TAYLOR SERIES EXPANSIONS FOR
STATIONARY MARKOV CHAINS

BERND HEIDERGOTT,* *Vrije Universiteit Amsterdam and Tinbergen Institute*

ARIE HORDIJK,** *Leiden University*

Abstract

We study Taylor series expansions of stationary characteristics of general-state-space Markov chains. The elements of the Taylor series are explicitly calculated and a lower bound for the radius of convergence of the Taylor series is established. The analysis provided in this paper applies to the case where the stationary characteristic is given through an unbounded sample performance function such as the second moment of the stationary waiting time in a queueing system.

Keywords: Markov chain; Taylor series; measure-valued differentiation; deviation matrix

AMS 2000 Subject Classification: Primary 60J10
Secondary 90C31

1. Introduction

Stationary performance indices of simple Markovian systems, such as the queue length probabilities for an M/M/1 queue, are analytic functions of the system's parameters: the service rate and the interarrival rate. This observation has initiated the study of analytical properties of stochastic systems. The approach predominant in the literature is to study Taylor series expansions of performance characteristics of stochastic networks with Poisson-λ-arrival stream with respect to λ. First results can be found in [30] and [13] and extensions to (max,+) linear stochastic networks are studied in [1]–[6]. Taylor series expansions of the stationary distribution of finite-state Markov chains have been studied in [8]. Sample path approaches, which are based on simulation, can be found in [13], [21], [22] and [29]. An analytical approach to computing stationary characteristics of Markovian queueing systems based on Taylor series expansion is the power series algorithm (PSA); see [28].

In this paper, we study Taylor series expansions of stationary characteristics of general-state-space Markov chains. In particular, the approach applies to unbounded performance characteristics. The flexibility and generality of the framework used in this paper comprises the aforementioned approaches.

The results presented in this paper serve a twofold purpose. Firstly, elaborating on a Jordan-type decomposition of transition kernels [15], the elements of the Taylor series are expressed in terms of measure-valued derivatives and can be translated into unbiased (higher-order) derivative estimators; see [14] for details. In particular, this allows us to approximate an entire performance function through simulation. This links our results to sample-path based approaches like perturbation analysis; see [17] and [9]. Secondly, we derive lower bounds for

the radius of convergence of the Taylor series. Knowledge of the radius of convergence is of importance for numerical procedures such as PSA.

The mathematical framework of this paper is that of measure-valued differentiation (MVD) and normed ergodicity. MVD extends the concept of weak differentiability as introduced by Pflug [27] so that performance measures from a predefined class $\mathcal{D}$ can be handled, and thereby overcomes the restriction to bounded functions implicit in the concept of weak differentiation. Normed ergodicity dates back to the early 1980s; see [18] for an early reference. It was originally used in analysis of Blackwell optimality; see [18] and the revised version which was published as [11]. Another early reference is [23] and a more recent publication on this topic is [7]. Since then, it has been used in various forms under different names in many subsequent papers. In [20] it was shown that, for a countable Markov chain which may have one or several classes of essential states (a so-called multichained Markov chain), normed ergodicity is equivalent to geometrical recurrence (for a similar result in Markov decision chains, see [12]). Inspired by this result for a countable Markov chain, a similar result was proved for a Harris chain in [26]. In this paper we use the recent results of [7].

The paper is organized as follows. Section 2 states the main concepts and results needed from the theory of MVD and ergodic theory. In Section 3, higher-order measure-valued derivatives of the stationary distribution are discussed and Section 4 establishes our main result on Taylor series expansions of the stationary distribution. We conclude the paper with an elaborate example.

2. Tools from MVD and ergodic theory

2.1. MVD

Let $(S, \mathcal{T})$ be a Polish measurable space. Let $\mathcal{M}(S, \mathcal{T})$ denote the set of finite (signed) measures on $(S, \mathcal{T})$ and $\mathcal{M}_1(S, \mathcal{T})$ the set of probability measures on $(S, \mathcal{T})$.

Definition 1. The mapping $P : S \times \mathcal{T} \to [0, 1]$ is called a (homogeneous) *transition kernel* on $(S, \mathcal{T})$ if

   (a) $P(s; \cdot) \in \mathcal{M}(S, \mathcal{T})$ for all $s \in S$; and

   (b) $P(\cdot; B)$ is $\mathcal{T}$-measurable for all $B \in \mathcal{T}$.

If, in condition (a), $\mathcal{M}(S, \mathcal{T})$ can be replaced by $\mathcal{M}_1(S, \mathcal{T})$, then P is called a *Markov kernel* on $(S, \mathcal{T})$.

Denote the set of transition kernels on $(S, \mathcal{T})$ by $\mathcal{K}(S, \mathcal{T})$ and the set of Markov kernels on $(S, \mathcal{T})$ by $\mathcal{K}_1(S, \mathcal{T})$. The product of two transition kernels Q and P on $(S, \mathcal{T})$ is defined as follows. For $s \in S$ and $B \in \mathcal{T}$, set

$$QP(s; B) = (P \circ Q)(s; B) = \int_S Q(s; \mathrm{d}z) P(z; B).$$

Moreover, write $P^n(s; B)$ for the measure obtained by the n-fold product of P in the above way.

Let $L^1(P) \subset \mathbb{R}^S$ denote the set of measurable mappings $g : S \to \mathbb{R}$, such that $\int_S P(s; \mathrm{d}u)|g(u)|$ is finite for all $s \in S$. For $\mathcal{D} \subset L^1(P)$, we say that a transition kernel $P \in \mathcal{K}(S, \mathcal{T})$ is $\mathcal{D}$-*preserving* if

$$\int_S P(\cdot; \mathrm{d}u)g(u) \in \mathcal{D} \quad \text{For all } g \in \mathcal{D}.$$

Thus, if $P, Q \in \mathcal{K}(S, \mathcal{T})$ are $\mathcal{D}$-preserving, so is their product, and the integral of any $g \in \mathcal{D}$ with respect to a finite combination of P and Q is defined.

Consider a family of Markov kernels $(P_\theta : \theta \in \Theta)$ on $(S, \mathcal{T})$, with $\Theta \subset \mathbb{R}$, and let

$$L^1(P_\theta; \Theta) := \bigcap_{\theta \in \Theta} L^1(P_\theta)$$

denote the set of measurable mappings $g : S \to \mathbb{R}$ such that $\int_S P_\theta(s; \mathrm{d}u)|g(u)|$ is finite for all $\theta \in \Theta$ and $s \in S$.

Definition 2. Let $\mathcal{D} \subset L^1(P_\theta; \Theta)$. We say that $P_\theta \in \mathcal{K}(S, \mathcal{T})$ is *n times differentiable* at θ with respect to $\mathcal{D}$, or *n times $\mathcal{D}$-differentiable* for short, if there exists a $P_\theta^{(n)} \in \mathcal{K}(S, \mathcal{T})$ such that, for any $s \in S$ and any $g \in \mathcal{D}$,

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \int_S P_\theta(s; \mathrm{d}u)g(u) = \int_S P_\theta^{(n)}(s; \mathrm{d}u)g(u) \tag{1}$$

and $P_\theta^{(n)}$ is $\mathcal{D}$-preserving. If the left-hand side of (1) equals zero for all $g \in \mathcal{D}$, then we say that $P_\theta^{(n)}$ is not significant. For $n = 1$, we also write P_θ' for the transition kernel defined in (1). In case θ is a boundary point, the limit in (1) has to be understood as a one-sided limit.

We denote the set of bounded continuous mappings from S to $\mathbb{R}$ by $C_b(S)$. If $C_b(S) \subset \mathcal{D}$, then $P_\theta^{(n)}$ in (1) is uniquely defined. For this reason, *we will assume throughout the paper that $C_b(S) \subset \mathcal{D}$ if not stated otherwise.*

Remark 1. Note that $P_\theta^{(n)}$ and P_θ^n have a different meaning. To unify the notation, we set $P_\theta^{(0)} = P_\theta$.

Remark 2. Consider the pointwise version of the above definition, that is, define n fold $\mathcal{D}$-differentiability as follows: for any $s \in S$, a (finite) signed measure $P_\theta^{(n)}(s; \cdot)$ exists such that (1) holds for any $g \in \mathcal{D}$. While this definition of $P_\theta^{(n)}$ reflects the way in which $P_\theta^{(n)}$ is actually calculated in applications, it raises the question of whether $P_\theta^{(n)}$ is a kernel on $(S, \mathcal{T})$ (i.e. whether $P_\theta^{(n)}(\cdot; A)$ is measurable for any $A \in \mathcal{T}$). While this fails to be true in general, we can identify two conditions both enforcing this independently.

1. Let $\mathcal{D}$ contain the indicator function of any $A \in \mathcal{T}$. If $\mathcal{T}$ is countable, then $P_\theta^{(n)}$ is indeed a kernel; see [14]. If $\mathcal{D}$ contains in addition $C_b(S)$ as a subset, then it is called a set of test functions in [14].

2. Suppose that $C_b(S) \subset \mathcal{D}$. If

$$\sup_{s \in S} \sup_{\substack{g \in C_b(S) \\ |g| \le 1}} \left| \int_S P_\theta^{(n)}(s; \mathrm{d}u)g(u) \right| < \infty,$$

   then $P_\theta^{(n)}$ is a kernel; see [15].

Example 1. Let $P, Q \in \mathcal{K}_1(S, \mathcal{T})$ and set

$$P_\theta = \theta P + (1 - \theta)Q, \qquad \theta \in [0, 1].$$

Note that $P_\theta \in \mathcal{K}_1(S, \mathcal{T})$ for $\theta \in [0, 1]$, and that $P_0 = Q$ and $P_1 = P$. Specifically, let $\mathcal{D}(P, Q) := L^1(P_\theta; \Theta)$ and assume that P and Q are $\mathcal{D}(P, Q)$-preserving. For any $g \in \mathcal{D}(P, Q)$ and any $s \in S$,

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \int_S P_\theta(s; \mathrm{d}u)g(u) = \frac{\mathrm{d}^n}{\mathrm{d}\theta^n}\left(\theta \int_S P(s; \mathrm{d}u)g(u) + (1-\theta) \int_S Q(s; \mathrm{d}u)g(u)\right)$$

$$= \begin{cases} \int_S P(s; \mathrm{d}u)g(u) - \int_S Q(s; \mathrm{d}u)g(u) & \text{for } n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $P'_\theta = P - Q$ satisfies (1) for $n = 1$ and, since P and Q are $\mathcal{D}(P, Q)$-preserving, it follows that P'_θ is $\mathcal{D}(P, Q)$-preserving. Notice that the Markov kernel P_θ is infinitely often $\mathcal{D}(P, Q)$-differentiable and only its first-order $\mathcal{D}(P, Q)$ derivative is significant.

