, and 2.3 hold. Then the following holds.*

i) *There exists a solution tuple (g^*, v^*), $g^* \in \mathbf{R}_+$, $v^* : \boldsymbol{S} \mapsto \mathbf{R}$, to the average cost optimality equation (DAOE)*

$$g + u_x = \min_{\phi_x \in \Phi'_x} \Big\{ c_x(\phi_x) + \sum_{y \in S} p_{xy}(\phi_x) u_y \Big\}, \tag{2.9}$$

with the property that (1) $g^* = \mathsf{g}$ is the minimum expected average cost (in Φ'), (2) any $\phi^* \in \Phi'$ with

$$\phi_x^* \in \arg \min_{\phi_x \in \Phi'_x} \Big\{ c_x(\phi_x) + \sum_{y \in S} p_{xy}(\phi_x) v_y \Big\}$$

is (average cost) optimal in Φ' and (3) there exists $x^* \in D$ with $v^*_{x^*} = \inf_x v^*_x$.

ii) *Let $x_0 \in \boldsymbol{S}$. Any sequence $\{\alpha_n\}_n$ with $\lim_n \alpha_n = 0$, has a subsequence, again denoted $\{\alpha_n\}_n$, along which the following limits exist:*

$$\begin{aligned} v'_x &= \lim_{n\to\infty} (v_x^{\alpha_n} - v_{x_0}^{\alpha_n}), \quad x \in \boldsymbol{S}, \\ g' &= \lim_{n\to\infty} \alpha_n v_x^{\alpha_n}, \quad x \in \boldsymbol{S} \\ \phi' &= \lim_{n\to\infty} \phi^{\alpha_n}. \end{aligned}$$

Further, the tuple (g', v') is a solution to (2.9) with the properties (1), (2) and (3), so that $g' = \mathsf{g}$. Moreover, ϕ' takes minimising actions in (2.9) for $g = g'$ and $u = v'$.

Theorem 2.8 is a slight extension from [40, Theorem 7.5.6], where the action space is assumed to be finite. Although the various necessary proof parts are scattered over [40, Chapter 7], we will merely indicate the necessary adjustments to allow for the compact parameter case. We would like to note that we use a completely analogous reasoning in the proof of Theorem 3.5, which contains all further details.

Proof. A close examination of the proof of [40, Theorem 7.5.6] shows that the assumption of a finite action space is not necessary. The proof can be adjusted in such a way that the statement holds for a compact action space as well. We briefly discuss the adjustments below. The existence of the limits along a sequence $\{\alpha_n\}_n$, $\alpha_n \downarrow 0$, $n \to \infty$, in assertion *ii)* is a direct result of Sennott [40].

Obtaining the average cost inequality (ACOI) for a limit point of α-discounted optimal policies, as $\alpha \to 0$, can be achieved by virtue of Fatou's lemma. This policy is shown to be optimal.

Further, one needs to show explicitly that there exists a policy realising the infimum of Equation 2.9. Since the limit policy satisfies the ACOI, a similar (very ingenious) reasoning as in the proof of Sennott [40, Theorem 7.4.3] yields that this policy satisfies the DAOE as well. In fact any policy satisfying the ACOI also satisfies the DAOE. It can then be shown by contradiction that this limit policy must attain the infimum. As a consequence, the limit tuple (g', v') from (ii) is a solution to (2.9). The rest directly follows from the proof of the afore mentioned theorem in [40]. □

Remark 2.3 In the literature the formulation of statements like Theorem 2.8 on the DAOE may sometimes have a misleading character. This may occur when the existence of a solution to (2.9) is stated first, and a subsequent claim is made that any minimising policy in (2.9) is average cost optimal. Strictly speaking, this may not be true. Examples 2.3 and 2.4 below, at the end of this section, illustrate that other 'wrong' solutions may exist. Unfortunately, Assumption 2.3 does not admit tools to select the 'right' solution among the set of all solutions. Thus, under Assumption 2.3 a solution to the DAOE should always be obtained via the vanishing discount approach, as in Theorem 2.8 (ii).

The next issue to be discussed is how to verify Assumption 2.3 (ii) and Assumption 2.3 (iv). This can be inferred from the following Lyapunov function criterion, which is a direct application of [26, Lemma 3.1]. The proof is a simple iteration argument.

Lemma 2.9. *Let $x_0 \in \boldsymbol{S}$ be given. Let $\phi \in \Phi$. Suppose that there exist functions $f, h : \boldsymbol{S} \to [0, \infty)$ with*

i) $f_x \geq \max\{1, c_x(\phi)\}$, $x \in \boldsymbol{S} \setminus \{x_0\}$;

ii) $f_x + \sum_{y \neq x_0} p_{xy}(\phi) h_y \leq h_x$, $x \in \boldsymbol{S}$.

Then $m_{xx_0}(\phi), c_{xx_0}(\phi) \leq h_x$, $x \in \boldsymbol{S}$.

To pave the way for developing a roadmap for obtaining structures of average cost optimal policies, we will shortly discuss the applicability of VI. Let us first state the algorithm. Again assume that $\Phi' \subset \Phi$ has the product property.

Algorithm 2 VI for an expected average cost optimal policy

1. Select v^0, set $n = 0$.
2. For each $x \in \boldsymbol{S}$, compute v_x^{n+1} by
$$v_x^{n+1} = \min_{\phi_x \in \Phi'_x} \Big\{ c_x(\phi_x) + \sum_{y \in S} p_{xy}(\phi_x) v_y^n \Big\},$$
and let
$$\phi^{n+1} \in \arg\min_{\phi \in \Phi'} \{ c(\phi) + P(\phi) v^n \}.$$
3. Increment n by 1 and return to step 2.

To our knowledge there are relatively few non-problem specific papers on the convergence of average cost VI for countable state space MDPs, cf. [27], [39], [4], and [2], the latter of which is based on the thesis [42]. The conditions in the first three papers are not restricted to conditions on the input parameters. In our opinion, the easiest verifiable ones are contained in the third [4], involving properties of the set of policies $\{\phi_n\}_n$. In case of well-structured problems, say ϕ_n are all equal, or have very specific structures, these conditions are easily be verifiable.[1] Here, we will limit to the conditions from [42] and [2] that are, as far as we know, the only ones formulated directly in terms of the input parameters of the process. The notation $\mathbf{e}$ stands for the function on $\boldsymbol{S}$ identically equal to 1.

Theorem 2.10. *Suppose that the following drift condition, called* V-geometric recurrence, *holds: there exist a function* $V : \boldsymbol{S} \to [1, \infty)$, *a finite set* $M \subset \boldsymbol{S}$ *and a constant* $\beta < 1$, *such that*
$$\sum_{y \notin M} p_{xy}(\phi) V_y \leq \beta V_x, \quad x \in \boldsymbol{S}, \phi \in \Phi'.$$

Suppose further that the following holds as well:

- *Assumption 2.1;*
- $\sup_{\phi \in \Phi'} \|c(\phi)\|_V < \infty$;
- $\phi \mapsto P(\phi)V$ *is componentwise continuous on* Φ';
- *the Markov process with transition matrix* $P(\phi)$ *is aperiodic and has one closed class,* $\phi \in \Phi'$.

Let $0 \in \boldsymbol{S}$. *There is a unique solution pair* (g^*, v^*) *with* $v^* \in \ell^\infty(\boldsymbol{S}, V)$, *and* $v_0^* = 0$, *to (2.9) with the properties in Theorem 2.8.*

Furthermore, average cost VI converges, in other words, $\lim_{n\to\infty}(v^n - v_0^n \mathbf{e}) \to v^*$, *and any limit point of the sequence* $\{\phi^n\}_n$ *is average cost optimal and a minimising policy in the DAOE (2.9) with solution tuple* (g^*, v^*).

[1] we still have a suspicion that there is a gap in the proofs in [4]

The V-uniform geometric recurrence condition in Theorem 2.10 has been introduced in [17], and shown in [18] to imply the assertion in Theorem 2.10. The paper [18], see also [42], has derived an equivalence of this condition (under extra continuity conditions) with V-uniform geometric ergodicity. The thesis [42] shows a similar implication for bounded jump Markov decision processes in continuous time, by uniformisation. Both properties have been extensively used both in the case of a parameter space consisting of one element only (cf. [31] and later works), and in the case of product parameter spaces in the context of optimal control. Together with the negative dynamic programming conditions developed by [38], the V-uniform geometric recurrence and ergodicity, developed in [16] and [17], have become 'standard' conditions in many papers and books. See for instance [24], [23], and [34], as well as references therein, for a survey and two books using both types of conditions. A parametrised version of [18] in both discrete and continuous time is currently in preparation.

The drawback of using V-geometric recurrence is its implying each associated Markov process to be positive recurrent. This is a major disadvantage for many models and therefore our motivation for using Assumption 2.3. Note that customer abandonment has a strong stabilising effect on the associated Markov processes, and then V-geometric recurrence typically may apply.

Roadmap to structural properties Below we formulate a scheme for deriving the structure of an optimal policy and value function, if the optimisation criterion is to minimise the expected average cost. Let $\Phi' = \Phi = \mathcal{D}$ be the set of all stationary, deterministic policies.

Roadmap for average cost MDPs in discrete time

1. Check the conditions of Theorem 2.10.
   If satisfied then:
   - perform VI Algorithm 2.
2. If not satisfied, then check Assumptions 2.1, 2.2 (α), for all $\alpha \in (0,1)$, and 2.3. If satisfied then:
   a) perform VI Algorithm 1 for the α-discounted cost criterion. If there exists $\alpha_0 > 0$ such that the desired structural properties hold for all $\alpha \in (0, \alpha_0)$ then
   b) apply the vanishing discount approach by taking the limit $\alpha \to 0$. This is justified by Theorem 2.8.
3. If not satisfied, or if no structural properties are concluded, then the outcome is inconclusive.

Note that the vanishing discount approach has the advantage of allowing a conclusion on Blackwell optimality of the limiting policy. Next we provide examples showing that the DAOE may have more than one solution.

Example 2.3. Consider a simple random walk on the state space $\boldsymbol{S} = \mathbf{Z}$ without any control. Thus Φ consists of one element ϕ, say $\phi_x = 1$ for all $x \in \boldsymbol{S}$. The transition mechanism is given by

$$p_{xy}(1) = \begin{cases} \lambda, & y = x+1 \\ \mu, & y = x-1, \end{cases}$$

where $\lambda < \mu$ and $\lambda + \mu = 1$. The cost in state $x \neq 0$ is equal to $c_x = 1$, and $c_0 = 0$. This is a transient Markov process, and hence it does not satisfy Assumption 2.3. However, it does satisfy the assumptions of [38], implying the assertion of Theorem 2.8 to hold.

This implies that the vanishing discount approach yields solution tuple $g = 1$ and

$$v_x = \begin{cases} \frac{1-(\mu/\lambda)^x}{\mu-\lambda}, & x < 0 \\ 0, & x \geq 0, \end{cases}$$

if $x_0 = 0$ is chosen. This can be deduced from boundedness conditions that will be discussed in a forthcoming paper [44].

However, other solutions $(g = 1, v')$ exist, namely for any $\theta \in \mathbf{R}$

$$v'_x = \begin{cases} (1-\theta)\frac{1-(\mu/\lambda)^x}{\mu-\lambda}, & x < 0 \\ 0, & x = 0 \\ \theta\frac{(\mu/\lambda)^x-1}{\mu-\lambda}, & x > 0. \end{cases}$$

There is no a priori tool to determine which solution is the one obtained from the vanishing discount approach.

