Mathematics of Control, Signals, and Systems
https://doi.org/10.1007/s00498-026-00436-0

ORIGINAL ARTICLE

Stochastic approximation in a Markovian framework revisited: Lipschitz continuity of the Poisson equation

Algo Carè[2] · Balázs Csanád Csáji[1,4] · Balázs Gerencsér[3,4] · László Gerencsér[1] · Miklós Rásonyi[3,4]

Received: 30 September 2025 / Accepted: 27 January 2026

Abstract

In this paper, we revisit a fundamental technical issue within the theory of stochastic approximation (SA) in a Markovian framework, first proposed in the book by Derevitskii and Fradkov (Applied theory of discrete adaptive control systems, Nauka, 1981), and further developed in much detail in the book by Benveniste, Métivier, and Priouret (Adaptive algorithms and stochastic approximations, Springer, Berlin, 1990). This theory is instrumental in many application areas such as the statistical analysis of Hidden Markov Models arising in telecommunication, quantized linear stochastic systems, and reinforcement learning. The problem at hand is the verification of the existence, uniqueness, and Lipschitz continuity of the solution of a parameter-dependent Poisson equation, in an appropriate weighted sup-norm, associated with a collection of Markov chains on general state spaces. Verification of the above facts is vital in the analysis of SA processes presented in the cited book via the ODE (ordinary differential equations) method, requiring substantial technical effort. The motivation and focus of the paper

✉ Balázs Csanád Csáji
csaji.balazs@sztaki.hun-ren.hu

Algo Carè
algo.care@unibs.it

Balázs Gerencsér
gerencser.balazs@renyi.hun-ren.hu

László Gerencsér
gerencser.laszlo@sztaki.hun-ren.hu

Miklós Rásonyi
rasonyi.miklos@renyi.hun-ren.hu

1 HUN-REN Institute for Computer Science and Control, Hungarian Research Network, Kende u. 13-17., Budapest 1111, Hungary

2 Dipartimento di Ingegneria dell'Informazione, Università di Brescia, via Branze 38, Brescia 25123, Italy

3 HUN-REN Alfréd Rényi Institute of Mathematics, Hungarian Research Network, Reáltanoda u. 13-15., Budapest 1053, Hungary

4 Department of Probability Theory and Statistics, Institute of Mathematics, Eötvös Loránd University (ELTE), Pázmány Péter sétány 1/C, Budapest 1117, Hungary

Published online: 27 March 2026

is to address this technical issue, by presenting a simple set of conditions, under which the above properties of the Poisson equation at hand can be conveniently established. A distinctive feature of our work is that it is based on a remarkable result of Hairer and Mattingly (2011), proving that by tilting standard conditions of mainstream stability theory for Markov chains, the transition kernels prove to be contractions in the space of differences of probability measures in a suitable metric. To demonstrate the applicability of our results, the proposed conditions are verified for a class of queuing system with open-loop control.

Keywords Stochastic approximation · Markov chains on general state spaces · Weighted total variation norm · Contractive Markov kernels · Poisson equation · Lipschitz continuity

1 Introduction

Stochastic approximation (SA) is a fundamental methodology for real-time statistical analysis in important application areas such as signal processing, control systems, and more recently in machine learning. Specifically, many algorithms in adaptive filtering, recursive system identification, adaptive methods in input design, adaptive control, or reinforcement learning rely on ideas of classical stochastic approximation theory initiated by the celebrated paper of Robbins and Monro [1] back in 1951.

Stochastic approximation in a Markovian framework, first proposed in [2, 3], and extensively developed in the book [4], was a significant contribution to the area, allowing the construction and analysis of statistical estimation methods for a wide class of nonlinear systems, such as Hidden Markov Models (HMM-s), arising in telecommunication, or quantized linear stochastic systems. In machine learning, studying SA in a Markovian framework is instrumental for reinforcement learning (RL), such as TD-learning [5] and Q-learning [6]. The theory presented in [4] is in a sense complementary to the widely used theory for recursive identification of linear stochastic systems, developed by Ljung in [7], with focus on mixing properties of the driving noise.