2.2. Normed ergodicity

Let $X(\theta) = \{X_\theta(n)\} = \{X_\theta(s, n)\}$ for $\theta \in \Theta$ be the Markov chain with initial state s and transition kernel P_θ, and set, for any $B \in \mathcal{T}$,

$$P_\theta^n(s, B) := P_\theta(s, n, B) = \Pr(X_\theta(s, n) \in B).$$

The joint state space of $X(\theta)$ for $\theta \in \Theta$ is denoted by S. However, for any specific θ, the chain $X(\theta)$ may not be irreducible on S but only on a subset of S. For the following ergodicity analysis we will require that the state space is indeed irreducible and we denote by $S_\theta \subset S$ the class of states such that $X(\theta)$ becomes irreducible as a Markov chain on S_θ. Furthermore, we denote by $\mathcal{T}_\theta$ the intersection of $\mathcal{T}$ and S_θ. Consequently, $(S_\theta, \mathcal{T}_\theta)$ is a measurable space for any $\theta \in \Theta$.

Condition 1. *There exists a function $g : S_\theta \to \mathbb{R}$ such that $g(s) \geq 0$ for $s \in S_\theta$ and*

$$E[g(X_\theta(s, m_\theta))] - g(s) \leq -\varepsilon + c\,\mathbf{1}_{V_\theta}(s)$$

for some $m_\theta \geq 1$, $\varepsilon > 0$ and $c < \infty$, where, for some $d < \infty$,

$$V_\theta = \{s \in S_\theta : g(s) \leq d\}.$$

Note that the function g is the same for each θ.

Condition 2. *There exist $n_\theta \geq 0$, a probability measure $\phi_\theta(\cdot)$ on $(S_\theta, \mathcal{T}_\theta)$, and $p_\theta \in (0, 1)$ such that*

$$\inf_{x \in V_\theta} \Pr(X_\theta(x, n_\theta) \in B) \geq p_\theta \phi_\theta(B)$$

for all $B \in \mathcal{T}_\theta$.

Assuming that Condition 1 holds for some $\theta \in \Theta$ and letting

$$\xi_\theta(s) := g(X_\theta(s, 1)) - g(s), \qquad s \in S_\theta,$$

we can introduce the following condition.

Condition 3. *The random variable $\xi_\theta(s)$ is uniformly integrable in s and in θ on a neighborhood Θ_0 of θ, and there exists a $\lambda > 0$ such that $\xi_\theta(s)e^{\lambda \xi_\theta(s)}$ is uniformly integrable (in s and in θ on Θ_0).*

Recall that uniform integrability of $\xi_\theta(s)$ in s and θ is defined as

$$\lim_{c \to \infty} \sup_{s,\theta} \int_{|t|>c} \Pr(\xi_\theta(s) \in \mathrm{d}t) = 0,$$

and similarly the uniform integrability of $\xi_\theta(s)\mathrm{e}^{\lambda\xi_\theta(s)}$ requires that

$$\lim_{c \to \infty} \sup_{s,\theta} \int_{|t|>c} \Pr(\xi_\theta(s)\mathrm{e}^{\lambda\xi_\theta(s)} \in \mathrm{d}t) = 0.$$

For a function $v : S \to \mathbb{R}$, let $\mathcal{V}_v$ denote the normed space of real-valued functions f on S with the finite v-norm

$$\|f\|_v = \sup_{s \in S} \frac{|f(s)|}{|v(s)|}.$$

The associated operator norm for a linear operator $T : \mathcal{V}_v \longrightarrow \mathcal{V}_v$ is defined by

$$\|T\|_v = \sup_{\|f\|_v \le 1} \|Tf\|_v.$$

For μ a (signed) measure, the associated norm is

$$\|\mu\|_v = \sup_{\|f\|_v \le 1} |\mu f|$$

and, for a kernel P, the associated norm reads

$$\|P\|_v = \sup_{s \in S} \sup_{\|f\|_v \le 1} \frac{|\int f(z) P(s; \mathrm{d}z)|}{|v(s)|}.$$

It is straightforward to check that, for kernels P, Q,

$$\|PQ\|_v \le \|P\|_v \|Q\|_v.$$

For our analysis, we choose v to be the following mapping:

$$v(s) := \mathrm{e}^{\lambda g(s)}, \qquad s \in S, \tag{2}$$

for some positive λ, where g is defined in Condition 1.

Let π_θ denote the stationary distribution under kernel P_θ. Denote by Π_θ the kernel that maps any distribution to π_θ, that is, $\mu\Pi_\theta = \pi_\theta$ for $\mu \in \mathcal{M}_1(S, \mathcal{T})$. In order to simplify the notation, we will—with a slight abuse of notation—identify π_θ with Π_θ.

The following result, which was established in [7] (see also [16]), shows that Conditions 1–3 imply that convergence of the Markov chain towards its stationary distribution happens at a geometrical rate. Specifically, the following results were established in [16].

Result 1. (Heidergott *et al.* [16, Theorem 2].) *Provided that Conditions 1–3 hold at θ, there exist a finite c_θ and a ρ_θ with $0 < \rho_\theta < 1$ such that, for λ small enough,*

$$\|P_\theta^n - \Pi_\theta\|_v \le c_\theta(\rho_\theta)^n.$$

Result 2. (Heidergott *et al.* [16, Lemma 1].) *If Conditions 1 and 3 hold at θ, then a neighborhood Θ_0 of θ exists such that*

$$\sup_{\hat{\theta} \in \Theta_0} \|P_{\hat{\theta}}\|_v < \infty.$$

Measure-valued derivatives are defined with respect to a predefined reference set $\mathcal{D}$. The set $\mathcal{D}$ that we will use for analyzing Π_θ is characterized through the function $v(\cdot)$ defined in (2), and will be denoted by $\mathcal{D}_v$. Specifically, let

$$\mathcal{D}_v := \{g{:}S \to \mathbb{R} \mid g \text{ is measurable and there is an } r \in \mathbb{R}$$
$$\text{such that } |g(s)| \leq r \cdot v(s) \text{ for } s \in S\}.$$

In other words, $\mathcal{D}_v$ is the set of measurable mappings g from S to $\mathbb{R}$ that are bounded by $r \cdot v$ for some finite number r. In accordance with the discussion following Definition 2, we assume that $C_b(S) \subset \mathcal{D}_v$.

Other choices for $\mathcal{D}$ are of course possible. For example, a choice for $\mathcal{D}$ that is of importance in applications is the set

$$\mathcal{C}_v := \{g \in \mathcal{D}_v \mid g \text{ is continuous}\}. \tag{3}$$

In particular, for $v \equiv 1$, we obtain the important case $\mathcal{D} = C_b(S)$; see [27]. In this paper, we will focus on $\mathcal{D} = \mathcal{D}_v$ and we will only briefly comment on the implications of another choice for $\mathcal{D}$ in Remark 10.

We require that $\mathcal{D}$ is a vector or linear space over $\mathbb{R}$, which is satisfied by $\mathcal{D}_v$. In addition to that, $\mathcal{D} = \mathcal{D}_v$ has to be a subset of $L^1(P_\theta : \Theta)$, a sufficient condition for which is given in the following corollary.

Corollary 1. *Let* $P_\theta \in \mathcal{K}(S, \mathcal{T})$. *If* $\|P_\theta\|_v < \infty$, *then* $\mathcal{D}_v \subset L^1(P_\theta, \Theta)$ *and* P_θ *is* $\mathcal{D}_v$-*preserving.*

Proof. By the definition of the norm $\| \cdot \|_v$, $\|P_\theta\|_v < \infty$ implies that, for any $g \in \mathcal{D}_v$,

$$\left| \int g(z) P_\theta(s; \mathrm{d}z) \right| < cv(s), \qquad s \in S,$$

for some finite number c, which shows that

$$\int g(z) P_\theta(\cdot; \mathrm{d}z) \in \mathcal{D}_v$$

and

$$\mathcal{D}_v \subset L^1(P_\theta, \Theta),$$

which concludes the proof. $\quad\blacksquare$

Remark 3. In accordance with Result 2, Condition 3 implies that $\|P_\theta\|_v < \infty$ and, by Corollary 1, this shows that P_θ is $\mathcal{D}_v$-preserving. Consequently, $\int g(z) P_\theta^m(s, \mathrm{d}z)$ is finite for any $m \geq 0$ and any $g \in \mathcal{D}_v$ and $s \in S$.

Definition 3. The transition kernel P_θ is said to be $\| \cdot \|_v$-*Lipschitz continuous* at $\theta \in \Theta$ if a neighborhood $\Theta_0 \subset \Theta$ of θ exists such that, for some finite K,

$$\|P_{\theta+\Delta} - P_\theta\|_v \leq |\Delta| K \quad \text{for all } \theta + \Delta \in \Theta_0.$$

The constant K is called the Lipschitz constant.

Remark 4. Let P_θ be $\|\cdot\|_v$-Lipschitz continuous at θ with Lipschitz constant K. This yields for any $g \in \mathcal{D}_v$ and Δ sufficiently small that

$$\left| \int g(z) P_{\theta+\Delta}(s; dz) - \int g(z) P_\theta(s; dz) \right| \leq \Delta K v(s).$$

Notice that $Kv(\cdot)$ lies in $\mathcal{D}_v$. Hence, by $\|\cdot\|_v$-Lipschitz continuity, for any $g \in \mathcal{D}_v$, a $K_g \in \mathcal{D}_v$ exists such that, for Δ sufficiently small,

$$\left| \int P_{\theta+\Delta}(\cdot; ds)g(s) - \int P_\theta(\cdot; ds)g(s) \right| \leq |\Delta| K_g(\cdot).$$

This property is called $\mathcal{D}_v$-Lipschitz continuity of P_θ at θ in [16] and we have thus shown that $\|\cdot\|_v$-Lipschitz continuity implies $\mathcal{D}_v$-Lipschitz continuity.