Example 2.4. Next we restrict the simple random walk to $\boldsymbol{S} = \mathbf{Z}_+$, and associate the corresponding transitions with ϕ^1. In other words, putting $\phi^1_x = 1$, $x \in \boldsymbol{S}$, gives

$$p_{xy}(1) = \begin{cases} \lambda, & y = x+1 \\ \mu, & y = (x-1)^+, \end{cases}$$

where $\lambda < \mu$ and $\lambda + \mu = 1$. Suppose that holding cost x is incurred per (discrete) unit time, when the number of customers in the system is x, and action 1 is used:

$$c_x(\phi^1_x) = c_x(1) = x, \quad x \in \boldsymbol{S}.$$

In [8] it was shown that the equation $v + g = c(\phi^1) + P(\phi^1)v$ has the following solutions: to any $g \in \mathbf{R}$ there is a solution tuple (g, v_g) with $v^g : \boldsymbol{S} \to \mathbf{R}$ the function given by

$$v^g_x = -\frac{x-1}{\mu-\lambda}g + \frac{(x-1)(x-2)}{2(\mu-\lambda)} + \mu\frac{x-1}{(\mu-\lambda)^2} + \frac{g}{\lambda} + \mu\frac{(\mu/\lambda)^{x-1}-1}{\mu-\lambda}\Big\{\frac{g}{\mu-\lambda} + \frac{g}{\lambda} - \frac{\mu}{(\mu-\lambda)^2}\Big\}, \quad (2.10)$$

for $x \geq 1$ and $v^g_0 = 0$. The solution obtained from a vanishing discount approach, is the one for which the expression between curly brackets is 0, i.e. for which

$$\frac{g}{\mu-\lambda} + \frac{g}{\lambda} - \frac{\mu}{(\mu-\lambda)^2} = 0,$$

in other words

$$g = \frac{\lambda}{\mu-\lambda},$$

and

$$v^g = -\frac{x-1}{\mu-\lambda}g + \frac{(x-1)(x-2)}{2(\mu-\lambda)} + \mu\frac{x-1}{(\mu-\lambda)^2} + \frac{g}{\lambda}.$$

This can also be derived from boundedness conditions analysed in [9]. Thus, $\mathsf{g}(\phi^1) = \lambda/(\mu-\lambda)$.

Next, in state 1 there is a further option to choose action 2, with corresponding transition probabilities

$$p_{1,3}(2) = \lambda = 1 - p_{1,0}(2).$$

This yields parameter ϕ^2, with $\phi^2_1 = 2$, and $\phi^2_x = 1$, $x \neq 1$. The corresponding cost $c_1(2)$ is chosen small enough (possibly negative) so that $\mathsf{g}(\phi^2) < \mathsf{g}(\phi^1)$. This yields an MDP satisfying Assumption 2.3. Although the direct cost possibly is not non-negative, it is bounded below.

Next, we claim that we may choose a constant g in (2.10) so large that

$$v_1^g + g = c_1(1) + \sum_y p_{1y}(1)v_y^g < c_1(2) + \sum_y p_{1y}(2)v_y^g = c_x(2) + \lambda v_3^g + \mu v_0^g,$$

in other words, the minimisation prescribes to choose action 1 in state 1. Indeed, this choice is possible if

$$c_x(2) + \lambda v_3^g > 1 + \lambda v_2^g,$$

or

$$1 - c_x(2) < \lambda(v_3^g - v_2^g). \tag{2.11}$$

It can be checked that

$$v_3^g - v_2^g > \frac{\mu^2}{\lambda^2}\Big(\frac{g}{\lambda} - \frac{\mu}{(\mu-\lambda)^2}\Big).$$

Therefore, one may choose $g > \mathsf{g}(\phi^1)$ large enough for (2.11) to be true. Hence (g, v^g) is a solution to (2.9) for the MDP with minimising policy ϕ^1. However, by construction $g > \mathsf{g}(\phi^1) > \mathsf{g}(\phi^2)$. Thus, (g, v^g) is a solution to the DAOE, where g is not the minimum expected average cost, and the policy choosing minimising action is not the optimal policy.

3 Continuous time parametrised Markov processes

In this section we will consider continuous time parametrised Markov processes. The setup is analogous to the discrete time case. Again we consider a parameter space Φ and a countable state space $\boldsymbol{S}$. With each $\phi \in \Phi$ we associate an $\boldsymbol{S} \times \boldsymbol{S}$ generator matrix or q-matrix $Q(\phi)$ and a cost rate vector $c(\phi) : \boldsymbol{S} \to \mathbf{R}$.

Following the construction in [28], see also [32], one can define a basic measurable space $(\Omega, \mathcal{F})$, a stochastic process $X : \Omega \to \{f : [0,\infty) \to \boldsymbol{S} \,|\, f \text{ right-continuous}\}$, a filtration $\{\mathcal{F}_t\}_t \subset \mathcal{F}$ to which X is adapted, and a probability distribution P_ν^ϕ on $(\Omega, \mathcal{F})$, such that X is the minimal Markov process with q-matrix $Q(\phi)$, for each initial distribution ν on $\boldsymbol{S}$ and $\phi \in \Phi$. Denote by $P(\phi) = \{\, p_{t,xy}(\phi)\}_{x,y\in S},\ t \geq 0$, the corresponding minimal transition function and by E_ν^ϕ the expectation operator corresponding to P_ν^ϕ.

Assumption 3.1 **i)** $Q(\phi)$ is a a conservative, stable q-matrix, i.e. for $x \in \boldsymbol{S}$ and $\phi \in \Phi$

- $0 \leq q_x(\phi) = -q_{xx}(\phi) < \infty$;
- $\sum_y q_{xy}(\phi) = 0$.

ii) $\{P_t(\phi)\}_{t\geq 0}$ is standard, i.e. $\lim_{t\downarrow 0} p_{t,xy}(\phi) = \delta_{xy}$, with δ_{xy} the Kronecker delta.

iii) $\phi \mapsto q_{xy}(\phi)$ and $\phi \mapsto c_x(\phi)$ are continuous, $x, y \in \boldsymbol{S}$;

iv) Φ is locally compact.

Let $\Phi' \subset \Phi$. The definition of the product property of $\{Q(\phi)\}_\phi$ and $\{c(\phi)\}_\phi$ with respect to Φ' is completely analogous to Definition 2.1. This entails Φ' to be compact in the product topology. Again, for easy reference, we say that Φ' has the product property if $\{Q(\phi)\}_\phi$ and $\{c(\phi)\}_\phi$ both have the product property with respect to Φ'. If Φ has the product property, then the parametrised Markov process is an MDP. Analogously to Remark 2.1, Φ may represent the collection of stationary policies or the stationary, deterministic ones.

Suppose furthermore, that a lump cost is charged, in addition to a cost rate incurred per unit time. Say at the moment of a jump x to y lump cost $d_{xy}(\phi)$ is charged, when the parameter is ϕ. This can be modelled as a (marginal) cost rate $c_x(\phi) = \sum_{y \neq x} d_{xy}(\phi) q_{xy}(\phi)$.

Below we give the definitions of various performance measures and optimality criteria. Lateron we will provide conditions under which these exist.

For $\alpha > 0$, under parameter $\phi \in \Phi$ the *expected total α-discounted cost value function* v^α is given by

$$v_x^\alpha(\phi) = \mathsf{E}_x^\phi\Big[\int_{t=0}^{\infty} e^{-\alpha t} c_{X_t} dt\Big], \quad x \in \boldsymbol{S}. \tag{3.1}$$

Suppose that $\Phi' \subset \Phi$ has the product property. The *minimum expected total α-discounted cost* w.r.t Φ' is defined

$$v_x^\alpha = \inf_{\phi \in \Phi'} \{v_x^\alpha(\phi)\}, \quad x \in \boldsymbol{S}. \tag{3.2}$$

If $v^\alpha(\phi) = v^\alpha$, then ϕ is said to be optimal in Φ'.

The *expected average cost* under parameter ϕ is given by

$$\mathsf{g}_x(\phi) = \limsup_{T \to \infty} \frac{1}{T} \mathsf{E}_x^\phi\Big[\int_{t=0}^{T} c_{X_t} dt\Big], \quad x \in \boldsymbol{S}. \tag{3.3}$$

Suppose that $\Phi' \subset \Phi$ has the product property. The *minimum expected average cost* is defined as

$$\mathsf{g}_x = \inf_{\phi \in \Phi'} \{\mathsf{g}_x(\phi)\}, \quad x \in \boldsymbol{S}. \tag{3.4}$$

If $\mathsf{g}(\phi) = \mathsf{g}$ for some $\phi \in \Phi'$ then ϕ is said to be average cost optimal in Φ'.

The notions of *Blackwell optimality* and *strong Blackwell optimality* are defined completely analogously to the discrete time versions.

A well-known procedure to determine the structure of an optimal policy in the continuous time case, is to reduce the continuous time MDP to a discrete time MDP in order to be able to apply VI. There are different time-discretisation methods. One is to consider the embedded jump process. Sometimes this is a viable method, see [25] where this approach has been taken. In Section 3.4 we give an example where the embedded jump approach seems to be less amenable to apply.

Instead, one may use uniformisation. However, applicability hinges on models, where the jumps are bounded as a function of parameter and state:

$$q := \sup_{x \in S, \phi \in \Phi} q_x(\phi) < \infty. \tag{3.5}$$

This property is violated in models with reneging customers, population models etc, and we will consider how to handle this next.

Let us first recall the uniformisation procedure.

3.1 Uniformisation

A detailed account of the uniformisation procedure and proofs can be found in [41]. If a continuous time parametrised Markov process has bounded transition rates (cf. (3.5)), it admits a transformation to an equivalent discrete time parametrised Markov process. Below we list the transformations for the α-discounted and average cost cases.

For the discounted cost criterion the equivalent discrete time process is given by

$$P(\phi) = I + \frac{1}{q} Q(\phi), \qquad c^d(\phi) = \frac{c(\phi)}{\alpha + q}, \qquad \alpha^d = \frac{\alpha}{\alpha + q}, \qquad \phi \in \Phi. \tag{3.6}$$

Denote the discrete time α^d-discounted cost under policy ϕ as $v^{d,\alpha^d}(\phi)$. Both the discrete-time and continuous time processes have equal expected cost, i.e. $v^{d,\alpha^d}(\phi) = v^{\alpha}(\phi)$. If $\Phi' \subset \Phi$ has the product property, then this implies that the optimal α- and α^d-discounted value functions with respect to Φ' are equal:

$$v^{d,\alpha^d} = v^{\alpha}.$$

For the average cost criterion the equivalent discrete time process is given by

$$P(\phi) = I + \frac{1}{q}Q(\phi), \qquad c^d(\phi) = \frac{c(\phi)}{q}, \qquad \phi \in \Phi.$$

Denote the discrete time average cost under parameter ϕ as $\mathsf{g}^d(\phi)$ and the value function as $v^d(\phi)$. Under the same parameter, the discrete-time and continuous time expected cost, relate to each other as follows

$$q\mathsf{g}^d(\phi) = \mathsf{g}(\phi).$$

The corresponding value functions are identical:

$$v^d(\phi) = v(\phi).$$

These relations apply similarly to optimal parameters in a product set $\Phi' \subset \Phi$.

The main concern is how to proceed in the case of unbounded jump rates: $q = \infty$.

3.2 Discounted cost

First we treat the discounted cost criterion. This section summarises the results of [13]. That paper only treats optimality within the class of stationary Markov policies, as we do in the present paper. We recall some definitions. These definitions are closely related to the conditions used in the discrete time analysis in Section 2.1.