The main *contribution* of this paper is a significant addition to the theory of SA in a Markovian framework, developed in [4], by introducing much simpler conditions for the Markov chain under which the Lipschitz Continuity of the Poisson Equation, a key technical issue required for the ODE analysis, can be resolved. With all technicalities explained in sufficient details, the paper is self-contained.[1]

To provide the context of the present paper, we briefly describe a few technical aspects central in [4]. Following their predecessors, the authors of [4] present a model that boils down to the solution of a nonlinear algebraic equation, specifically defined in terms of a strictly stationary parameter-dependent Markov process $(X_n(\theta))$, representing physical signals, their filtered values, or a combination of these. The process may take its values in an abstract measurable space $\mathbf{X}$, such as a Euclidean space $\mathbb{R}^m$,

[1] This paper is a significantly extended version of our paper published in the Proceedings of the 58th IEEE Conference on Decision and Control (CDC) [8]; a preprint of the current extension is available at [9].

a Hilbert space, or a finite discrete set. For the sake of simplicity, we assume that $(X_n(\theta))$ has a unique invariant distribution or measure $\mu^*(\theta)$ over $\mathbf{X}$.

The parameter $\theta \in \Theta$ may characterize the open-loop system dynamics, the effect of a controller, or the tentative value of the true parameter within system identification. In the context of [4], as in the theory of recursive system identification and adaptive control of linear stochastic systems, see [7], Θ is typically a subset of a Euclidean space, $\mathbb{R}^k$, while we assume that Θ is an open set of a (separable) Banach or Hilbert space, or more generally of a normed space, as required in some recent machine learning applications, see [10–12]. We will further discuss the relevance of our more general setting for applications in the Introduction below.

The dynamics of the Markov process $(X_n(\theta))$ would be classically described by its transition probability kernels $P_\theta(x, A)$, with $x \in \mathbf{X}$, $A \subset \mathbf{X}$, A being measurable, and indeed, we shall follow suit in the rest of the paper. But for now it is preferable to take a system's point of view and define the process explicitly via

$$X_{n+1}(\theta) = F(\theta; X_n(\theta), W_{n+1}), \qquad n \geq 0, \tag{1}$$

where F is a measurable mapping, and (W_n) is a sequence of i.i.d. (independent, identically distributed) random variables. The sequence (W_n) may represent exogenous system noise, measurement noise, or a dither injected by the user. The initial condition $X_0(\theta)$ is assumed to have stationary distribution $\mu^*(\theta)$. The dependence of $X_0(\theta)$ on parameter θ should not be discussed at this point.

The objective is to identify or tune the parameter so that some appropriately defined asymptotic cost function expressing reconstruction error or tracking error is minimized. The (problem-specific) instantaneous cost is a function of both θ and $x \in \mathbf{X}$. Assuming $\theta \in \mathbb{R}^k$, an $\mathbb{R}^k$-valued pseudo-gradient of the (instantaneous) cost function w.r.t. (with respect to) θ can be defined, denoted by $H(\theta; x)$. Define, for any n,

$$h(\theta) := \mathbb{E}\big[H(\theta; X_n(\theta))\big].$$

Then, the (asymptotic) estimation problem reduces to the problem of solving the nonlinear algebraic equation

$$h(\theta) = 0. \tag{2}$$

The practical objective is to find the root of (2), denoted by θ^*, via a recursive algorithm based on computable approximations of $H(\theta; X_n(\theta))$. Computability means that the r.h.s. (right-hand side) of (3)–(4) can be evaluated via the cyber-physical system at hand. The proposed SA algorithm of [4] is

$$X_{n+1} = F(\theta_n; X_n, W_{n+1}), \tag{3}$$

$$\theta_{n+1} = \theta_n + \frac{1}{n+1} H(\theta_n; X_{n+1}), \tag{4}$$

with $n \geq 0$, $\theta_0 \in \Theta$ and X_0 arbitrary subject to moment constraints. We note in passing that computability is critical in applications such as stochastic adaptive control or adaptive input design, see [13].

We note that an early variant of the above problem and the associated algorithm is presented in [14], further elaborated in much detail in [7], where $(X_n(\theta))$ is assumed to be defined via a linear stochastic system, driven by a weakly dependent process, exhibiting certain mixing properties and a weak form of stationarity. Therefore, in [7], typically, $(X_n(\theta))$ is not a Markov process.