Theorem 1. *Let P_θ be a $\mathcal{D}_v$-preserving transition kernel on $(S, \mathcal{T})$ such that P_θ is $\|\cdot\|_v$-Lipschitz continuous and $\mathcal{D}$-differentiable at θ. If $\|P_\theta\|_v$ is finite, then P_θ^n is $\|\cdot\|_v$-Lipschitz continuous and $\mathcal{D}$-differentiable at θ. Specifically, the following product rule holds:*

$$(P_\theta^n)' = \sum_{j=1}^n P_\theta^{n-j-1} P_\theta' P_\theta^{j-1}.$$

Proof. Notice that

$$\begin{aligned}
P_{\theta+\Delta} P_{\theta+\Delta} - P_\theta P_\theta &= (P_{\theta+\Delta} - P_\theta) P_\theta \\
&\quad + P_\theta (P_{\theta+\Delta} - P_\theta) \\
&\quad + (P_{\theta+\Delta} - P_\theta)(P_{\theta+\Delta} - P_\theta).
\end{aligned} \tag{4}$$

Letting K denote the $\|\cdot\|_v$-Lipschitz constant of P_θ and taking $\|\cdot\|_v$ norms on both sides yields that

$$\begin{aligned}
\|P_{\theta+\Delta} P_{\theta+\Delta} - P_\theta P_\theta\|_v &\leq \|P_{\theta+\Delta} - P_\theta\|_v \|P_\theta\|_v \\
&\quad + \|P_\theta\|_v \|P_{\theta+\Delta} - P_\theta\|_v \\
&\quad + \|P_{\theta+\Delta} - P_\theta\|_v \|P_{\theta+\Delta} - P_\theta\| \\
&\leq |\Delta| K \|P_\theta\|_v + \|P_\theta\|_v |\Delta| K + |\Delta|^2 K^2.
\end{aligned}$$

Since we have assumed that $\|P_\theta\|_v$ is finite, the above inequalities establish the $\|\cdot\|_v$-Lipschitz continuity of P_θ^2 at θ.

We now turn to the proof of $\mathcal{D}_v$-differentiability. Notice that $\|\cdot\|_v$-Lipschitz continuity of P_θ at θ implies $\mathcal{D}_v$-Lipschitz continuity of P_θ at θ (see Remark 4), which implies that

$$|(P_{\theta+\Delta} - P_\theta)g| \leq |\Delta| K_g,$$

with $K_g \in \mathcal{D}_v$. Hence, dominated convergence together with $\mathcal{D}_v$-differentiability of P_θ yields that, for any $g \in \mathcal{D}_v$,

$$\lim_{\Delta \to 0} \frac{1}{\Delta} (P_{\theta+\Delta} - P_\theta) P_\theta g = P_\theta' P_\theta g$$

and

$$\lim_{\Delta \to 0} \frac{1}{\Delta} P_\theta (P_{\theta+\Delta} - P_\theta) g = P_\theta P_\theta' g$$

for the first two terms on the right-hand side of (4). We now turn to the last term in (4). By $\| \cdot \|_v$-Lipschitz continuity,

$$\|(P_{\theta+\Delta} - P_\theta)(P_{\theta+\Delta} - P_\theta)\|_v \leq |\Delta|^2 K.$$

Dividing both sides by $|\Delta|$, we obtain that

$$\lim_{\Delta \to 0} \frac{1}{|\Delta|} \|(P_{\theta+\Delta} - P_\theta)(P_{\theta+\Delta} - P_\theta)\|_v = 0.$$

This proves the statement for $n = 2$. The proof now follows by induction.

Remark 5. An alternative version of the product rule in Theorem 1 was established in [14]. The main difference between this product rule and the one in [14] is that Theorem 1 is based on norm assumptions whereas the product rule in [14] elaborates on the condition that $\mathcal{D}$ is a set of test functions (and thus contains all indicator mappings of sets $A \in \mathcal{T}$).

3. MVD of the stationary distribution

Let

$$K_\theta(n) = \sum_{m=0}^{\infty} P_\theta^{(n)} P_\theta^m.$$

The kernel $P_\theta^{(n)}$ can be written as the re-scaled difference between two Markov kernels (see [15]), which implies that $P_\theta^{(n)} \Pi_\theta = 0$. The operator $K_\theta(n)$ can thus be written as

$$K_\theta(n) = P_\theta^{(n)} D_\theta,$$

where

$$D_\theta = \sum_{m=0}^{\infty} (P_\theta^m - \Pi_\theta)$$

denotes the *deviation operator* associated with P_θ.

Remark 6. Let Conditions 1–3 hold at θ. Result 1 implies that

$$\|D_\theta\|_v = \left\| \sum_{m=0}^{\infty} (P_\theta^m - \Pi_\theta) \right\|_v \leq c_\theta \frac{1}{1 - \rho_\theta}$$

and, by Corollary 1, the transition kernel D_θ is $\mathcal{D}_v$-preserving. Specifically, $\int g(z) D_\theta(s, \mathrm{d}z)$ is finite for any $g \in \mathcal{D}_v$ and $s \in S$.

We introduce the following condition.

Condition 4. *For $\theta \in \Theta$,*

$$\lim_{\Delta \to 0} \|P_{\theta+\Delta} - P_\theta\|_v = 0. \tag{5}$$

In other words, Condition 4 implies that P_θ is continuous at θ with respect to the norm $\| \cdot \|_v$. In case θ is a boundary point, we consider a one-sided limit in (5). Notice that $\| \cdot \|_v$-Lipschitz continuity of P_θ at θ is sufficient for Condition 4 to hold.

Example 2. We revisit the situation in Example 1 and assume that $\|P\|_v < \infty$ and $\|Q\|_v < \infty$. It is easily checked that, for any $\theta + \Delta \in [0, 1]$,

$$\|P_{\theta+\Delta} - P_\theta\|_v = |\Delta| \|P - Q\|_v < \infty,$$

which shows that P_θ is $\|\cdot\|_v$-Lipschitz at θ for any $\theta \in [0, 1]$, and Condition 4 is thus satisfied on the entire interval $[0, 1]$.

Lemma 1. *Let Conditions 1–3 be satisfied at θ.*

(i) *There exists a neighborhood Θ_0 of θ such that, for all $\theta + \Delta \in \Theta_0$,*

$$\Pi_{\theta+\Delta} - \Pi_\theta = \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta. \tag{6}$$

(ii) *If, in addition, Condition 4 is satisfied at θ, then*

$$\Pi_{\theta+\Delta} = \Pi_\theta \sum_{m=0}^{\infty} ((P_{\theta+\Delta} - P_\theta)D_\theta)^m$$

and

$$\sup_{\hat\theta \in \Theta_0} \|\Pi_{\hat\theta}\|_v < \infty. \tag{7}$$

Proof. For θ we have by Result 1 that

$$\|P_\theta^n - \Pi_\theta\|_v \le c_\theta \rho_\theta^n. \tag{8}$$

By Result 2, $\|P_\theta\|_v < \infty$ and we find with the triangle norm inequality that

$$\|\Pi_\theta\|_v \le \|P_\theta - \Pi_\theta\|_v + \|P_\theta\|_v < \infty. \tag{9}$$

Let I denote the identity kernel. It is easily seen that, for any k,

$$(I - P_\theta)\left(\sum_{l=0}^{k-1} P_\theta^l\right) = (I - P_\theta^k).$$

If $\theta + \Delta \in \Theta$, then

$$\Pi_{\theta+\Delta}(I - P_\theta)\left(\sum_{l=0}^{k-1} P_\theta^l\right) = \Pi_{\theta+\Delta}(I - P_\theta^k)$$

$$= \Pi_{\theta+\Delta} - \Pi_\theta P_\theta^k; \tag{10}$$

the inequalities (9) yield that the expressions in 10 are well defined. Since P_θ is unichained, we have that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$ and we obtain from 10 that

$$\Pi_{\theta+\Delta} - \Pi_\theta = \Pi_{\theta+\Delta} - \Pi_{\theta+\Delta} P_\theta^k + \Pi_{\theta+\Delta}(P_\theta^k - \Pi_\theta)$$

$$= \Pi_{\theta+\Delta}(I - P_\theta)\left(\sum_{l=0}^{k-1} P_\theta^l\right) + \Pi_{\theta+\Delta}(P_\theta^k - \Pi_\theta)$$

$$= \Pi_{\theta+\Delta}(I - P_\theta)\left(\sum_{l=0}^{k-1} P_\theta^l - \Pi_\theta\right) + \Pi_{\theta+\Delta}(P_\theta^k - \Pi_\theta).$$

We have already established the geometric ergodicity of P_θ (see (8)), and taking the limit as k tends to ∞ we find that

$$\Pi_{\theta+\Delta} - \Pi_\theta = \Pi_{\theta+\Delta}(I - P_\theta)\left(\sum_{l=0}^\infty P_\theta^l - \Pi_\theta\right),$$

which proves (i).