Definition 3.1

- The function $W : \boldsymbol{S} \to (0, \infty)$ 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 $\inf_{x \notin K_n} W_x \to \infty$, as $n \to \infty$.
- The function $V : \boldsymbol{S} \to (0, \infty)$ is called a (γ, Φ)-drift function if $Q(\phi)V \leq \gamma V$ for all $\phi \in \Phi$, where $QV_x := \sum_{y \in S} q_{xy} V_y$.

Assumption 3.2 ($\boldsymbol{\alpha}$) i) There exist a constant $\gamma < \alpha$ and function $V : \boldsymbol{S} \to (0, \infty)$ such that V is a (γ, Φ)-drift function;

ii) $\sup_\phi \|c(\phi)\|_V =: c_V < \infty$ for all $\phi \in \Phi$;

iii) There exist a constant θ and a function $W : \boldsymbol{S} \to (0, \infty)$ such that W is a (θ, Φ)-drift function and W/V is a moment function, where $(W/V)_x = W_x/V_x$, $x \in \boldsymbol{S}$.

Assumptions 3.2 (α) (i) and 3.2 (α)(ii) are the continuous time counterpart of Assumption 2.2 (α). Assumption 3.2(α)(iii) is sufficient to guarantee nonexplosiveness of the parametrised Markov process (cf. [43, Theorem 2.1]), and implies continuity properties of the map $\phi \mapsto (P_t(\phi)V)_x$, $x \in \boldsymbol{S}$.

Theorem 3.2 ([13, Theorem 4.1]). *Suppose Assumptions 3.1 and 3.2 (α) hold, then $\phi \mapsto v^{\alpha}(\phi)$ is component-wise continuous and $v^{\alpha}(\phi)$ is the unique solution in $\ell^{\infty}(\boldsymbol{S}, V)$ to*

$$\alpha u = c(\phi) + Q(\phi)u. \tag{3.7}$$

Theorem 3.3 ([13, Theorem 4.2]). *Assume that $\Phi' \subset \Phi$ has the product property. Suppose further that Assumptions 3.1, and 3.2(α) hold. Then v^α is the unique solution in $\ell^\infty(\boldsymbol{S}, V)$ to the α-discount optimality equation (CDOE)*

$$\alpha u_x = \inf_{\phi_x \in \Phi'_x} \{c_x(a_x) + \sum_y q_{xy}(\phi_x) u_y\}, \quad x \in \boldsymbol{S}. \tag{3.8}$$

There exists $\phi^* \in \Phi'$ with $\phi^*_x \in \arg\min_{\phi_x \in \Phi'_x} \{c_x(a_x) + \sum_y q_{xy}(\phi_x) u_y\}$, $x \in \boldsymbol{S}$. Any policy $\phi' \in \Phi'$ that minimises (3.8) is optimal in Φ', and it holds that $v^\alpha(\phi') = v^\alpha$.

As discussed in Section 2.2, the parameter set may contain a perturbation component. Introducing a perturbation yields a parameter set of the following form $\Phi = \mathcal{N} \times \mathcal{D}$, where N is a perturbation parameter and $\mathcal{D}$ the set of deterministic stationary (or merely stationary) policies.

Corollary 3.4 (cf. [13, Theorem 5.1]). *Suppose that Assumptions 3.1 and 3.2 (α) hold. Let $\Phi = \mathcal{N} \times \mathcal{D}$. Assume that $\{N\} \times \mathcal{D}$ has the product property for $N \in \mathcal{N}$. Then,*

i) $\lim_{N \to N_0} v^\alpha(N) = v^\alpha(N_0)$;

ii) *any limit point of $(\delta^*_N)_{N \to N_0}$ is optimal in $\{N_0\} \times \mathcal{D}$.*

iii) *Suppose that the MDP with parameter set $\{N\} \times \mathcal{D}$ is uniformisable, i.e.*

$$q^N := \sup_{x \in S, \delta \in \mathcal{D}} |q_{xx}(N, \delta)| < \infty.$$

Consider the discount discrete-time uniformised MDP, with transition matrices, cost and discount factor given by (cf. (3.6))

$$P^(N, \delta) = I + \frac{1}{q^N} Q(N, \delta), \quad c(N, \delta) = \frac{c(N, \delta)}{\alpha + q^N}, \quad \alpha^d = \frac{\alpha}{\alpha + q^N}.$$

Then the MDP satisfies Assumptions 2.1 and 2.2(α^d), for the same function V.

Proof. Assertions (i) and (ii) are in fact [13, Theorem 5.1], but they follow easily from Theorems 3.2 and 3.3. Assertion (iii) is a direct verification. □

Roadmap to structural properties We finally have collected the tools to provide a scheme for the derivation of structural properties of an optimal policy and value function for a continuous time MDP with unbounded jump rates, provided the required conditions hold. Applications of this scheme are discussed in [13], and [14].

Let $\Phi' = \mathcal{D}$ be the set of stationary, deterministic policies, and $\Phi = \mathcal{N} \times \mathcal{D}$. Each set $\{N\} \times \mathcal{D}$ is assumed to have the product property, $N \in \mathcal{N}$.

Roadmap for α-discounted MDPs in continuous time

1. If Assumptions 3.1 and 3.2(α) hold, and $q < \infty$ do
   - **a)** perform a uniformisation;
   - **b)** use VI Algorithm 1 to verify the structural properties of an optimal policy and value function;

c) using the equivalence of uniformised and non-uniformised systems yields the structure of an optimal policy and value function of the non-uniformised continuous time MDP.

2. If Assumptions 3.1 and 3.2(α) hold, and $q = \infty$ do

   i) perform a bounded jump perturbation leaving the structural properties intact and satisfying Assumptions 3.1 and 3.2(α). For instance, one might apply SRT (see Section 2.2) or try a brute force perturbation;

   ii) do steps a,b,c. This potentially yields structural properties of an optimal policy and the value function for each N-perturbed MDP;

   iii) take the limit for the perturbation parameter to vanish. Corollary 3.4 gives the structural results for an optimal policy and value function.

3. If the assumptions do not hold, or if no structural properties can be concluded, then the outcome is inconclusive.

As has been mentioned already, one might apply discounted VI directly to the associated discrete time MDP, embedded on the jumps of the continuous time MDP (cf. e.g. [23, Theorem 4.12]). In the example of Section 3.4 we discuss some problems with the application of this procedure.

3.3 Average cost

We finally turn to studying the average cost criterion in continuous time. The assumptions that we make are Assumption 3.1 and the analogon to Assumption 2.3 that we used in Section 2.3 for analysing the average cost criterion in discrete time. In fact, Assumption 2.3 can be used unaltered. However, one has to use the continuous time definitions of the hitting time of a state, and total expected cost incurred till the hitting time.

The hitting time τ_z of a state $z \in \boldsymbol{S}$ is defined by:

$$\tau_z = \inf_{t>0}\{X_t = z, \exists s \in (0,t) \text{ such that } X_s \neq z\}. \tag{3.9}$$

Then, $m_{xz}(\phi) = \mathsf{E}_x^\phi \tau_z$ and $c_{xz} = \mathsf{E}_x^\phi \int_0^{\tau_z} c_{X_t} dt$, where either expression may be infinite.

The following theorem is completely analogous to the discrete time equivalent, with the only difference that the CAOE has a slightly different form.

Theorem 3.5. *Suppose Assumptions 3.1, 3.2(α), $\alpha > 0$, and 2.3 hold.*

i) *There exists a solution tuple (g^*, v^*) to the average cost optimality equation (CAOE)*

$$g = \min_{\phi_x \in \Phi'_x} \{c_x(\phi_x) + \sum_{y \in S} q_{xy}(\phi) u_y\}, \tag{3.10}$$

with the property that (1) $g^* = \mathsf{g}$ is the minimum expected average cost (in Φ'), (2) $\phi^* \in \Phi'$ with

$$\phi_x^* \in \arg\min_{\phi_x \in \Phi'_x} \{c_x(\phi_x) + \sum_{y \in S} q_{xy}(\phi_x) u_y\}$$

is (average cost) optimal in Φ', and (3) there exists $x^* \in D$ with $v^*_{x^*} = \inf_x v^*_x$.

ii) *Let $x_0 \in \boldsymbol{S}$. Any sequence $\{\alpha_n\}_n$ with $\lim_{n\to\infty} \alpha_n = 0$, has a subsequence, again denoted $\{\alpha_n\}_n$, along which the following limits exist:*

$$\begin{aligned} v'_x &= \lim_{n\to\infty} \{v_x^{\alpha_n} - v_{x_0}^{\alpha_n}\}, \quad x \in \boldsymbol{S}, \\ g' &= \lim_{n\to\infty} \alpha_n v_x^{\alpha_n}, \quad x \in \boldsymbol{S}, \\ \phi' &= \lim_{n\to\infty} \phi^{\alpha_n}. \end{aligned}$$

Furthermore, the tuple (g', v') is a solution to (3.10) with the properties (1), (2) and (3), so that $g' = \mathtt{g}$. Moreover, ϕ' takes minimising actions in (3.10) for $g = g'$ and $v = v'$.

We have not encountered the above result in this form. However, the derivations are analogous to the discrete time variant, cf. [40, Chapter 7], and to the proofs in [25], where continuous time variants of Sennott's discrete time conditions have been assumed. In fact, Assumption 2.3 implies [25, Assumption 5.4]. Although one could piece together the proof of Theorem 3.5 from these references, we prefer to give it explicitly in Appendix A.

For the verification of the validity, one may use the following lemma, that is analogous to Lemma 2.9. The proof is similar to the proof of [47, Theorem 1].

Lemma 3.6. *Let $x_0 \in \boldsymbol{S}$ be given. Let $\phi \in \Phi'$.*

Suppose that there exist functions $f, h : \boldsymbol{S} \to [0, \infty)$ with

i) $f_x \geq \max\{1, c_x(\phi)\}$, $x \in \boldsymbol{S} \setminus \{x_0\}$;

ii) $f_x + \sum_{\substack{y: y\neq x_0 \\ \text{if } x\neq x_0}} q_{xy}(\phi) h_y \leq 0$, $x \in \boldsymbol{S}$.

Then $m_{xx_0}(\phi), c_{xx_0}(\phi) \leq h_x$, $x \in \boldsymbol{S}$.

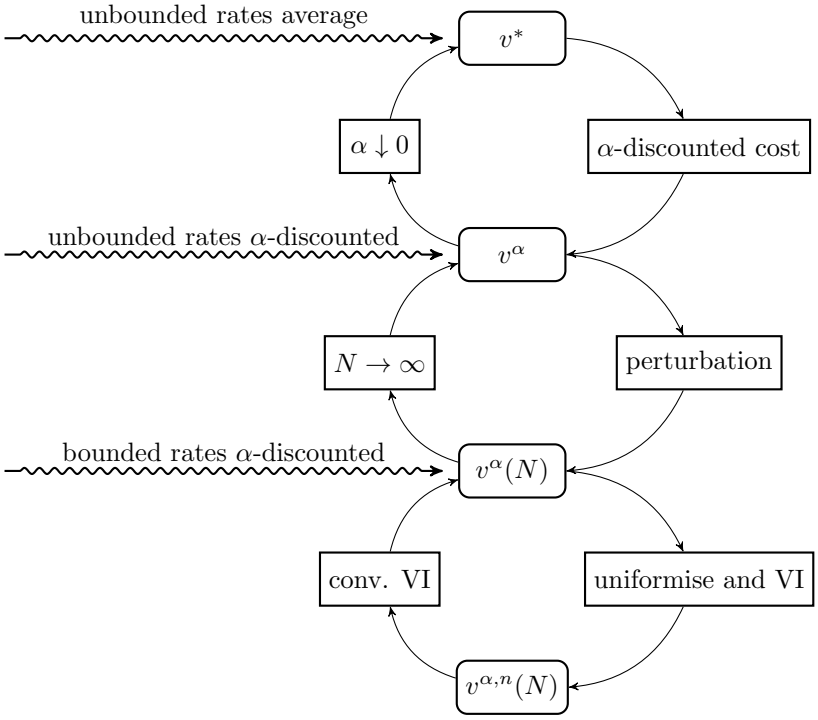
3.4 Roadmap to structural properties

First we present a roadmap for determining structural properties of average cost MDPs in continuous time. We illustrate it with a simple example. More complicated examples can be found in [6] and [14]. Then a table will summarise the schematic approach that we have presented through the various roadmaps, including references to the required conditions and results.

Let $\Phi' = \mathcal{D}$ be the set of stationary, deterministic policies, and $\Phi = \mathcal{N} \times \mathcal{D}$. Assume that $\{N\} \times \mathcal{D}$ has the product property for $N \in \mathcal{N}$.

Roadmap for average cost MDPs in continuous time

1. If Assumptions 3.1, 3.2(α), for all $\alpha > 0$, and 2.3 hold then do
   - apply the roadmap for α-discounted MDPs in continuous time; if the outcome is that the α-discounted problem has the desired structural properties for all $0 < \alpha < \alpha_0$, for some $\alpha_0 > 0$, then do
     - apply the vanishing discount approach of Theorem 3.5 (ii).
2. If the assumptions do not hold, or structural properties can not be shown, the outcome is inconclusive.



The picture uses the following, so far not explained, notation. If the perturbation parameter is N, then the α-discounted value function is denoted by $v^\alpha(N)$. Applying discounted cost VI to the N-perturbation, yield the iterates $v^{\alpha,n}(N)$, $n = 0, \ldots$. The limit as the perturbation parameter vanishes is represented by $N \to \infty$.

Arrival control of the M/M/1+M-queue As an application of this final roadmap, we consider arrival control of the M/M/1+M-queue. Customers arrive in a single server unit with infinite buffer size according to a Poisson (λ) process. Each customer requires an exponentially distributed service time with parameter μ, but he may also renege after an exponentially distributed amount of time with parameter β (service is not exempted from reneging). Arrival process, service times and reneging times are all independent.

Due to reneging, the process associated with the number of customers in the server unit is ergodic at exponential rate. However, this is not desirable from a customer service point of view. Therefore, the following arrival control is exercised. Per unit time and per customer a holding cost of size 1 is incurred. The controller can decide to accept (decision A) or reject (decision R) an arriving customer. If he takes decision A, then a lump reward of size K is incurred.

The goal is to select the control policy with minimum expected average cost. We wish to show that a control-limit acceptance policy is optimal. In other words, that there exists $x^* \in \boldsymbol{S}$, such that accepting in state $x \leq x^*$ and rejecting in state $x > x^*$ is average cost optimal.

This leads to the following MDP on the state space $\boldsymbol{S} = \mathbf{Z}_+$, where state x corresponds to x customers being present in the system. The collection of stationary, deterministic policies is given by $\mathcal{D} = \{\mathsf{A}, \mathsf{R}\}^\infty$. The transition rates are as follows: for $x \in \boldsymbol{S}$

$$q_{xy}(\mathsf{A}) = \begin{cases} \lambda, & y = x+1 \\ \mu \mathbf{1}_{\{x>0\}} + x\beta, & y = x-1 \\ -(\lambda + \mu \mathbf{1}_{\{x>0\}} + x\beta), & y = x, \end{cases} \qquad q_{xy}(\mathsf{R}) = \begin{cases} \mu \mathbf{1}_{\{x>0\}} + x\beta, & y = x-1 \\ -(\mu \mathbf{1}_{\{x>0\}} + x\beta), & y = x. \end{cases}$$

The lump reward can be modelled as a cost rate, and we get for $x \in \boldsymbol{S}$

$$c_x(\mathsf{A}) = x - \lambda K, \qquad c_x(\mathsf{R}) = x.$$

This is an unbounded rate MDP. Denote the never accepting policy by δ_0, then this generates a Markov process with absorbing state 0, and finite expected average cost $\mathsf{g}(\delta_0) = 0$. One can check for $f_x = x$, $x \in \boldsymbol{S}$, that $h_x = e^{\theta x}$, $x \in \boldsymbol{S}$, satisfies Lemma 3.6 (ii), if we choose $\theta > \ln(1 + \beta^{-1})$. Let $\epsilon > 0$. It then follows that Assumption 2.3 is satisfied with set $D = \{x \,|\, x - \lambda K \leq 0 + \epsilon\}$.

It is not difficult to verify that Assumptions 3.1 and 3.2 (α), $\alpha > 0$, are satisfied. Indeed, for given $\alpha > 0$, there exists $\kappa_\alpha > 0$, such that $V_x^\alpha = e^{\kappa_\alpha x}$, $x \in \boldsymbol{S}$, is a $(\gamma_\alpha, \mathcal{D})$-drift function, for some $\gamma_\alpha > 0$.

It follows that there exists a solution tuple (g^*, v^*) of the CAOE (3.10) with the properties (1), (2), (3). This CAOE takes the form

$$g^* = x + \big(\mu \mathbf{1}_{\{x>0\}} + (x \wedge N)\beta\big) v^*_{x-1} - \big(\lambda + \mu \mathbf{1}_{\{x>0\}} + (x \wedge N)\beta\big) v^*_x + \lambda \min\{-K + v^*_{x+1}, v^*_x\},$$

where we have already rearranged the terms in such a way that the equation is amenable to analysis. It is easily deduced that it is optimal to accept in state x if

$$v^*_{x+1} - v^*_x \leq K.$$

Hence, in order that a control-limit acceptance policy be average cost optimal, it is sufficient to show that v^* is convex.

To this end, we will use the roadmap to show that there exists a solution pair (g^*, v^*) to the CAOE (3.10) with properties (1), (2) and (3), with v^* a convex function. Theorem 3.5 justifies using the vanishing discount approach, and so it is sufficient to show convexity of the α-discount value function v^α, for all $\alpha > 0$ sufficiently small. Note that the imposed conditions for the roadmap for α-discount MDPs are satisfied, since these are imposed as well for the assertions in Theorem 3.5, and these have been checked to hold.

The roadmap for the verification of structural properties of v^α prescribes to choose suitable perturbations. We consider a simple perturbation, where the reneging rates are truncated at $N\beta$ in states $x \geq N$, $N \geq 1$. The value $N = \infty$ then corresponds to the original MDP. Thus, $q_{xy}(N, \delta_x) = q_{Ny}(\delta_x)$, for $x \geq N$. A simple verification implies for $\Phi = \{1, 2, \ldots, \infty\} \times \mathcal{D}$, that V^α is a (γ_α, Φ)-drift function and that Assumptions 3.1 and 3.2 (α) are satisfied, $\alpha > 0$, for this extended parameter space.

Fix $\alpha > 0$ and $N \in \{1, 2, \ldots\}$. By virtue of Corollary 3.4 it is sufficient to check convexity of the α-discount value function $v^\alpha(N)$, for the N-perturbation. Finally, by Theorem 2.6 it is sufficient to check convexity of $v^{\alpha,n}(N)$, which is the n-horizon approximation of $v^\alpha(N)$. Convexity of $v^{\alpha,n}(N)$ follows iteratively by putting $v^{\alpha,0}(N) \equiv 0$, and checking that convexity is propagated through the iteration step: for $x \in \boldsymbol{S}$

$$\begin{aligned} v_x^{\alpha,n+1}(N) \;=\; & x - \alpha\big(\mu \mathbf{1}_{\{x>0\}} + (x \wedge N)\beta\big) v^{\alpha,n}_{x-1}(N) + \alpha(1 - \lambda - \mu \mathbf{1}_{\{x>0\}} + (x \wedge N)\beta) v^{\alpha,n}_x(N) \\ & + \lambda\alpha \min\{-K + v^{\alpha,n}_{x+1}(N), v^{\alpha,n}_x(N)\}. \end{aligned} \tag{3.11}$$

Even based dynamic programming [29] applied to (3.11) yields precisely the propagation of convexity.

Associated embedded jump MDP Instead of introducing a perturbation, we could have applied discounted VI to the associated α-discounted embedded jump MDP. The assumptions that we have

made, imply convergence to the α-discounted value function (cf. [23, Theorem 4.14]). This yields the following VI-scheme:

$$\bar{v}_x^{\alpha,n+1} = \min\Big[\frac{1}{\alpha+\lambda+\mu\mathbf{1}_{\{x>0\}}+x\beta}\big\{x-\lambda K+\lambda\bar{v}_{x+1}^{\alpha,n}+(\mu\mathbf{1}_{\{x>0\}}+x\beta)\bar{v}_{x-1}^{\alpha,n}\big\},\\ \frac{1}{\alpha+\mu\mathbf{1}_{\{x>0\}}+x\beta}\big\{x+(\mu\mathbf{1}_{\{x>0\}}+x\beta)\bar{v}_{x-1}^{\alpha,n}\big\}\Big]$$

First note that starting the iterations with the simple function $\bar{v}^{\alpha,0}\equiv 0$, only yields a convex function $\bar{v}^{\alpha,1}$,

$$\bar{v}_x^{\alpha,1}=\frac{x-\lambda K}{\alpha+\lambda+\mu\mathbf{1}_{\{x>0\}}+x\beta},\quad x=0,1,\dots,$$

under restrictions on the input parameters. In the minimisation one has to compare terms with different denominators. For showing convexity this is even more complicated, since one has to show that

$$\bar{v}_{x+2}^{\alpha,n+1}-\bar{v}_{x+1}^{\alpha,n+1}\geq\bar{v}_{x+1}^{\alpha,n+1}-\bar{v}_x^{\alpha,n+1},$$

given convexity of $\bar{v}^{\alpha,n}$, where each of these terms is a minimisation of two terms with different denominators. Already for this simple example it is not clear that this will work.

Note that applying VI on the average cost embedded jump MDP has the same disadvantages. Additionally, one needs extra conditions (cf. Theorem 2.10) to ensure that average VI converges at all.

Summary

The next table summarises the different roadmaps, with the appropriate references to the results justifying the various steps.

	Dynamics	Criterion	Roadmap			
1	DT-time	α-discounted	VI (Alg. 1) Thm. 2.6			
2.1	DT-time Vgeo	average	VI (Alg. 2) Thm. 2.10			
2.2	DT-time no Vgeo	average	VDA Thm. 2.8	then VI (Alg. 1) Thm. 2.6		
3.1	CT-time bdd. rates	α-discounted	UNI § 3.1	then VI (Alg. 1) Thm. 2.6		
3.2	CT-time unb. rates	α-discounted	PB Cor. 3.4	then UNI § 3.1	then VI (Alg. 1) Thm. 2.6	
4.1	CT-time bdd. rates	average	VDA Thm. 3.5	then UNI § 3.1	then VI (Alg. 1) Thm. 2.6	
4.2	CT-time unb. rates	average	VDA Thm. 3.5	then PB Cor. 3.4	then UNI § 3.1	then VI (Alg. 1) Thm. 2.6

Abbreviations to summarising table

Discrete time: DT-time
Continuous time: CT-time
Algorithm: Alg.
Value Iteration: VI
Vanishing Discount Approach: VDA
Uniformisation: UNI

Perturbation: PB
Conditions Theorem 2.10: VGeo
bounded: bdd.
unbounded: unb.