By Remark 6, $\|D_\theta\|_v < \infty$, and Condition 4 gives

$$\lim_{\Delta \to \theta} \|(P_{\theta+\Delta} - P_\theta)D_\theta\|_v = 0.$$

Choose $\delta > 0$ such that $\|(P_{\theta+\Delta} - P_\theta)D_\theta\|_v = \varepsilon < 1$ for $|\Delta| < \delta$ then, for $g \in \mathcal{D}$,

$$\frac{g}{\|g\|_v} \in \mathcal{D}_1, \quad \text{with } \mathcal{D}_1 = \{g \in \mathcal{D}_v : \|g\|_v \le 1\},$$

and

$$\|(P_{\theta+\Delta} - P_\theta)D_\theta g\|_v \le \|(P_{\theta+\Delta} - P_\theta)D_\theta\|_v \|g\|_v \le \varepsilon\|g\|_v.$$

Since

$$\frac{(P_{\theta+\Delta} - P_\theta)D_\theta g}{\|(P_{\theta+\Delta} - P_\theta)D_\theta g\|_v} \in \mathcal{D}_1,$$

we have

$$\|((P_{\theta+\Delta} - P_\theta)D_\theta)^2 g\|_v \le \|(P_{\theta+\Delta} - P_\theta)D_\theta\|_v \|(P_{\theta+\Delta} - P_\theta)D_\theta g\|_v$$
$$\le \varepsilon^2\|g\|_v.$$

Similarly,

$$\|((P_{\theta+\Delta} - P_\theta)D_\theta)^k g\|_v \le \varepsilon^k\|g\|_v.$$

Hence, the Neumann series

$$\sum_{m=0}^\infty ((P_{\theta+\Delta} - P_\theta)D_\theta)^m g$$

is convergent and the series can be bounded through

$$\left\|\sum_{m=0}^\infty ((P_{\theta+\Delta} - P_\theta)D_\theta)^m g\right\|_v \le \frac{1}{1 - \varepsilon}\|g\|_v. \tag{11}$$

Elaborating on the equality (6) we find that

$$\Pi_\theta = \Pi_{\theta+\Delta}(I - (P_{\theta+\Delta} - P_\theta)D_\theta)$$

and it follows from our choice of δ that, when $|\Delta| < \delta$,

$$\Pi_{\theta+\Delta} = \Pi_\theta \sum_{m=0}^\infty ((P_{\theta+\Delta} - P_\theta)D_\theta)^m. \tag{12}$$

The proof now follows directly from the above expansion for $\Pi_{\theta+\Delta}$ together with (11). Indeed, for $g \in \mathcal{D}_v$ and $\theta + \Delta$ with $|\Delta| < \delta$,

$$\|\Pi_{\theta+\Delta} g\|_v \leq \|\Pi_\theta\|_v \left\| \sum_{m=0}^{\infty} ((P_{\theta+\Delta} - P_\theta) D_\theta)^m g \right\|_v$$

$$\leq \|\Pi_\theta\|_v \frac{1}{1 - \varepsilon} \|g\|_v$$

and

$$\sup_{\hat{\theta} \in \Theta} \|\Pi_{\hat{\theta}} g\|_v \leq \|\Pi_\theta\|_v \frac{1}{1 - \varepsilon},$$

for $\Theta \subset (\theta - \delta, \theta + \delta)$, which proves the claim.

Remark 7. Inspecting the proof of Lemma 1, we can notice that, if Conditions 1–4 hold at θ, then

$$\Pi_{\hat{\theta}} = \Pi_\theta \sum_{m=0}^{\infty} ((P_{\hat{\theta}} - P_\theta) D_\theta)^m \tag{13}$$

for $\hat{\theta}$ in a neighborhood of θ; cf. (12) above. This relation is well known in Markov decision chains where it is called a Newton-type representation; see Proposition 3.3(i) in [19].

Remark 8. Under the conditions put forward in Lemma 1, applying Corollary 1 yields that Π_θ is $\mathcal{D}_v$-preserving. Consequently, for any $s \in S$, $\int g(z) \Pi_\theta(s, dz)$ is finite for any $g \in \mathcal{D}_v$, and so is $\int g(z) \pi_\theta(dz)$.

As the following lemma shows, Conditions 1–3 together with $\| \cdot \|_v$-continuity of P_θ are sufficient for uniform boundedness of $\|P_\theta^n - \Pi_\theta\|_v$.

Lemma 2. *Let Conditions 1–3 together with Condition 4 hold at θ. Then there exist a finite $c < \infty$, a ρ with $0 < \rho < 1$ and a neighborhood Θ_0 of θ such that, for λ small enough,*

$$\sup_{\hat{\theta} \in \Theta_0} \|P_{\hat{\theta}}^n - \Pi_{\hat{\theta}}\|_v \leq c\rho^n. \tag{14}$$

Proof. Let $\theta + \Delta \in \Theta$. Using norm inequalities we find that

$$\|P_{\theta+\Delta}^n - \Pi_{\theta+\Delta}\|_v \leq \|P_{\theta+\Delta}^n - P_\theta^n\|_v + \|P_\theta^n - \Pi_\theta\|_v + \|\Pi_\theta - \Pi_{\theta+\Delta}\|_v.$$

We bound the three terms on the right-hand side of this inequality. From Result 2, we may choose Θ such that, when $1 \leq K < \infty$,

$$\sup_{\hat{\theta} \in \Theta} \|P_{\hat{\theta}}\|_v \leq K.$$

It is easily verified that the following telescope sum holds:

$$P_{\theta+\Delta}^n - P_\theta^n = \sum_{k=1}^{n} P_{\theta+\Delta}^{k-1} (P_{\theta+\Delta} - P_\theta) P_\theta^{n-k}.$$

This gives the norm inequality

$$\|P_{\theta+\Delta}^n - P_\theta^n\|_v \le \|P_{\theta+\Delta} - P_\theta\|_v \sum_{k=0}^{n-1} \|P_{\theta+\Delta}^k\|_v \|P_\theta^{n-1-k}\|_v$$

$$\le \|P_{\theta+\Delta} - P_\theta\|_v \sum_{k=0}^{n-1} \|P_{\theta+\Delta}\|_v^k \|P_\theta\|_v^{n-1-k}$$

$$\le \|P_{\theta+\Delta} - P_\theta\|_v n K^n.$$

Choose an $\varepsilon < 1$. Result 1 implies that

$$\|P_\theta^n - \Pi_\theta\|_v \le c_\theta \rho_\theta^n.$$

Choose an n such that $c_\theta \rho_\theta^n < \varepsilon/3$. From Lemma 1 we find that there exists a $\delta > 0$ such that, for $|\Delta| < \delta$,

$$\|\Pi_\theta - \Pi_{\theta+\Delta}\|_v \le \frac{\varepsilon}{3}$$

and

$$\|P_{\theta+\Delta} - P_\theta\|_v \le \frac{\varepsilon}{3n K^n}.$$

Then, for

$$\Theta_0 := \{\hat{\theta} : |\hat{\theta} - \theta| < \delta\},$$

we find from the above inequalities that

$$\rho_1 := \sup_{\hat{\theta} \in \Theta_0} \|P_{\hat{\theta}}^n - \Pi_{\hat{\theta}}\|_v < 1.$$

Notice that $P_{\hat{\theta}}^m \Pi_{\hat{\theta}} = \Pi_{\hat{\theta}} P_{\hat{\theta}}^m = \Pi_{\hat{\theta}}$ implies that

$$P_{\hat{\theta}}^m - \Pi_{\hat{\theta}} = (P_{\hat{\theta}} - \Pi_{\hat{\theta}})^m. \tag{15}$$

Again using norm inequalities we obtain for $m \ge 1$ that

$$\|P_{\hat{\theta}}^m - \Pi_{\hat{\theta}}\|_v = \|(P_{\hat{\theta}} - \Pi_{\hat{\theta}})^m\|_v$$

$$= \|(P_{\hat{\theta}} - \Pi_{\hat{\theta}})^{n\lfloor m/n \rfloor + m - n\lfloor m/n \rfloor}\|_v$$

$$\le (\|(P_{\hat{\theta}} - \Pi_{\hat{\theta}})^n\|_v)^{\lfloor m/n \rfloor} \|(P_{\hat{\theta}} - \Pi_{\hat{\theta}})^{m - n\lfloor m/n \rfloor}\|_v$$

$$= (\|P_{\hat{\theta}}^n - \Pi_{\hat{\theta}}\|_v)^{\lfloor m/n \rfloor} \|P_{\hat{\theta}}^{m - n\lfloor m/n \rfloor} - \Pi_{\hat{\theta}}\|_v$$

$$\le \rho_1^{\lfloor m/n \rfloor} \max_{k=0,\dots,n-1} \|P_{\hat{\theta}}^k - \Pi_{\hat{\theta}}\|_v, \tag{16}$$

the first and third equalities following from (15) and $\lfloor \cdot \rfloor$ being the integer-part function. The inequality (14) with

$$c := \sup_{\hat{\theta} \in \Theta_0} \max_{k=0,\dots,n-1} \|P_{\hat{\theta}}^k - \Pi_{\hat{\theta}}\|_v$$

and $\rho := \rho_1^{1/n}$ follows (16).

The following theorem establishes a recurrence relation for the $\mathcal{D}_v$-derivatives of $K_\theta(n)$.

Theorem 2. *Assume that Conditions 1–3 are satisfied at θ. If P_θ is $\|\cdot\|_v$-Lipschitz continuous at θ, then Π_θ is $\|\cdot\|_v$-Lipschitz continuous at θ. If, in addition, P_θ is $\mathcal{D}_v$-differentiable at θ, then Π_θ is $\mathcal{D}_v$-differentiable at θ with*

$$\Pi'_\theta = \Pi_\theta K_\theta(1).$$

Proof. We elaborate on Lemma 1(i):

$$\Pi_{\theta+\Delta} - \Pi_\theta = \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

implies that

$$\|\Pi_{\theta+\Delta} - \Pi_\theta\|_v \leq \|\Pi_{\theta+\Delta}\|_v \|(P_{\theta+\Delta} - P_\theta)D_\theta\|_v.$$

Let P_θ be $\|\cdot\|_v$-Lipschitz continuous at θ with constant K. The conditions of the theorem then imply that $\|D_\theta\|_v < \infty$; see Remark 6. We have

$$\|\Pi_{\theta+\Delta} - \Pi_\theta\|_v \leq \|\Pi_{\theta+\Delta}\|_v |\Delta| K \|D_\theta\|_v$$

and, because $\sup_{\hat{\theta}\in\Theta} \|\Pi_{\hat{\theta}}\|_v < \infty$ (see Lemma 1), the $\|\cdot\|_v$-Lipschitz continuity of Π_θ follows.