A Proof of Theorem 3.5

For the proof of Theorem 3.5 we will need a number of prepatory lemmas.

Lemma A.1. *Suppose that Assumptions 3.1, 3.2*(α)*,* $\alpha > 0$*, and 2.3 hold. The following hold.*

i) *Let* $\mu(\phi_0)$ *denote the stationary distribution under parameter* ϕ_0*, where* ϕ_0 *has been specified in Assumption 2.3. Then* ϕ_0 *has one closed class,* R *say, that is positive recurrent. It holds, that*

$$\mathsf{g}(\phi_0) = \alpha \sum_R \mu_x(\phi_0) v_x^\alpha(\phi_0). \tag{A.1}$$

ii) *Let* $\phi \in \Phi'$*. Let* $x \notin D$*, and put* $\tau := \tau_D$ *to be the hitting time of* D *(cf. (3.9)). Then*

$$v_x^\alpha(\phi) \geq \mathsf{E}_x^\phi \Big[\mathbb{1}_{\{\tau=\infty\}} \frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + \mathbb{1}_{\{\tau<\infty\}} \big((1-e^{-\alpha\tau}) \frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + e^{-\alpha\tau} v_{X_\tau}^\alpha(\phi) \big) \Big]. \tag{A.2}$$

iii) *There exists* $x_\alpha \in D$ *with* $v_{x_\alpha}^\alpha = \inf_x v_x^\alpha$*.*

Proof. First we prove (i). By virtue of Assumption 2.3 (ii) the Markov process associated with ϕ_0 has one closed class, which is positive recurrent. Furthermore, absorption into this class takes place in finite expected time and with finite expected cost, for any initial state $x \notin R$, since necessarily $x_0 \in R$.

Then we get

$$\begin{aligned}
\sum_{x\in R} \mu_x(\phi_0) \mathsf{E}_x^{\phi_0}[c_{X_t}] &= \sum_{x\in R} \mu_x(\phi_0) \sum_{y\in R} p_{t,xy}(\phi_0) c_y(\phi_0) \\
&= \sum_{y\in R} c_y(\phi_0) \sum_{x\in R} \mu_x(\phi_0)\, p_{t,xy}(\phi_0) \\
&= \sum_{y\in R} c_y(\phi_0) \mu_y(\phi_0) = \mathsf{g}(\phi_0),
\end{aligned}$$

where the interchange of summation is allowed by nonnegativity. This is used as well to justify the next derivation

$$\begin{aligned}
\alpha \sum_{x\in R} \mu_x(\phi_0) v_x^\alpha(\phi_0) &= \alpha \sum_{x\in R} \mu_x(\phi_0) \mathsf{E}_x^{\phi_0} \big[\int_{t=0}^\infty e^{-\alpha t} c_{X_t} dt \big] \\
&= \alpha \int_{t=0}^\infty e^{-\alpha t} \sum_{x\in R} \mu_x(\phi_0) \mathsf{E}_x^{\phi_0}[c_{X_t}] dt \\
&= \alpha \int_{t=0}^\infty e^{-\alpha t} g(\phi_0) dt = \mathsf{g}(\phi_0).
\end{aligned}$$

The proof of (ii) follows by splitting the α-discounted cost into three terms, the first two of which represent the α-discounted cost till τ, in the respective cases $\tau = \infty$ and $\tau < \infty$, and the third is the cost starting from $\tau < \infty$:

$$\begin{aligned} v_x^\alpha(\phi) &= \mathsf{E}_x^\phi\Big[\int_{t=0}^{\infty} e^{-\alpha t} c_{X_t} dt\Big] \\ &\geq \mathsf{E}_x^\phi\Big[\mathbb{1}_{\{\tau=\infty\}} \int_{t=0}^{\infty} e^{-\alpha t} dt(\mathsf{g}(\phi_0)+\epsilon) + \mathbb{1}_{\{\tau<\infty\}}\Big(\int_{t=0}^{\tau} e^{-\alpha t} dt(\mathsf{g}(\phi_0)+\epsilon) \int_{t=\tau}^{\infty} e^{-\alpha t} c_{X_t} dt\Big)\Big] \\ &= \mathsf{E}_x^\phi\Big[\mathbb{1}_{\{\tau=\infty\}} \frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + \mathbb{1}_{\{\tau<\infty\}}\big((1-e^{-\alpha\tau})\frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + e^{-\alpha\tau} v_{X_\tau}^\alpha(\phi)\big)\Big]. \end{aligned}$$

The inequality is due to the definitions of D and τ.

We finally prove (iii). Part (i) implies the existence of $z_\alpha \in R$ such that $\mathsf{g}(\phi_0) \geq \alpha v_{z_\alpha}^\alpha(\phi_0)$. Then there also exists a $y_\alpha \in D$ with $\mathsf{g}(\phi_0) \geq \alpha v_{y_\alpha}^\alpha(\phi_0)$. Indeed, suppose such $y_\alpha \in D$ does not exist. Then $v_y^\alpha(\phi_0) > \frac{\mathsf{g}(\phi_0)}{\alpha}$ for all $y \in D$. This leads to a contradiction, since by virtue of part (ii)

$$\frac{\mathsf{g}(\phi_0)}{\alpha} \geq v_{z_\alpha}^\alpha(\phi_0) \geq \mathsf{E}_{z_\alpha}^\phi\Big[(1-e^{-\alpha\tau})\frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + e^{-\alpha\tau} v_{X_\tau}^\alpha(\phi_0)\Big] > \frac{\mathsf{g}(\phi_0)}{\alpha}.$$

Let $x_\alpha = \arg\min_{y\in D} v_y^\alpha$, and so $v_{x_\alpha}^\alpha \leq v_{x_\alpha}^\alpha(\phi_0) \leq \frac{g(\phi_0)}{\alpha}$. Then $x_\alpha = \arg\min_y v_y^\alpha$, because by Eqn. (A.2) for any $x \notin D(\phi_0)$ and α-discount optimal policy ϕ_α

$$\begin{aligned} v_x^\alpha &= v_x^\alpha(\phi_\alpha) \\ &\geq \mathsf{E}_x^\phi\Big[\mathbb{1}_{\{\tau=\infty\}} \frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + \mathbb{1}_{\{\tau<\infty\}}\big((1-e^{-\alpha\tau})\frac{\mathsf{g}(\phi_0)+\epsilon}{\alpha} + e^{-\alpha\tau} v_{X_\tau}^\alpha\big)\Big] \\ &\geq \mathsf{E}_x^\phi\Big[\mathbb{1}_{\{\tau=\infty\}} v_{x_\alpha}^\alpha + \mathbb{1}_{\{\tau<\infty\}}\big((1-e^{-\alpha\tau}) v_{x_\alpha}^\alpha + e^{-\alpha\tau} v_{x_\alpha}^\alpha\big)\Big] \\ &= v_{x_\alpha}^\alpha. \end{aligned}$$

□

Lemma A.2. *Suppose that Assumptions 3.1, 3.2(α), $\alpha > 0$, and 2.3 hold. Let $\{\alpha_n\}_n$ be a positive sequence converging to 0. The following hold.*

i) *There exist a subsequence, call it $\{\alpha_n\}_n$ again, and $x_0 \in D$ such that $\alpha_n v_{x_0}^{\alpha_n} \leq \mathsf{g}(\phi_0)$, $n = 1, 2, \ldots$.*

ii) *There exist a constant L and a function $M : \boldsymbol{S} \to (0, \infty)$, such that $-L \leq v_x^\alpha - v_z^\alpha \leq M_x$, $\alpha > 0$.*

Proof. To prove (i), note that Lemma A.1 (iii) implies for all n the existence of $x_{\alpha_n} \in D$, such that $v_{x_{\alpha_n}}^{\alpha_n} \leq v_x^{\alpha_n}$, $x \in \boldsymbol{S}$. By Assumption 2.3 (iii) D is finite, and so there exists $x_0 \in D$ and a subsequence of $\{\alpha_n\}_n$, that we may call $\{\alpha_n\}_n$ again, such that $x_{\alpha_n} = x_0$. Therefore by Lemma A.1 (i), for all n

$$\alpha_n v_{x_0}^{\alpha_n} \leq \alpha_n \sum_x \mu_x(\phi_0) v_x^{\alpha_n} \leq \alpha_n \sum_x \mu_x(\phi_0) v_x^{\alpha_n}(\phi_0) = \mathsf{g}(\phi_0).$$

For the proof of (ii), take

$$M_x = c_{xz}(\phi_0), \quad L = \max_{y\in D} c_{zy}(\phi^y),$$

with z and ϕ^y from Assumptions 2.3 (ii) and (iv). Let $\alpha > 0$. Let strategy ϕ follow ϕ_0 until z is reached, from then onwards it follows the α-discount optimal policy ϕ_α. Then again by Assumption 2.3 (ii) we have

$$v_x^\alpha \leq v_x^\alpha(\phi) \leq c_{xz}(\phi_0) + v_z^\alpha(\phi_\alpha) = c_{xz}(\phi_0) + v_z^\alpha.$$

Notice that Assumptions 2.3 (iii) and (iv) yield $L < \infty$. According to Lemma A.1 (iv) there is a minimum cost starting state $x_\alpha \in D$. Let ϕ' be the policy that uses policy ϕ^{x_α} of Assumption 2.3 (iv) until hitting x_α, after which ϕ' follows the α-discount optimal policy ϕ_α. This yields,

$$v_z^\alpha - v_x^\alpha \le v_z^\alpha - \min_x v_x^\alpha \le v_z^\alpha(\phi') - v_{x_\alpha}^\alpha \le c_{zx_\alpha}(\phi^{x_\alpha}) \le L.$$

□

Lemma A.3. *Suppose that Assumptions 3.1, 3.2(α), $\alpha > 0$, and 2.3 hold. Then,*

$$\limsup_{\alpha \downarrow 0} \alpha v_x^\alpha \le \mathsf{g}_x(\phi), \quad x \in \boldsymbol{S}, \phi \in \Phi'. \tag{A.3}$$

Proof. Let $\phi \in \Phi'$. We wish to apply Theorem B.3 for $s(t) = \sum_y p_{t,xy}(\phi) c_y(\phi)$. First, Assumption 3.1, Assumption 3.2 (α) and the dominated convergence theorem yield that $t \mapsto \sum_y p_{t,xy}(\phi) c_y(\phi)$ is continuous and $|v_x^\alpha(\phi)| < \infty$ (cf. [13, Theorem 3.2]). By Assumption 2.3 (i),

$$\sum_y p_{t,xy}(\phi) c_y(\phi), \quad v_x^\alpha(\phi) \ge 0, \quad x \in \boldsymbol{S}.$$

Then, $S(\alpha) = v_x^\alpha(\phi)$ and $\mathsf{g}_x(\phi) = \limsup_{T\to\infty} \frac{1}{T} S_T$. Hence Theorem B.3 (1c) implies

$$\limsup_{\alpha \downarrow 0} \alpha v_x^\alpha(\phi) \le \mathsf{g}_x(\phi).$$

□

Lemma A.4 ([25, Theorem 5.2])**.** *Suppose that Assumptions 3.1, 3.2(α), $\alpha > 0$, and 2.3 hold. Let (g, v) be a tuple, with $g \in \mathbf{R}$ and $v : \boldsymbol{S} \to [-L, \infty)$, $x \in \boldsymbol{S}$, and $\phi \in \Phi'$ be such that*

$$g \ge c_x(\phi) + \sum_y q_{xy}(\phi) v_y, \quad x \in \boldsymbol{S}. \tag{A.4}$$

Then $\mathsf{g}_x(\phi) \le g$, $x \in \boldsymbol{S}$.