We now turn to the $\mathcal{D}_v$-differentiability of Π_θ. Again by Lemma 1,

$$\begin{aligned}
\Pi_{\theta+\Delta} - \Pi_\theta &= \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta \\
&= (\Pi_{\theta+\Delta} - \Pi_\theta)(P_{\theta+\Delta} - P_\theta)D_\theta + \Pi_\theta(P_{\theta+\Delta} - P_\theta)D_\theta.
\end{aligned}$$

By the $\|\cdot\|_v$-Lipschitz continuity of Π_θ and P_θ, we have

$$\|(\Pi_{\theta+\Delta} - \Pi_\theta)(P_{\theta+\Delta} - P_\theta)D_\theta\|_v \leq |\Delta|^2 K$$

for some finite number K. Hence, for any $g \in \mathcal{D}_v$,

$$\lim_{\theta\to\theta} \frac{1}{\Delta}(\Pi_{\theta+\Delta} - \Pi_\theta)(P_{\theta+\Delta} - P_\theta)D_\theta g = 0.$$

By the $\mathcal{D}_v$-Lipschitz continuity of P_θ (which is implied by the $\|\cdot\|_v$-Lipschitz continuity of P_θ; see Remark 4), we obtain for any $g \in \mathcal{D}_v$ that

$$\frac{1}{|\Delta|} |(P_{\theta+\Delta} - P_\theta)D_\theta g| \leq K_g,$$

with $K_g \in \mathcal{D}_v$. The $\mathcal{D}_v$-differentiability of P_θ implies that, for any $g \in \mathcal{D}_v$, the limit

$$\lim_{\Delta\to 0} \frac{1}{\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta g = P'_\theta D_\theta g$$

exists and is again in $\mathcal{D}_v$. Since $\Pi_\theta K_g$ is finite (from (7) in Lemma 1(ii) together with Corollary 1), the dominated convergence theorem applies and we obtain that

$$\lim_{\Delta\to 0} \frac{1}{\Delta}\Pi_\theta(P_{\theta+\Delta} - P_\theta)D_\theta g = \Pi_\theta P'_\theta D_\theta g.$$

We have thus shown that

$$\Pi'_\theta = \Pi_\theta P'_\theta D_\theta = \Pi_\theta K_\theta(1),$$

which concludes the proof of the induction hypothesis.

Remark 9. In [16], differentiability of Π_θ with respect to an arbitrary set $\mathcal{D}$ was established under a different set of conditions. Studying the relation between the settings in [16] and those chosen for the present analysis is a topic of further research.

The following theorem will establish sufficient conditions for $\mathcal{D}_v$-differentiability of $K_\theta(n)$. To simplify the presentation, we summarize the set of conditions needed for the theorem by a single condition, Condition C^n, where $n \geq 0$.

Condition C^n. (i) *The kernel P_θ is $n + 1$ times $\mathcal{D}_v$-differentiable at θ.*

(ii) *For some $r \geq 0$, the neighborhood $\Theta_0 = (\theta - r, \theta + r)$ of θ is such that, when $0 \leq m \leq n$, $\|P_\theta^{(m)}\|_v < \infty$ and $P_\theta^{(m)}$ is $\|\cdot\|_v$-Lipschitz continuous at θ.*

Example 3. We revisit the situation in Example 1 and assume that $\|P\|_v < \infty$ and $\|Q\|_v < \infty$. In accordance with Example 1, P_θ is infinitely often $\mathcal{D}_v$-differentiable at any $\theta \in [0, 1]$ (with only the first-order $\mathcal{D}_v$-derivative significant) and

$$\|P_\theta'\|_v \leq \|P\|_v + \|Q\|_v < \infty.$$

Moreover, by Example 2, P_θ is $\|\cdot\|_v$-Lipschitz continuous at any $\theta \in [0, 1]$ and $\mathcal{D}_v$-preserving. Because any (higher-order) $\mathcal{D}_v$-derivative of P_θ is independent of θ, they are $\|\cdot\|_v$-Lipschitz continuous at any $\theta \in [0, 1]$. Hence, Condition C^n is satisfied for any n on the entire interval $[0, 1]$.

Theorem 3. *If Conditions 1–3 and C^n for some $n \geq 1$ are satisfied at θ then $K_\theta(n)$ is $\mathcal{D}_v$-differentiable with*

$$K_\theta(n)' = K_\theta(n)K_\theta(1) + K_\theta(n + 1).$$

Proof. Let $Q_\theta = P_\theta^{(n)}$. Then

$$K_\theta(n) = \lim_{M \to \infty} \sum_{m=0}^{M} Q_\theta P_\theta^m.$$

By Theorem 1,

$$\left(\sum_{m=0}^{M} Q_\theta P_\theta^m \right)' = \sum_{m=0}^{M} Q_\theta' P_\theta^m + \sum_{m=0}^{M} \sum_{j=0}^{m} Q_\theta P_\theta^{m-j-1} P_\theta' P_\theta^j$$

and, taking the limit as M tends to infinity,

$$\lim_{M \to \infty} \left(\sum_{m=0}^{M} Q_\theta P_\theta^m \right)' = \sum_{m=0}^{\infty} Q_\theta' P_\theta^m + \sum_{m=0}^{\infty} \sum_{j=0}^{m} Q_\theta P_\theta^{m-j-1} P_\theta' P_\theta^j.$$

Rearranging terms gives

$$\lim_{M \to \infty} \left(\sum_{m=0}^{M} Q_\theta P_\theta^m \right)' = K_\theta(n + 1) + K_\theta(n)K_\theta(1). \tag{17}$$

It remains to be shown that

$$\lim_{M \to \infty} \left(\sum_{m=0}^{M} Q_\theta P_\theta^m g \right)' = K_\theta(n)' g, \qquad g \in \mathcal{D}_v.$$

To see this, note that, for any $M > 0$,

$$K_\theta(n)' = \left(\sum_{m=0}^{M} Q_\theta P_\theta^m \right)' + \left(\sum_{m>M} Q_\theta P_\theta^m \right)' \tag{18}$$

and the proof of the claim thus follows from

$$\lim_{M \to \infty} \left(\sum_{m>M} Q_\theta P_\theta^m \right)' = 0. \tag{19}$$

By Condition C^n, we may choose Θ_0 small enough such that

$$\sup_{\hat{\theta} \in \Theta_0} \|Q_{\hat{\theta}}\|_v = K_1 < \infty, \tag{20}$$

which is a consequence of the $\|\cdot\|_v$-Lipschitz continuity of $Q_\theta = P_\theta^{(n)}$. Indeed, since $\|Q_\theta\|_v$ is finite, $\|\cdot\|_v$-Lipschitz continuity of Q_θ at θ with constant K on, say, Θ_0, implies for $\theta + \Delta \in \Theta_0$ that

$$\|Q_{\theta+\Delta}\|_v = \|Q_{\theta+\Delta} - Q_\theta + Q_\theta\|_v \le \|Q_\theta\|_v |\Delta| K < \infty$$

and (20) holds for $K_1 = K \|Q_\theta\|_v$, where we assume without loss of generality that $|\Delta| \le 1$.

We consider the difference quotient for any $\theta + \Delta \in \Theta_0$, the neighborhood as defined in Lemma 2. Since, for $m \ge 1$ and each $\hat{\theta} \in \Theta_0$,

$$\sum_{m>M} Q_{\hat{\theta}} P_{\hat{\theta}}^m = \sum_{m>M} Q_{\hat{\theta}} (P_{\hat{\theta}}^m - \Pi_{\hat{\theta}}) = \sum_{m>M} Q_{\hat{\theta}} (P_{\hat{\theta}} - \Pi_{\hat{\theta}})^m,$$

we have that

$$\begin{aligned}
\Delta(M) &:= \sum_{m>M} Q_{\theta+\Delta} P_{\theta+\Delta}^m - \sum_{m>M} Q_\theta P_\theta^m \\
&= \sum_{m>M} Q_{\theta+\Delta} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} Q_\theta (P_\theta - \Pi_\theta)^m \\
&= \sum_{m>M} Q_{\theta+\Delta} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} Q_{\theta+\Delta} (P_\theta - \Pi_\theta)^m \\
&\quad + \sum_{m>M} Q_{\theta+\Delta} (P_\theta - \Pi_\theta)^m - \sum_{m>M} Q_\theta (P_\theta - \Pi_\theta)^m.
\end{aligned}$$

We derive bounds for the difference of the first and the second terms and for the difference of the third and fourth terms in the right-hand side of this equality. Using (20), we find that

$$\left\| \sum_{m>M} Q_{\theta+\Delta} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} Q_{\theta+\Delta} (P_\theta - \Pi_\theta)^m \right\|_v$$

$$\le K_1 \left\| \sum_{m>M} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} (P_\theta - \Pi_\theta)^m \right\|_v.$$

By Theorem 2, $\Pi_{\theta+\Delta}$ is $\|\cdot\|_v$-Lipschitz continuous at θ with constant K_2, say. Together with the $\|\cdot\|_v$-Lipschitz continuity of $P_{\theta+\Delta}$ with constant, say, K_3, we obtain from the above inequality that

$$\left\| \sum_{m>M} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} (P_\theta - \Pi_\theta)^m \right\|_v$$

$$= \left\| \sum_{m>M} ((P_{\theta+\Delta} - P_\theta) - (\Pi_{\theta+\Delta} - \Pi_\theta)) \sum_{k=0}^{m-1} (P_{\theta+\Delta} - \Pi_{\theta+\Delta})^k (P_\theta - \Pi_\theta)^{m-1-k} \right\|_v$$

$$\leq \|(P_{\theta+\Delta} - P_\theta) - (\Pi_{\theta+\Delta} - \Pi_\theta)\|_v \sum_{m>M} cm\rho^m$$

$$\leq |\Delta|(K_2 + K_3)K_1(M),$$

where the first inequality follows from Lemma 2 and where

$$K_1(M) = \sum_{m>M} cm\rho^m.$$

Note that $\lim_{M\to\infty} K_1(M) = 0$. Hence,

$$\left\| \sum_{m>M} Q_{\theta+\Delta}(P_{\theta+\Delta} - \Pi_{\theta+\Delta})^m - \sum_{m>M} Q_{\theta+\Delta}(P_\theta - \Pi_\theta)^m \right\|_v \leq |\Delta| K_1 (K_2 + K_3)K_1(M).$$

Similarly, with the $\|\cdot\|_v$-Lipschitz continuity of $Q_{\theta+\Delta}$, say

$$\|Q_{\theta+\Delta} - Q_\theta\|_v \leq |\Delta|K_4,$$

we have that

$$\left\| \sum_{m>M} Q_{\theta+\Delta}(P_\theta - \Pi_\theta)^m - \sum_{m>M} Q_\theta(P_\theta - \Pi_\theta)^m \right\|_v \leq |\Delta| K_4 (K_2 + K_3) \sum_{m>M} c\rho^m.$$

We conclude that

$$\Delta(M) \leq |\Delta| K(M),$$

with

$$K(M) := (K_4 + K_1)(K_2 + K_3)\left(c\frac{\rho^{M+1}}{1-\rho} + K_1(M) \right),$$

and

$$\lim_{M\to\infty} K(M) = 0,$$

which establishes (19). Taking limits in (18),

$$K_\theta(n)' = \lim_{M\to\infty} \left(\sum_{m\leq M} Q_\theta P_\theta^m \right)'$$

$$= K_\theta(n+1) + K_\theta(n)K_\theta(1),$$

where the last equality follows from (17).