Now we have all results at hand to finish the proof of Theorem 3.5. The most important difficulty is to obtain the CAOE from a continuous time average cost optimality inequality (CAOI). To achieve this we have translated a very interesting argument used in [40, Chapter 7] for the discrete time case to continuous time.

Proof of Theorem 3.5. Let $\{\alpha_n\}_n > 0$ be a positive sequence converging to 0. Lemma A.2(ii) implies that $-L \le v_x^{\alpha_n} - v_z^{\alpha_n} \le M_x$, for a constant L and a function $M : \boldsymbol{S} \to (0, \infty)$, and $x \in \boldsymbol{S}$. Note that $[-L, M_x]$ is compact. By a diagonalisation argument, the sequence has a convergent subsequence, denoted $\{\alpha_n\}_n$ again, along which the limit exists for any $x \in \boldsymbol{S}$, say $v_x^{\alpha_n} - v_z^{\alpha_n} \to v'_x$, $x \in \boldsymbol{S}$.

Lemma A.2(i) implies that there exists a further subsequence, again denoted $\{\alpha_n\}_n$, such that $0 \le \alpha_n v_{x_0}^{\alpha_n} \le \mathsf{g}(\phi_0)$, for some $x_0 \in D$. Compactness of $[0, \mathsf{g}(\phi_0)]$ implies existence of a limit point, say g', along a subsequence, that in turn is denoted by $\{\alpha_n\}_n$.

By the above, $\alpha_n(v_y^{\alpha_n} - v_{x_0}^{\alpha_n}) \to 0$, and thus $\alpha_n v_y^{\alpha_n} \to g'$ for all $y \in \boldsymbol{S}$.

Since Φ' is compact, there is a final subsequence of $\{\alpha_n\}_n$, denoted likewise, such that $\{\phi^{\alpha_n}\}_n$, with ϕ^{α_n} an α-discount optimal policy, has a limit point ϕ' say. The tuple (g', v') has property (3) from part (i) of the Theorem.

We will next show that this tuple is a solution to the following inequality:

$$g' \geq c_x(\phi') + \sum_y q_{xy}(\phi')v'_y \geq \inf_{\phi\in\Phi'}\{c_x(\phi) + \sum_y q_{xy}(\phi)v'_y\}. \tag{A.5}$$

Indeed, the α-DDOE (3.8) yields for all $x \in \boldsymbol{S}$

$$\alpha v_x^\alpha \quad = \quad c_x(\phi_\alpha) + \sum_y q_{xy}(\phi_\alpha)v_y^\alpha.$$

Then we use Fatou's lemma and obtain

$$\begin{aligned}
g' &= \liminf_{n\to\infty}\{\alpha_n v_x^{\alpha_n}\} \\
&= \liminf_{n\to\infty}\left\{c_x(\phi_{\alpha_n}) + \sum_{y\neq x} q_{xy}(\phi_{\alpha_n})[v_y^{\alpha_n} - v_z^{\alpha_n}] - q_x(\phi_{\alpha_n})[v_x^{\alpha_n} - v_z^{\alpha_n}]\right\} \\
&\geq c_x(\phi') + \sum_{y\neq x}\liminf_{n\to\infty}\{q_{xy}(\phi_{\alpha_n})[v_y^{\alpha_n} - v_z^{\alpha_n}]\} - \liminf_{n\to\infty}\left\{q_x(\phi_{\alpha_n})[v_x^{\alpha_n} - v_z^{\alpha_n}]\right\} \\
&= c_x(\phi') + \sum_y q_{xy}(\phi')v'_y \\
&\geq \inf_{\phi\in\Phi'}\{c_x(\phi) + \sum_y q_{xy}(\phi)v'_y\},
\end{aligned}$$

where subtraction of $v_z^{\alpha_n}$ is allowed, since $Q(\phi)$ has row sums equal to zero. In the third equation we use continuity of $\phi \mapsto c_x(\phi)$ and $\phi \mapsto q_{xy}(\phi)$.

This allows to show that (g', v') has property (1) from the Theorem and that ϕ' is optimal in Φ'. Indeed, Lemma A.3 and Lemma A.4 yield for all $x \in \boldsymbol{S}$

$$\mathsf{g}_x(\phi') \leq g' = \lim_{n\to\infty}\alpha_n v_x^{\alpha_n} \leq \mathsf{g}_x \leq \mathsf{g}_x(\phi'). \tag{A.6}$$

Hence $\mathsf{g}_x(\phi') = \mathsf{g}_x = g'$, $x \in \boldsymbol{S}$, and so ϕ' is optimal in Φ', and g' is the minimum expected average cost.

The following step is to show that both inequalities in Eqn. (A.5) are in fact equalities. To this end, it is sufficient to show that (g', v') is a solution tuple to the CAOE (3.10). Then Eqn. (A.5) immediately implies that ϕ' takes minimising actions in (3.10) for the solution (g', v').

Hence, let us assume the contrary. If $g' > \inf_{\phi\in\Phi'}\{c_x(\phi)+\sum_y q_{xy}(\phi)v'_y\}$ then there exists $\bar{\phi}_x \in \Phi'_x$, such that $g' > c_x(\bar{\phi}_x) + \sum_y q_{xy}(\bar{\phi}_x)v'_y$. Put $d_x \geq 0$ to be the corresponding discrepancy

$$g' = c_x(\bar{\phi}_x) + d_x + \sum_y q_{xy}(\bar{\phi}_x)v'_y.$$

As a consequence, if the inequality in (A.5) is not an equality, then there exists $\bar{\phi} \in \Phi'$ and a discrepancy function $d : \boldsymbol{S} \to [0, \infty)$, $d \not\equiv 0$, such that

$$g' = c_x(\bar{\phi}) + d_x + \sum_y q_{xy}(\bar{\phi})v'_y, \quad x \in \boldsymbol{S}. \tag{A.7}$$

In other words

$$0 = c_x(\bar{\phi}) + d_x - g' + \sum_y q_{xy}(\bar{\phi})v'_y, \quad x \in \boldsymbol{S}.$$

For $x \notin D$, $c_x(\bar{\phi}) + d_x - g' \geq \mathsf{g}(\phi_0) + \epsilon - g' \geq \epsilon$, and so $v' + L\mathbf{e}$ is a non-negative solution to the equation

$$\sum_y q_{xy}(\bar{\phi})(v'_y + L) \leq -\epsilon, \quad y \notin D.$$

This is precisely the condition in [47, Theorem 1] with $\lambda = 0$[2]. Following the proof of that theorem and using that $q_x(\bar{\phi}) > 0$ for $x \notin D$ (otherwise $\mathsf{g}_x(\bar{\phi}) = c_x(\bar{\phi}) > g'$), we can conclude that

$$v'_x + L \geq m_{xD}(\bar{\phi}), \quad x \notin D,$$

so that $m_{xD}(\bar{\phi}) < \infty$, for $x \notin D$.

For $x \in D$, either $q_x(\bar{\phi}) = 0$, or $q_x(\bar{\phi}) > 0$ and

$$m_{xD}(\bar{\phi}) = \frac{1}{q_x(\bar{\phi})} + \sum_{y \notin D} \frac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})} m_{yD}(\bar{\phi}) \leq \frac{1}{q_x(\bar{\phi})} + \sum_y \frac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})}(v'_y + L) < \infty,$$

by virtue of (A.7). We now will perform an iteration argument along the same lines as the proof of [47, Theorem 1].

First consider the case that $q_x(\bar{\phi}) > 0$. Dividing Eqn. (A.7) for state x by $q_x(\bar{\phi})$ we get, after reordering,

$$v'_x \geq \frac{c_x(\bar{\phi}) + d_x - g'}{q_x(\bar{\phi})} + \sum_{y \neq x} \frac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})} v'_y.$$

Introduce the substochastic matrix P on $\boldsymbol{S} \setminus D$ by

$$p_{xy} = \begin{cases} \dfrac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})} & y \notin D \cup \{x\} \\ 0 & \text{otherwise.} \end{cases}$$

Denote the n iterate by $P^{(n)}$, where $P^{(0)}$ is the $\boldsymbol{S} \times \boldsymbol{S}$ identity matrix. Then, for $x \notin D$ we get

$$\begin{aligned} v'_x &\geq \frac{c_x(\bar{\phi}) + d_x - g'}{q_x(\bar{\phi})} + \sum_y p_{xy} v'_y + \sum_{y \in D} \frac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})} v'_y \\ &\geq \frac{c_x(\bar{\phi}) + d_x - g'}{q_x(\bar{\phi}} + + \sum_{y \in D} \frac{q_{xy}(\bar{\phi})}{q_x(\bar{\phi})} v'_y + \sum_y p_{xy} \Big[\frac{c_y(\bar{\phi}) + d_y - g'}{q_y(\bar{\phi})} + \sum_w p_{yw} v'_w + \sum_{w \in D} \frac{q_{yw}(\bar{\phi})}{q_y(\bar{\phi})} v'_w \Big] \\ &\geq \sum_{n=0}^{N-1} \sum_y p^{(n)}_{xy} \frac{c_y(\bar{\phi}) + d_y - g'}{q_y(\bar{\phi})} + \sum_{n=0}^{N-1} \sum_y p^{(n)}_{xy} \sum_{w \in D} \frac{q_{yw}(\bar{\phi})}{q_y(\bar{\phi})} v'_w + \sum_y p^{(N)}_{xy} v'_y. \end{aligned}$$

Taking the liminf $N \to \infty$, we get

$$v'_x \geq \sum_{n=0}^{\infty} \sum_y p^{(n)}_{xy} \frac{c_y(\bar{\phi}) + d_y - g'}{q_y(\bar{\phi})} + \sum_{n=0}^{\infty} \sum_y p^{(n)}_{xy} \sum_{w \in D} \frac{q_{yw}(\bar{\phi})}{q_y(\bar{\phi})} v'_w + \liminf_{N \to \infty} \sum_y p^{(N)}_{xy} v'_y.$$

Clearly

$$\liminf_{N \to \infty} \sum_y p^{(N)}_{xy} v'_y \geq \liminf_{N \to \infty} \sum_y p^{(N)}_{xy} (-L).$$

[2]Factor λ in front of y_i in that paper has been mistakenly omitted

However, since $m_{xD}(\bar{\phi}) < \infty$, $x \notin D$, we get that $\liminf_{N\to\infty} \sum_y p_{xy}^{(N)} = 0$. Hence, for $\tau := \tau_D$

$$\begin{aligned} v'_x &\geq \sum_{n=0}^{\infty} \sum_y p_{xy}^{(n)} \frac{c_y(\bar{\phi}) + d_y - g'}{q_y(\bar{\phi})} + \sum_{n=0}^{\infty} \sum_y p_{xy}^{(n)} xy \sum_{w\in D} \frac{q_{yw}(\bar{\phi})}{q_y(\bar{\phi})} v'_w \\ &\geq \mathsf{E}_x^{\bar{\phi}}\Big[\int_{t=0}^{\tau} (c_{X_t} + d_{X_t} - g')dt\Big] + \mathsf{E}_x^{\bar{\phi}}\Big[v'_{X_\tau}\Big] \\ &= c_{xD}(\bar{\phi}) + \mathsf{E}_x^{\bar{\phi}}\Big[\int_{t=0}^{\tau} d_{X_t} dt\Big] - m_{xD}(\bar{\phi})g' + \mathsf{E}_x^{\bar{\phi}}\Big[v'_{X_\tau}\Big], \end{aligned} \tag{A.8}$$

for $x \notin D$. For $x \in D$ we can derive the same inequality. Note that we assumed $q_x(\bar{\phi}) > 0$. On the other hand, we have that

$$v_x^\alpha \leq c_{xD}(\bar{\phi}) + \mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha\tau} v_{X_\tau}^\alpha\big].$$