We come to the main result of this section, namely that Π_θ is n times $\mathcal{D}_v$-differentiable under appropriate conditions and $\Pi_\theta^{(n)}$ can be obtained in a closed-form expression.

Theorem 4. *Assume that Conditions 1–3 and C^n for some $n \geq 1$ are satisfied at θ. Then Π_θ is n times $\mathcal{D}_v$-differentiable with*

$$\Pi_\theta^{(n)} = \sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq n \\ l_1 + \cdots + l_m = n}} \frac{n!}{l_1! \cdots l_m!} \Pi_\theta \prod_{k=1}^{m} K_\theta(l_k),$$

or, equivalently,

$$\Pi_\theta^{(n)} = \sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq n \\ l_1 + \cdots + l_m = n}} \frac{n!}{l_1! \cdots l_m!} \Pi_\theta \prod_{k=1}^{m} (P_\theta^{(l_k)} D_\theta).$$

Proof. Notice that $K_\theta(m)$ is a transition kernel for $m \leq n$.
We give a proof by induction. For $n = 1$, the proof follows from Theorem 2.
For $n = 2$, we combine Theorems 2 and 1 as follows:

$$\begin{aligned}
\Pi_\theta^{(2)} &= (\Pi_\theta K_\theta(1))' \\
&= \Pi_\theta' K_\theta(1) + \Pi_\theta K_\theta(1)' \\
&= \Pi_\theta K_\theta(1) K_\theta(1) + \Pi_\theta K_\theta(1)',
\end{aligned}$$

and, invoking Theorem 3, this is equal to

$$\Pi_\theta K_\theta(1) K_\theta(1) + \Pi_\theta K_\theta(1) K_\theta(1) + \Pi_\theta K_\theta(2) = 2\Pi_\theta K_\theta(1) K_\theta(1) + \Pi_\theta K_\theta(2).$$

To continue the induction, suppose that the statement of the theorem is true for n. Taking the $(n + 1)$th derivative of Π_θ yields

$$\Pi_\theta^{(n+1)} = \sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq n \\ l_1 + \cdots + l_m = n}} \frac{n!}{l_1! \cdots l_m!} \left(\Pi_\theta \prod_{k=1}^{m} K_\theta(l_k) \right)'.$$

For given $l = (l_1, l_2, \ldots, l_m)$, Theorem 1 gives

$$\left(\Pi_\theta \prod_{k=1}^{m} K_\theta(l_k) \right)' = \Pi_\theta' \prod_{k=1}^{m} K_\theta(l_k) + \Pi_\theta \left(\prod_{k=1}^{m} K_\theta(l_k) \right)'.$$

By Theorem 1 together with Theorem 2,

$$\Pi_\theta' \prod_{k=1}^{m} K_\theta(l_k) = \Pi_\theta K_\theta(1) \prod_{k=1}^{m} K_\theta(l_k)$$

and

$$\left(\prod_{k=1}^{m} K_\theta(l_k) \right)' = \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_\theta(l_k)(K_\theta(l_j))' \prod_{k=j+1}^{m} K_\theta(l_k),$$

where we set $\prod_{\alpha}^{\beta} = 1$ if $\alpha > \beta$. Inserting the explicit formula for $K_{\theta}(l_j)'$ gives

$$\left(\prod_{k=1}^{m} K_{\theta}(l_k)\right)' = \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_{\theta}(l_k)(K_{\theta}(l_j)K_{\theta}(1) + K_{\theta}(l_j + 1)) \prod_{k=j+1}^{m} K_{\theta}(l_k).$$

Hence,

$$\left(\Pi_{\theta} \prod_{k=1}^{m} K_{\theta}(l_k)\right)' = \Pi_{\theta} K_{\theta}(1) \prod_{k=1}^{m} K_{\theta}(l_k)$$

$$+ \Pi_{\theta} \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_{\theta}(l_k)K_{\theta}(l_j)K_{\theta}(1) \prod_{k=j+1}^{m} K_{\theta}(l_k)$$

$$+ \Pi_{\theta} \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_{\theta}(l_k)K_{\theta}(l_j + 1) \prod_{k=j+1}^{m} K_{\theta}(l_k).$$

Observe that a vector $(l_1, l_2, \ldots, l_m)$ is translated by

$$\Pi_{\theta} K_{\theta}(1) \prod_{k=1}^{m} K_{\theta}(l_k)$$

into $(1, l_1, l_2, \ldots, l_m)$, by

$$\Pi_{\theta} \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_{\theta}(l_k)K_{\theta}(l_j)K_{\theta}(1) \prod_{k=j+1}^{m} K_{\theta}(l_k)$$

into the vectors $(l_1, \ldots, l_j, 1, l_{j+1}, l_m)$ for $1 \leq j \leq m$, and by

$$\Pi_{\theta} \sum_{j=1}^{m} \prod_{k=1}^{j-1} K_{\theta}(l_k)K_{\theta}(l_j + 1) \prod_{k=j+1}^{m} K_{\theta}(l_k)$$

into the vectors $(l_1, \ldots, l_{j-1}, l_j + 1, l_{j+1}, l_m)$ for $1 \leq j \leq m$. Because

$$\sum_{k=1}^{n} l_k = n,$$

it follows that the elements of the new vectors add up to $n + 1$. Moreover, by the above procedure, any vector $\tilde{l}$ with $1 \leq \tilde{l}_k \leq n + 1$ and $\sum_k \tilde{l}_k = n + 1$ is obtained. Finally, we take into account that the number of ways to achieve a certain vector $\tilde{l}$ increases with the growing n. This concludes the proof of the theorem.

Remark 10. For the analysis in this section, we have worked with $\mathcal{D}_v$ as the set of performance mappings. In applications we might also be interested in carrying out the analysis for subsets of $\mathcal{D}_v$, such as $\mathcal{C}_v$, as defined in (3).

Let $\hat{\mathcal{D}}$ be a subset of $\mathcal{D}_v$. In order to apply the analysis in this section to $\hat{\mathcal{D}}$, the following issues have to be addressed:

- P_θ, D_θ and Π_θ are $\hat{\mathcal{D}}$-preserving,

- $\| \cdot \|_v$-Lipschitz continuity implies $\hat{\mathcal{D}}$-Lipschitz continuity.

For example, if $\hat{\mathcal{D}} = \mathcal{C}_v$, then it has to be checked on an individual basis if P_θ is $\mathcal{C}_v$-preserving. However, for the remaining issues, sufficient conditions can be established. To see this, let Conditions 1–3 hold and assume that P_θ is $\| \cdot \|_v$-continuous. By Lemma 1, $\|\Pi_\theta\|_v$ is finite and, since $\Pi_\theta g$ is a constant mapping, Π_θ is $\mathcal{C}_v$-preserving. Moreover, if P_θ is $\mathcal{C}_v$-preserving, then P_θ^n is $\mathcal{C}_v$-preserving, and it follows that $P_\theta^m - \Pi_\theta$ is $\mathcal{C}_v$-preserving. Since $\sum_{m=0}^N (P_\theta^m - \Pi_\theta) g(s)$ converges uniformly to $D_\theta g(s)$ for any $s \in S$ as N tends to ∞, it readily follows that D_θ is $\mathcal{C}_v$-preserving.

As for $\mathcal{C}_v$-Lipschitz continuity, if v is continuous, then, following the line of argument put forward in Remark 4, it follows that $\| \cdot \|_v$-Lipschitz continuity indeed implies $\mathcal{C}_v$-Lipschitz continuity.

4. Taylor series expansions

When Π_θ is to be developed into a Taylor series, we have to control the growth rate of the elements of the series, that is, of $\{\Pi_\theta^{(n)}\}$. By Theorem 4, the key ingredient for obtaining an upper bound for the elements of the Taylor series is to bound the operator $K_\theta(m)$, $0 \le m \le n$, in the expression for $\Pi_\theta^{(n)}$. Elaborating on the $\| \cdot \|_v$ norm, such a bound can be obtained from

$$\|K_\theta(n)\|_v = \left\| \sum_{m=0}^\infty P_\theta^{(n)} (P_\theta^m - \Pi_\theta) \right\|_v$$
$$\le \|P_\theta^{(n)} D_\theta\|_v$$
$$\le \|P_\theta^{(n)}\|_v \|D_\theta\|_v.$$

Theorem 4 yields that

$$\|\Pi_\theta^{(n)}\|_v \le \sum_{\substack{1 \le m \le n \\ 1 \le l_k \le n \\ l_1 + \cdots + l_m = n}} \frac{n!}{l_1! \cdots l_m!} \|\Pi_\theta\|_v \prod_{k=1}^m \|P_\theta^{(l_k)} D_\theta\|_v$$
$$= \|\Pi_\theta\|_v H_\theta(n),$$

where

$$H_\theta(n) := \sum_{\substack{1 \le m \le n \\ 1 \le l_k \le n \\ l_1 + \cdots + l_m = n}} \frac{n!}{l_1! \cdots l_m!} \prod_{k=1}^m \|P_\theta^{(l_k)} D_\theta\|_v.$$

We summarize the above analysis in the following theorem.