This is equivalent to

$$v_x^\alpha - v_z^\alpha \leq c_{xD}(\bar{\phi}) - v_z^\alpha(1 - \mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha\tau}\big]) + \mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha\tau}(v_{X_\tau}^\alpha - v_z^\alpha)\big].$$

Hence, for the sequence $\{\alpha_n\}_n$ we have

$$v_x^{\alpha_n} - v_z^{\alpha_n} \leq c_{xD}(\bar{\phi}) - \alpha_n v_z^{\alpha_n} \frac{1 - \mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha_n\tau}\big]}{\alpha_n} + \mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha_n\tau}(v_{X_\tau}^{\alpha_n} - v_z^{\alpha_n})\big].$$

Taking the limit of n to infinity yields

$$\begin{aligned} v'_x &\leq c_{xD}(\bar{\phi}) - g' \cdot m_{xD}(\bar{\phi}) + \lim_{n\to\infty} \big\{\mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha_n\tau}(v_{X_\tau}^{\alpha_n} - v_z^{\alpha_n})\big]\big\} \\ &= c_{xD}(\bar{\phi}) - g' \cdot m_{xD}(\bar{\phi}) + \mathsf{E}_x^{\bar{\phi}}\big[v'(X_\tau)\big]. \end{aligned} \tag{A.9}$$

Taking the limit through the expectation is justified by the dominated convergence theorem, since

$$|\mathsf{E}_x^{\bar{\phi}}\big[e^{-\alpha_n\tau}(v_{X_\tau}^{\alpha_n} - v_z^{\alpha_n})\big]| \leq \mathsf{E}_x^{\bar{\phi}} e^{-\alpha_n\tau} |v_{X_\tau}^{\alpha_n} - v_z^{\alpha_n}| \leq \mathsf{E}_x^{\bar{\phi}}(M_{X_\tau} \vee L) < \infty.$$

Combining Eqns. (A.8) and (A.9) yields $d \equiv 0$, for x with $q_x(\bar{\phi}) > 0$.

For x with $q_x(\bar{\phi}) = 0$, necessarily $g' = c_x(\bar{\phi})$ and then equality in Eqn. (A.5) immediately follows. Since there is equality for ϕ', this also implies that the inf is a min, and so we have obtained that (g', v') is a solution to the CAOE (3.10).

The only thing left to prove, is that the solution tuple (g', v') has property (2), that is, every minimising policy in (3.10) is average cost optimal. But this follows in the same manner as the argument leading to Eqn. (A.6) yielding optimality of ϕ'. This finishes the proof.

B Tauberian Theorem

This section develops a Tauberian theorem that is used to provide the necessary ingredients for proving Theorem 3.5. This theorem is the continuous time counterpart of Theorem A.4.2 in Sennott [40]. A related assertion can be found in [25, Proposition A.5], however, in a weaker variant (without the Karamata implication, see Theorem B.3, implication $(i) \implies (iii)$). The continuous time version seems deducible from Chapter 5 of the standard work on this topic [50]. We give a direct proof here.

Let $s : [0,\infty) \to \mathbf{R}$ be a function that is bounded below by $-L$ say and $(\mathcal{B}([0,\infty)), \mathcal{B})$-measurable, where $\mathcal{B}$ denotes the Borel-σ-algebra on $\mathbf{R}$, and $\mathcal{B}([0,\infty))$ the Borel-σ-algebra on $[0,\infty)$. Assume for any $\alpha > 0$ that

$$S(\alpha) = \int_{t=0}^{\infty} s(t) e^{-\alpha t} dt < \infty.$$

Furthermore, assume for any $T > 0$ that

$$S_T = \int_{t=0}^{T} s(t)dt < \infty.$$

Lemma B.1. *Suppose that $L = 0$, i.e. s is a nonnegative function. Then for all $\alpha > 0$ it holds that*

$$\frac{1}{\alpha} S(\alpha) = \int_{T=0}^{\infty} e^{-\alpha T} S_T dT. \tag{B.1}$$

Furthermore, for all $\alpha > 0, U \geq 0$ the following inequalities hold true:

$$\alpha S(\alpha) \geq \inf_{T \geq U} \left\{ \frac{S_T}{T} \right\} \left(1 - \alpha^2 \int_{T=0}^{U} T e^{-\alpha T} dT \right), \tag{B.2}$$

and

$$\alpha S(\alpha) \leq \alpha^2 \int_{T=0}^{U} e^{-\alpha T} S_T dT + \sup_{T \geq U} \left\{ \frac{S_T}{T} \right\}. \tag{B.3}$$

Proof. We first prove Equation (B.1). To this end, let $\alpha > 0$. Then,

$$\begin{aligned}
\frac{1}{\alpha} S(\alpha) &= \int_{u=0}^{\infty} e^{-\alpha u} du \int_{t=0}^{\infty} s(t) e^{-\alpha t} dt \\
&= \int_{t=0}^{\infty} \int_{u=0}^{\infty} s(t) e^{-\alpha(u+t)} du dt \\
&= \int_{T=0}^{\infty} \int_{t=0}^{T} s(t) e^{-\alpha T} du dT \\
&= \int_{T=0}^{\infty} e^{-\alpha T} \int_{t=0}^{T} s(t) du dT \\
&= \int_{T=0}^{\infty} e^{-\alpha T} S_T dT.
\end{aligned}$$

Interchange of integrals, change of variables are allowed, since the integrands are non-negative and the integrals are finite.

Next, we prove Equation (B.2). To this end, we use Equation (B.1). Then, for all $\alpha > 0$, $U \geq 0$

$$\begin{aligned}
\alpha S(\alpha) &= \alpha^2 \int_{T=0}^{\infty} S_T e^{-\alpha T} dT \\
&\geq \alpha^2 \int_{T=U}^{\infty} \frac{S_T}{T} T e^{-\alpha T} dT \\
&\geq \alpha^2 \inf_{t \geq U} \left\{ \frac{S_T}{T} \right\} \int_{T=U}^{\infty} T e^{-\alpha T} dT \\
&= \alpha^2 \inf_{t \geq U} \left\{ \frac{S_T}{T} \right\} \left(\int_{T=0}^{\infty} T e^{-\alpha T} dT - \int_{T=0}^{U} T e^{-\alpha T} dT \right) \\
&= \inf_{T \geq U} \left\{ \frac{S_T}{T} \right\} \left(1 - \alpha^2 \int_{T=0}^{U} T e^{-\alpha T} dT \right).
\end{aligned}$$

The first inequality uses explicitly that the integrand is non-negative.

Similarly we expand from Eqn. (B.1) to get Inequality (B.3) as follows. Let $\alpha > 0, U \geq 0$. Then,

$$\begin{aligned}
\alpha S(\alpha) &= \alpha^2 \int_{T=0}^{\infty} e^{-\alpha T} S_T dT \\
&= \alpha^2 \int_{T=0}^{U} e^{-\alpha T} S_T dT + \alpha^2 \int_{T=U}^{\infty} e^{-\alpha T} T \frac{S_T}{T} dT \\
&\leq \alpha^2 \int_{T=0}^{U} e^{-\alpha T} S_T dT + \sup_{T \geq U} \left\{ \frac{S_T}{T} \right\} \alpha^2 \int_{T=U}^{\infty} T e^{-\alpha T} dT \\
&\leq \alpha^2 \int_{T=0}^{U} e^{-\alpha T} S_T dT + \sup_{T \geq U} \left\{ \frac{S_T}{T} \right\} \alpha^2 \int_{T=0}^{\infty} T e^{-\alpha T} dT \\
&= \alpha^2 \int_{T=0}^{U} e^{-\alpha T} S_T dT + \sup_{T \geq U} \left\{ \frac{S_T}{T} \right\}.
\end{aligned}$$

□

Let $f : [0,1] \to \mathbf{R}$ be an integrable function, and define

$$S_f(\alpha) = \int_{t=0}^{\infty} e^{-\alpha t} f(e^{-\alpha t}) s(t) dt.$$

Lemma B.2. *Assume that $L = 0$ and*

$$W := \liminf_{\alpha \downarrow 0} \alpha S(\alpha) = \limsup_{\alpha \downarrow 0} \alpha S(\alpha) < \infty.$$

Let $r : [0,1] \to \mathbf{R}$ be given by

$$r(x) = \begin{cases} 0 & x < 1/e \\ 1/x & x \geq 1/e. \end{cases}$$

Then

$$\lim_{\alpha \downarrow 0} \alpha S_r(\alpha) = \left(\int_{x=0}^{1} r(x) dx \right) \lim_{\alpha \downarrow 0} \alpha S(\alpha). \tag{B.4}$$

Proof. We first prove (B.4) for polynomial functions, then for continuous functions and finally for r. To show that (B.4) holds for polynomials, it is sufficient to prove it for $p(x) = x^k$. Thus,

$$\begin{aligned}
\alpha S_p(\alpha) &= \alpha \int_{t=0}^{\infty} e^{-\alpha t} (e^{-\alpha t})^k s(t) dt \\
&= \frac{1}{k+1} \left[\alpha(k+1) \int_{t=0}^{\infty} e^{-\alpha(k+1)t} s(t) dt \right] \\
&= \int_{x=0}^{1} p(x) dx \left[\alpha(k+1) S(\alpha(k+1)) \right].
\end{aligned}$$

Taking the limit of $\alpha \downarrow 0$ proves (B.4) for polynomials. This is allowed because W is finite. Next we show Eqn. (B.4) for continuous functions. The Weierstrass approximation theorem (see [46, Section 13.33], [3]) yields that a continuous function q on a closed interval can be arbitrary closely approximated by polynomials. Let p such that $p(x) - \epsilon \leq q(x) \leq p(x) + \epsilon$ for $0 \leq x \leq 1$. Then,

$$\int_{x=0}^{1} p(x) dx - \epsilon \leq \int_{x=0}^{1} q(x) dx \leq \int_{x=0}^{1} p(x) dx + \epsilon.$$

$$\begin{aligned} S_{p-\epsilon}(\alpha) &= \int_{t=0}^{\infty} e^{-\alpha t}(p(e^{-\alpha t}) - \epsilon)s(t)dt \\ &= \int_{t=0}^{\infty} e^{-\alpha t}p(e^{-\alpha t})s(t)dt - \epsilon\int_{t=0}^{\infty} e^{-\alpha t}s(t)dt \\ &= S_p(\alpha) - \epsilon S(\alpha) \end{aligned}$$

This implies

$$0 \leq S_{p+\epsilon}(\alpha) - S_{p-\epsilon}(\alpha) \leq 2\epsilon S(\alpha),$$