Theorem 5. *Let Conditions 1–3 and C^{n+1} for some n be satisfied at θ. Then, for any $g \in D_v$,*

$$\int g \, d\Pi_{\theta+\Delta} = \sum_{m=0}^{n} \frac{\Delta^m}{m!} \int g \, d\Pi_{\theta}^{(m)} + r_{n+1,\Delta}$$

for any Δ with $|\Delta| < r$, where r is given in Condition C^{n+1} and

$$r_{n+1,\Delta} \leq \frac{|\Delta|^{n+1}}{(n+1)!} \|\Pi_{\theta}\|_v \sup_{|\Delta| \leq r} H_{\theta+\Delta}(n+1).$$

Suppose that the conditions of Theorem 5 hold. In particular, assume that Condition C^n holds for any n. Let r_{θ} be given through

$$\frac{1}{r_{\theta}} = \limsup_{n} \left(\frac{1}{n!} \|\Pi_{\theta}\|_v H_{\theta}(n) \right)^{1/n}$$

$$= \limsup_{n} \left(\frac{1}{n!} H_{\theta}(n) \right)^{1/n}. \tag{21}$$

Then, for any $g \in \mathcal{D}_v$, $\int g \, d\Pi_{\theta}$ can be developed into a Taylor series at θ whose radius of convergence is at least r_{θ} and we obtain the following result.

Corollary 2. *Assume that Condition 1–3 hold at θ. If Condition C^n is satisfied at θ for all $n \geq 1$, then*

$$\Pi_{\theta+\Delta} = \sum_{n=0}^{\infty} \frac{\Delta^n}{n!} \Pi_{\theta}^{(n)}$$

for $|\Delta| \leq r_{\theta}$, with r_{θ} as given in (21).

Example 4. We revisit the situation of Example 3 and we assume that Conditions 1–3 hold at θ. Under the conditions in Example 3, Theorem 4 applies and we obtain that

$$\Pi_{\theta}^{(n)} = n! \, \Pi_{\theta}(K_{\theta}(1))^n.$$

Because only the first-order derivative is significant, we have

$$H_{\theta}(n) = n! \left(\|(Q-P)D\|_v \right)^n,$$

where

$$D = \sum_{m=0}^{\infty} (P^m - \Pi_P)$$

is the deviation operator associated with P and Π_P the operator that maps any distribution to the unique invariant distribution to P. The radius of convergence of the Taylor series at θ is therefore given by

$$\frac{1}{r_{\theta}} = \limsup_{n} \left(\frac{1}{n!} \|\Pi_{\theta}\|_v H_{\theta}(n) \right)^{1/n}$$

$$= \|(Q-P)D\|_v.$$

Hence, by Corollary 2, the Taylor series for Π_θ reads

$$\Pi_{\theta+\Delta} = \sum_{n=0}^{\infty} \Delta^n \Pi_\theta (K_\theta(1))^n$$

$$= \sum_{n=0}^{\infty} \Delta^n \Pi_\theta ((Q-P)D_\theta)^n \qquad (22)$$

for any $\theta \in [0,1]$ and $|\Delta| < r_\theta$. The representation (22) was shown to hold for finite-state Markov chains in [8]. In Section 5 we will compute r_θ for an example from queueing theory.

In the previous example, we calculated the radius of convergence of the Taylor series for a Markov chain that had only one significant derivative. The general situation is as follows. Notice that

$$H_\theta(n) \leq \sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq n \\ l_1+\cdots+l_m=n}} \frac{n!}{l_1! \cdots l_m!} \prod_{k=1}^{m} \|P_\theta^{(l_k)}\|_v \|D_\theta\|_v.$$

If the chain satisfies Conditions 1–3 at θ, then $\|D_\theta\|$ is bounded by the expression given in Remark 6. Assume that the first $s(P_\theta)$ $\mathcal{D}_v$-derivatives of P_θ are significant and set $s(P_\theta) = \infty$ if all higher-order $\mathcal{D}_v$-derivatives are significant. Let

$$\eta_\theta = \sup_{n \leq s(P_\theta)} \|P_\theta^{(n)}\|_v.$$

Then

$$\|H_\theta(n)\|_v \leq \left(c_\theta \frac{\eta_\theta}{1-\rho_\theta} \right)^n \sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq \min(n,s(P_\theta)) \\ l_1+\cdots+l_m=n}} \frac{n!}{l_1! \cdots l_m!}.$$

Notice that

$$\sum_{\substack{1 \leq m \leq n \\ 1 \leq l_k \leq \min(n,s(P_\theta)) \\ l_1+\cdots+l_m=n}} \frac{n!}{l_1! \cdots l_m!} \leq \sum_{\substack{l \in \{0,\ldots,n\}^n \\ l_1+\cdots+l_n=n}} \frac{n!}{l_1! \cdots l_m!} \leq n^n.$$

Hence,

$$\|\Pi_\theta^{(n)}\|_v \leq \|\Pi_\theta\|_v \left(\frac{c_\theta \eta_\theta}{1-\rho_\theta} \right)^n n^n,$$

and an upper bound for the radius of convergence is obtained from the Cauchy–Hadamard theorem to be

$$\frac{1}{r_\theta} = \frac{c_\theta \eta_\theta}{1-\rho_\theta} \limsup_n \left(\frac{n^n}{n!} \right)^{1/n},$$

which gives

$$r_\theta = \frac{1-\rho_\theta}{c_\theta \eta_\theta e}.$$

Unfortunately, even for simple systems the values for c_θ, ρ_θ and η_θ cannot be computed. Numerical procedures for computing c_θ, ρ_θ and η_θ (approximately) are topics of further research.

Remark 11. Substituting the Taylor expansion of $P_{\hat\theta} - P_\theta$ into (13) and collecting the terms of $\hat\theta^k$ for given k provides an other derivation of the Taylor series of $\Pi_{\hat\theta}$.

5. Example

Let P denote the Markov kernel of the queue-length process in an M/M/1 queue with arrival rate η_1 and service rate μ, where we assume that $\eta_1 < \mu$. Furthermore, let Q denote the Markov kernel of the queue-length process in a G/M/1 queue with service rate μ and whose interarrival times consist of two stages: the first stage is exponentially distributed with rate η_1 and the second stage is independent of the first stage exponentially distributed with rate η_2. For $\theta \in [0, 1]$, set

$$P_\theta = (1 - \theta)P + \theta Q.$$

Then P_θ is the Markov kernel of the embedded queue-length process in a G/M/1 queue with service rate μ. Interarrival times follow a Cox distribution with rates η_j for $j = 1, 2$ and parameter θ, that is, the interarrival times consist with probability $1 - \theta$ of a single exponentially distributed stage with rate η_1, and a second stage with rate η_2 follows with probability θ. Let $X_\theta(n) = (X_\theta(1, n), X_\theta(2, n))$ be the state of the embedded jump process, with $X_\theta(1, n) \in \mathbb{N}$ the total number of customers in the system and $X_\theta(2, n) \in \{1, 2\}$ the stage of the interarrival time. Let

$$P_\theta((k, i); (k', i')) = \Pr(X_\theta(m + 1) = (k', i') \mid X_\theta(m) = (k, i))$$

for $(k, i), (k', i') \in \mathbb{N} \times \{1, 2\}$. Set

$$\mathcal{D}_v = \{g : \mathbb{N} \times \{1, 2\} \to \mathbb{R} \mid \text{for some } r|g(k, i)| \leq re^{\lambda k} \text{ for all } (k, i) \in \mathbb{N} \times \{1, 2\}\}$$

for λ sufficiently small.

It is shown in [16] that P_θ satisfies Conditions 1–3 at any $\theta \in [0, 1]$. Moreover, it is easily checked that $\|P\|_v$ and $\|Q\|_v$ are finite and, by Corollary 1, P and Q are therefore $\mathcal{D}_v$-preserving. In accordance with Example 3, Condition C^n holds for any n at any $\theta \in [0, 1]$. In the remainder of this section, we compute a lower bound for the radius of convergence of the Taylor series of Π_θ developed at $\theta = 0$.

The deviation matrix for the M/M/1 queue, denoted by D, was derived in [24], [25] and [10]. For the embedded process it is

$$D(i, j) := \frac{\rho^{\max\{j-i,0\}} - (i + j + 1)(1 - \rho)\rho^j}{\mu(1 - \rho)},$$

where

$$\rho := \frac{\eta_1}{\mu} \tag{23}$$

denotes the traffic intensity of the M/M/1 queue.

Let $D(i, 1; j) = D(i, j)$ for any $i, j \geq 0$. Note that, in principle, we also have to define $D(i, 2; j)$, but since these values are not required for our analysis we do not dwell on this issue here. In accordance with Example 4, the radius of convergence is determined through

$$\|(Q - P)D\|_v,$$

where, from Section 5.2.1 of [16], we may take $v(i, k) := \rho_1^i$ for $\rho_1 > 1$ and $\rho_1 - 1 < \delta$ for some $\delta > 0$. In particular,

$$(Dv)(i) = \sum_{j \geq 0} \frac{\rho^{\max\{j-i,0\}} - (i + j + 1)(1 - \rho)\rho^j}{\mu(1 - \rho)} \rho_1^j.$$

Note that P is not defined on states $(k, 2)$. We may define P on these states as Q, then $Q - P = 0$ on these states. Hence,

$$((Q - P)Dv)(i, 2) = 0.$$

For state $(i, 1)$, we find the following transition probabilities

$$P(i, 1; i + 1, 1) = Q(i, 1; i, 2) = \frac{\eta_1}{\eta_1 + \mu 1_{i>0}}$$

for $i \geq 0$; see Section 5.2.1 of [16] where a complete specification of P and Q is given. By calculation,

$$
\begin{aligned}
((Q - P)Dv)(i, 1) &= (Q(i, 1; i + 1, 1) - P(i, 1; i + 1, 1))(Dv)(i + 1) \\
&\quad + (Q(i, 1; i - 1, 1) - P(i, 1; i - 1, 1))(Dv)(i - 1) \\
&\quad + (Q(i, 1; i, 2) - P(i, 1; i, 2))(Dv)(i).
\end{aligned}
$$

In particular,

$$Q(i, 1; i + 1, 1) = 0 = P(i, 1; i, 2)$$

and

$$Q(i, 1; i - 1, 1) = P(i, 1; i - 1, 1),$$

which gives

$$
\begin{aligned}
((Q - P)Dv)(i, 1) &= Q(i, 1; i, 2)(Dv)(i) - P(i, 1; i + 1, 1)(Dv)(i + 1) \\
&= \frac{\eta_1}{\eta_1 + \mu 1_{i>0}}((Dv)(i) - (Dv)(i + 1)) \\
&= \frac{\eta_1}{\eta_1 + \mu 1_{i>0}} \left(\sum_{j \geq 0} \frac{\rho^{\max\{j-i,0\}} - (i + j + 1)(1 - \rho)\rho^j}{\mu(1 - \rho)} \rho_1^j \right. \\
&\qquad\qquad\qquad\left. - \sum_{j \geq 0} \frac{\rho^{\max\{j-i-1,0\}} - (i + j + 2)(1 - \rho)\rho^j}{\mu(1 - \rho)} \rho_1^j \right) \\
&= \frac{\eta_1}{\eta_1 + \mu 1_{i>0}} \frac{1}{\mu(1 - \rho)} \left(\sum_{j \geq 0} \rho^{\max\{j-i,0\}} \rho_1^j - \sum_{j \geq 0} \rho^{\max\{j-i-1,0\}} \rho_1^j \right) \\
&\quad + \frac{\eta_1}{\eta_1 + \mu 1_{i>0}} \frac{1}{\mu} \left(\sum_{j \geq 0} -(i + j + 1)\rho^j \rho_1^j - \sum_{j \geq 0} -(i + j + 2)\rho^j \rho_1^j \right).
\end{aligned}
$$

The summations in the above equality are

$$
\begin{aligned}
\sum_{j \geq 0} (\rho^{\max\{j-i,0\}} - \rho^{\max\{j-i-1,0\}})\rho_1^j &= \sum_{j \geq i+1} (\rho^{j-i} - \rho^{j-i-1})\rho_1^j \\
&= (\rho - 1)\rho_1^{i+1} \sum_{j \geq i+1} \rho^{j-i-1} \rho_1^{j-i-1} \\
&= -(1 - \rho)\rho_1^{i+1} \frac{1}{1 - \rho\rho_1}
\end{aligned}
$$

and

$$\sum_{j\geq 0} -(i+j+1)\rho^j \rho_1^j - \sum_{j\geq 0} -(i+j+2)\rho^j \rho_1^j = \sum_{j\geq 0}(-1)\rho^j \rho_1^j$$

$$= -\frac{1}{1-\rho\rho_1}.$$

Hence, we obtain for the overall expression

$$((Q-P)Dv)(i,1) = \frac{\eta_1}{\eta_1 + \mu 1_{i>0}} \frac{1}{\mu}\left(\rho_1^{i+1}\left(-\frac{1}{1-\rho\rho_1}\right) - \frac{1}{1-\rho\rho_1}\right).$$

Recall that the norm of a matrix A is defined as

$$\|A\|_v := \sup_i \sum_j |A(i,j)|\frac{v(j)}{v(i)}.$$

Using this and (23), we find that

$$\|(Q-P)D\|_v = \sup_i \frac{\eta_1}{\eta_1 + \mu 1_{i>0}} \frac{1}{\mu}\left(\frac{\rho_1}{1-\rho\rho_1} + \frac{1}{\rho_1^i(1-\rho\rho_1)}\right)$$

$$= \frac{1}{\mu}\left(\frac{1+\rho_1}{1-\rho\rho_1}\right)$$

$$= \frac{\rho}{\eta_1}\left(\frac{1+\rho_1}{1-\rho\rho_1}\right).$$

Since the radius of convergence r_0 is independent of ρ_1, we have that

$$r_0 \geq \limsup_{\rho_1 \downarrow 1}\left(\frac{\rho}{\eta_1}\left(\frac{1+\rho_1}{1-\rho\rho_1}\right)\right)^{-1} = \frac{\eta_1}{2}\left(\frac{1-\rho}{\rho}\right).$$

Note that if we take a different time scale, then η_1 will change but ρ is invariant. Hence, we may choose a time scale for which $r_0 > 1$, and consequently the Taylor series converges on the complete interval $[0,1]$. The Taylor series then gives the analytic expansion of the stationary distribution Π_θ on $[0,1]$. Recalling that Π_P denotes the unique invariant distribution on P, we have thus shown that

$$\Pi_\theta = \sum_{n=0}^{\infty} \theta^n \Pi_P ((Q-P)D)^n$$

$$= \sum_{n=0}^{\infty} \theta^n \Pi_P \left((Q-P)\sum_{m=0}^{\infty}(P^m - \Pi_P)\right)^n,$$

for any $\theta \in [0,1]$.

Acknowledgement

References

[1] AYHAN, H. AND BACCELLI, F. (2001). Expansions for joint Laplace transforms of stationary waiting times in (max,+)-linear systems with Poisson input. *Queueing Systems* **37,** 291–328.

[2] AYHAN, H. AND SEO, D. (2001). Laplace transform and moments of waiting times in Poisson driven (max,+)-linear systems. *Queueing Systems* **37,** 405–436.

[3] AYHAN, H. AND SEO, D. (2002). Tail probability of transient and stationary waiting times in (max,+)-linear systems. *IEEE Trans. Automatic Control* **47,** 151–157.

[4] BACCELLI, F. AND SCHMIDT, V. (1996). Taylor series expansions for Poisson-driven (max,+)-linear systems. *Ann. Appl. Prob.* **6,** 138–185.

[5] BACCELLI, F., HASENFUSS, S. AND SCHMIDT, V. (1997). Transient and stationary waiting times in (max,+)-linear systems with Poisson input. *Queueing Systems* **26,** 301–342.

[6] BACCELLI, F., HASENFUSS, S. AND SCHMIDT, V. (1998). Expansions for steady-state characteristics of (max,+)-linear systems. *Commun. Statist. Stoch. Models* **14,** 1–24.

[7] BOROVKOV, A. AND HORDIJK, A. (2004). Characterization and sufficient conditions for normed ergodicity of Markov chains. To appear in *Adv. Appl. Prob.* **36,** No. 1.

[8] CAO, X.-R. (1998). The Maclaurin series for performance functions of Markov chains. *Adv. Appl. Prob.* **30,** 676–692.

[9] CASSANDRAS, C. AND LAFORTUNE, S. (1999). *Introduction to Discrete Event Systems.* Kluwer, Norwell, MA.

[10] COOLEN-SCHRIJNER, P. AND VAN DOORN, E. A. (2002). The deviation matrix of a continuous-time Markov chain. *Prob. Eng. Inf. Sci.* **16,** 351–366.

[11] DEKKER, R. AND HORDIJK, A. (1988). Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Math. Operat. Res.* **13,** 395–421.

[12] DEKKER, R., HORDIJK, A. AND SPIEKSMA, F. M. (1994). On the relation between recurrence and ergodicity properties in denumerable Markov decision chains. *Math. Operat. Res.* **19,** 539–559.

[13] GONG, W.-B. AND HU, J.-Q. (1992). The Maclaurin series of the GI/G/1 queue. *J. Appl. Prob.* **29,** 176–184.

[14] HEIDERGOTT, B. AND VÁZQUEZ-ABAD, F. (2000). Measure-valued differentiation for stochastic processes: the finite horizon case. EURANDOM Report 2000-033. Available at http://staff.feweb.vu.nl/bheidergott/.

[15] HEIDERGOTT, B., HORDIJK, A. AND WEISSHAUPT, H. (2002). Derivatives of Markov kernels and their Jordan decomposition. EURANDOM Report 2003-001. Available at http://staff.feweb.vu.nl/bheidergott/.

[16] HEIDERGOTT, B., HORDIJK, A. AND WEISSHAUPT, H. (2002). Measure-valued differentiation for stationary Markov chains. EURANDOM Report 2002-027. Available at http://staff.feweb.vu.nl/bheidergott/.

[17] HO, Y. AND CAO, X. (1991). *Perturbation Analysis of Discrete Event Systems.* Kluwer, Boston, MA.

[18] HORDIJK, A. AND DEKKER, R. (1983). Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. Rep. 83–36, Institute of Applied Mathematics and Computing Science, Leiden University.

[19] HORDIJK, A. AND PUTERMAN, M. L. (1987). On the convergence of policy iteration in finite state undiscounted Markov decision processes: the unichain case. *Math. Operat. Res.* **12,** 163–176.

[20] HORDIJK, A. AND SPIEKSMA, F. M. (1992). On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Adv. Appl. Prob.* **24,** 343–376.

[21] HU, J.-Q. (1995). Analyticity of single-server queues in light traffic. *Queueing Systems* **19,** 63–80.

[22] HU, J.-Q. (1996). The departure process of the GI/G/1 queue and its Maclaurin series. *Operat. Res.* **44,** 810–815.

[23] KARTOSCHOV, N. (1985). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. *Theory Prob. Appl.* **30,** 247–259.

[24] KOOLE, G. M. (1998). The deviation matrix of the M/M/1/∞ and M/M/1/N queue, with applications to controlled queueing models. In *Proc. 37th IEEE Conf. Decision Control* (Tampa, FL), IEEE Press, pp. 56–59.

[25] KOOLE, G. M. AND SPIEKSMA, F. M. (2001). On deviation matrices for birth–death process. *Prob. Eng. Inf. Sci.* **15,** 239–258.

[26] MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

[27] PFLUG, G. (1996). *Optimisation of Stochastic Models.* Kluwer, Boston, MA.

[28] VAN DEN HOUT, W. (1996). *The power-series-algorithm.* Doctoral Thesis, Center for Economic Research, Tilburg University.

[29] ZHU, Y. AND LI, H. (1993). The Maclaurin expansion for a G/GI/1 queue with Markov-modulated arrivals and services. *Queueing Systems* **14,** 125–134.

[30] ZAZANIS, M. (1992). Analyticity of Poisson-driven stochastic systems. *Adv. Appl. Prob.* **24,** 532–541.