As ϵ approaches 0, finiteness of W yields the result for continuous functions. In a similar manner r can be approximated by continuous functions q, q' with $q' \leq r \leq q$ as follows

$$q(x) = \begin{cases} 0 & x < \frac{1}{e} - \delta \\ \frac{e}{\delta}x + e - \frac{1}{\delta} + & \frac{1}{e} - \delta \leq x < \frac{1}{e} \\ \frac{1}{x} & x \geq \frac{1}{e} \end{cases} \qquad q'(x) = \begin{cases} 0 & x < \frac{1}{e} \\ \frac{e}{\gamma+\gamma^2 e}x + \frac{1}{\gamma+\gamma^2 e} & \frac{1}{e} \geq x > \frac{1}{e} + \gamma \\ 1/x & \frac{1}{e} + \gamma \geq x. \end{cases}$$

This proves Equality (B.4). □

Theorem B.3. *The following assertions hold.*

1. $\liminf_{T\to\infty} \frac{S_T}{T} \overset{(a)}{\leq} \liminf_{\alpha\downarrow 0} \alpha S(\alpha) \overset{(b)}{\leq} \limsup_{\alpha\downarrow 0} \alpha S(\alpha) \overset{(c)}{\leq} \limsup_{T\to\infty} \frac{S_T}{T}$;
2. *the following are equivalent*

$$\begin{aligned} &i) \quad \liminf_{\alpha\downarrow 0} \alpha S(\alpha) &&= \limsup_{\alpha\downarrow 0} \alpha S(\alpha) < \infty; \\ &ii) \quad \liminf_{T\to\infty} \tfrac{S_T}{T} &&= \limsup_{T\to\infty} \frac{S_T}{T} < \infty; \\ &iii) \quad \lim_{\alpha\downarrow 0} \alpha S(\alpha) &&= \lim_{T\to\infty} \frac{S_T}{T} < \infty. \end{aligned}$$

Proof. This proof is based on Sennott [40]. Clearly inequality (b) holds. So this leaves to prove inequalities (a) and (c). Proof of inequality (a). First notice, that if we take $s \equiv M$ a constant function, then

$$\liminf_{T\to\infty} \frac{S_T}{T} = \liminf_{\alpha\downarrow 0} \alpha S(\alpha) = \limsup_{\alpha\downarrow 0} \alpha S(\alpha) = \limsup_{T\to\infty} \frac{S_T}{T}.$$

Therefore adding a constant M to the function s does not influence the result. Hence it is sufficient to prove the theorem for nonnegative functions s. This means the assumptions of Lemma B.1 hold and we may use inequality (B.2). Thus,

$$\inf_{T\geq U} \left\{ \frac{S_T}{T} \right\} \left(1 - \alpha^2 \int_{T=0}^{U} Te^{-\alpha T} dT \right) \leq \alpha S(\alpha).$$

Notice that $\lim_{\alpha\downarrow 0} \alpha^2 \int_{T=0}^{U} Te^{-\alpha T} dT = 0$, hence taking the lim inf as $\alpha \downarrow 0$ gives

$$\inf_{T\geq U} \frac{S_T}{T} \leq \liminf_{\alpha\downarrow 0} \alpha S(\alpha).$$

Now taking the limit $U \to \infty$ on both sides gives

$$\liminf_{T\to\infty} \frac{S_T}{T} \leq \liminf_{\alpha\downarrow 0} \alpha S(\alpha),$$

which yield the result. Using Inequality (B.3) of Lemma B.1 and apply the same reasoning proves inequality (c).

Next we prove part 2. Part 1 implies that $i) \Longleftarrow ii) \Longleftrightarrow iii)$. So it is sufficient to prove that $i) \Longrightarrow iii)$.
Assume that $i)$ holds, then we may invoke Lemma B.2. First notice that $\int_{x=0}^{1} r(x)dx = 1$, hence Eqn. (B.4) reduces to

$$\lim_{\alpha\downarrow 0} \alpha S_r(\alpha) = \lim_{\alpha\downarrow 0} \alpha S(\alpha).$$

Moreover,

$$\alpha S_r(\alpha) = \alpha \int_{t=0}^{\infty} e^{-\alpha t} s(t) e^{\alpha t} \mathbb{1}_{\{e^{-\alpha t} \geq e^{-1}\}} dt = \alpha \int_{t=0}^{1/\alpha} s(t)dt = \alpha S_{1/\alpha}$$

To complete the proof, we have

$$\lim_{\alpha\downarrow 0} \alpha S(\alpha) = \lim_{\alpha\downarrow 0} \alpha S_r(\alpha) = \lim_{\alpha\downarrow 0} \alpha S_{1/\alpha} = \lim_{T\to\infty} \frac{S_T}{T}.$$

□

Acknowledgement

Acknowledgement

Thank Sandjai for nice pictures

References

[1] I.J.B.F. Adan, V.G. Kulkarni, and A.C.C. van Wijk. Optimal control of a server farm. *INFOR*, 51(4):241–252, 2013.

[2] E. Altman, A. Hordijk, and F.M. Spieksma. Contraction conditions for average and α-discount optimality in countable Markov games with unbounded rewards. *Math. Operat. Res.*, 22:588–619, 1997.

[3] T.M. Apostol. *Mathematical analysis*. Addison Wesley Publishing Company, 1974.

[4] Y. Aviv and A. Federgruen. The value iteration method for countable state Markov decision processes. *Operat. Res. Let.*, 24:223–234, 1999.

[5] R. Bellman. A Markovian decision process. Technical report, DTIC Document, 1957.

[6] S. Bhulai, H. Blok, and F.M. Spieksma. Optimal control for K-competing queues with abandonments. In preparation, 2015.

[7] S. Bhulai, A.C. Brooms, and F.M. Spieksma. On structural properties of the value function for an unbounded jump Markov process with an application to a processor sharing retrial queue. *Queueing Syst.*, 76(4):425–446, 2014.

[8] S. Bhulai and G.M. Koole. On the structure of value functions for threshold policies in queueing models. *J. Appl. Prob.*, 40(3):613–622, 2003.

[9] S. Bhulai and F. M. Spieksma. On the uniqueness of solutions to the Poisson equations for average cost Markov chains with unbounded cost functions. *Math. Meth. Operat. Res.*, 58(2):221–236, 2003.

[10] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. J. Wiley & Sons, New York, 2d edition edition, 1999.

[11] H. Blok. Markov decision processes with unbounded transition rates: structural properties of the relative value function. Master's thesis, Utrecht University, 2011.

[12] H. Blok. *Parametrised Markov processes and optimal control (preliminary title)*. PhD thesis, Universteit Leiden, 2016.

[13] H. Blok and F.M. Spieksma. Countable state Markov decision processes with unbounded jump rates and discounted cost: optimality and approximations. *Adv. Appl. Prob.*, 40, 2015.

[14] H. Blok and F.M. Spieksma. Structural properties of the server farm model. In preparation, 2015.

[15] V.S. Borkar. *Topics in controlled Markov chains*. Longman Sc & Tech, 1991.

[16] R. Dekker and A. Hordijk. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Math. Operat. Res.*, 13:395–421, 1988.

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

[18] R. Dekker, A. Hordijk, and F.M. Spieksma. On the relation between recurrence and ergodicity properties in denumerable Markov decision chains. *Math. Operat. Res.*, 19:539–559, 1994.

[19] C. Derman. *Finite state Markovian decision processes*. Academic Press, New York, 1970.

[20] D.G. Down, G. Koole, and M.E. Lewis. Dynamic control of a single-server system with abandonments. *Queueing Syst.*, 67(1):63–90, 2011.

[21] E.A. Feinberg. Total reward criteria. In E.A. Feinberg and A. Shwartz, editors, *Handbook of Markov Decision Processes*, volume 40 of *International Series in Operations Research and Management Science*, chapter 5, pages 155–189. Kluwer Academic Publishers, Amsterdam, 2002.

[22] L. Fisher and S.M. Ross. An example in denumerable decision processes. *Ann. Math. Stat.*, 39(2):674–675, 1968.

[23] X. Guo and O. Hernández-Lerma. *Continuous-Time Markov Decision Processes*. Number 62 in Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2009.

[24] X. Guo, O. Hernández-Lerma, and T. Prieto-Rumeau. A survey of recent results on continuous-time markov decision processes. *Top*, 14(2):177–261, 2006.

[25] X.P. Guo and O. Hernández-Lerma. *Continuous-Time Markov Decision Processes*, volume 62 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, 2009.

[26] A. Hordijk. Regenerative Markov decision models. *Math. Program. Study*, pages 49–72, 1976.

[27] A. Hordijk, P.J. Schweitzer, and H.C. Tijms. The asymptotic behaviour of the minimal total expected cost for the denumerable state Markov decision model. *J. Appl. Prob.*, 12(298–305), 1975.

[28] M. Kitaev. Semi-Markov and jump Markov controlled models: average cost criterion. *Theory Prob. Appl.*, 30:272–288, 1986.

[29] G.M. Koole. *Monotonicity in Markov reward and decision chains: Theory and Applications*, volume 1. Now Publishers Inc., 2007.

[30] S.A. Lippman. On dynamic programming with unbounded rewards. *Management Science*, 21:1225–1233, 1975.

[31] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, Berlin, 1993.

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

[33] T. Prieto-Rumeau and O. Hernández-Lerma. Discounted continuous-time controlled Markov chains: convergence of control models. *J. Appl. Prob. Applied Probability*, 49(4):1072–1090, 2012.

[34] T. Prieto-Rumeau and O. Hernández-Lerma. *Selected topics on continuous-time controlled Markov chains and Markov games*, volume 5. World Scientific, 2012.

[35] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Programming*. J. Wiley & Sons, New Jersey, 2d edition edition, 2005.

[36] S.M. Ross. Non-discounted denumerable Markovian decision models. *Ann. Math. Statist.*, 39(2):412–423, 1968.

[37] H.L. Royden. *Real Analysis*. Macmillan Publishing Company, New York, 2d edition, 1988.

[38] L.I. Sennott. Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. *Operat. Res.*, 37(4):626–633, 1989.

[39] L.I. Sennott. Valute iteration in countable state average cost Markov decision processes with unbounded costs. *Ann. Oper. Res.*, 28:261–272, 1991.

[40] L.I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley Series in Probability and Statistics. Wiley, New York, 1999.

[41] R.F. Serfozo. An equivalence between continuous and discrete time Markov decision processes. *Operat. Res.*, 27(3):616–620, 1979.

[42] F.M. Spieksma. *Geometrically ergodic Markov Chains and the optimal Control of Queues*. PhD thesis, Leiden University, 1990. Available on request from the author.

[43] F.M. Spieksma. Countable state Markov processes: non-explosiveness and moment function. *Prob. Engin. Inform. Sci.*, 29(4):623–637, 2015.

[44] F.M. Spieksma. Parametrised countable state Markov processes in discrete and continuous time: drift conditions for exponential convergence. In preparation, 2015.

[45] R.E. Strauch. Negative Dynamic Programming. *Ann. Math. Stat.*, 37(4):871–890, 1966.

[46] E.C. Titchmarsh. *The Theory of Functions*. Oxford University Press, 2d edition, 1986.

[47] R.L. Tweedie. Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes. *J. Appl. Prob.*, pages 122–130, 1981.

[48] R.R. Weber and Sh. Stidham Jr. Optimal control of service rates in networks of queues. *Adv. Appl. Prob.*, 19(1):202–218, 1987.

[49] J. Wessels. Markov programming by successive approximations with respect to weighted supremum norms. *J. Math. An. Appl.*, 58(2):326–335, 1977.

[50] D.V. Widder. *The Laplace Transform*. Princeton University Press, London, 1941.