A long-standing challenge is to prevent the estimates θ_n from leaving the domain of definition Θ. A common remedy is that the parameter estimate is forced to stay in a compact domain using resetting, see [4] or [15] in the context of [7]. An alternative remedy is to restrict attention to global SA processes with $\Theta = \mathbb{R}^k$, and assuming the existence of a global stochastic Lyapunov function preventing θ_n from diverging [16].

Apart from this, a common standard approach for the convergence analysis of the above algorithm is the ODE (ordinary differential equation) method, see [4, 7, 17], in which the discrete-time estimate sequence (θ_n) is approximated by the continuous-time solution trajectory z_s of the following ODE,

$$\dot{z}_s = h(z_s), \qquad s \geq 0,$$

matching time indices via $n = [e^s]$, see [4]. Alternatively, we can define the ODE by

$$\dot{y}_t = \frac{1}{t} h(y_t), \qquad t \geq 1. \tag{5}$$

The advantage of this approach is that the change of time scale is inherent in the ODE, and $|\theta_n - y_n|$ can be conveniently bounded from above for any integer $n \geq 1$, see [15] in the context of [7]. A key condition for θ_n tracking y_n is that the ODE has an asymptotically stable equilibrium point at θ^* with a reasonable domain of attraction.

In order to capture the tracking error $|\theta_n - y_n|$ on finite intervals, following a sequence of intricate arguments, see [4, Part II, Chapters 1 and 2], we arrive at the problem of estimating the additive functional

$$\sum_{n=1}^{N} \Big(H(\theta; X_n(\theta)) - \mathbb{E}_{\mu_\theta^*}\big[H(\theta; X_n(\theta))\big] \Big), \tag{6}$$

where μ_θ^* is the assumed unique invariant measure of $X_n(\theta)$ under parameter θ. A well-known device in the theory of Markov processes is to express (6) using a Markovian version of the Newton–Leibniz formula of basic calculus by representing the individual terms via the solution of the Poisson equation

$$(I - P_\theta^*)\, u_\theta(x) \;=\; H(\theta; x) - \mathbb{E}_{\mu_\theta^*}\big[H(\theta; X(\theta))\big]. \tag{7}$$

Here, $u_\theta(x)$ is an unknown function, playing the role of a primitive function, and P_θ^*, the adjoint of P_θ, is given by (8). Thus, (6) will become the sum of martingale differences. For a historical perspective on the topic, see the early paper [18].

In the ODE analysis of [4, Part II, Chapter 2], a vital technical tool is the verification of the Lipschitz continuity of $u_\theta(x)$ w.r.t. θ. This requires substantial technical efforts, and the conditions under which a key technical result [4, Part II, Chapter 2, Theorem 6]) is derived are quite demanding, see the final note below at the end of Sect. 4.

The motivation and focus of the present paper is to provide a significantly simpler set of conditions, with Assumption 3 playing a central role, under which the existence, uniqueness, and Lipschitz continuity of the solution of parameter-dependent Poisson equations can be conveniently established.

Our methodology is based on a powerful result in the stability theory for Markov chains, developed in [19]. A major observation of [19] is that by tilting standard conditions of mainstream stability theory for Markov chains, see [20], the transition kernels prove to be contractions in the space of differences of probability measures in a suitable metric, see Proposition 3. We briefly present and interpret the main results of [19] in a self-contained manner in Sect. 2. Applying this mathematical framework to the problem at hand yields a transparent and flexible analysis.

The prospective advantage of using the methodology of [19] by Hairer and Mattingly is that (uniform) contractivity of the probability transition kernels for $\theta \in \Theta$ implies the stability of the inhomogeneous Markov chain given by (3) with arbitrary $\theta_n \in \Theta$. Ensuring some kind of stability of (3) is in fact a key issue in SA, see [4].

The recent interest in stochastic approximation in a Markovian framework is also reflected in [12] with a focus on the ODE method. However, in contrast to the general theory of [4] the Markov chain, driving the SA iteration, does not depend on the parameter. Therefore, it bypasses the particular technical issue settled in this paper.

The same limited scope has been chosen in the early versions of [21]. In this latest version, representing a major extension of prior versions, recently published as [16], the authors do consider SA processes driven by Markov chains depending on a parameter $\theta \in \mathbf{R}^d$, partially following the logic of [4]. In particular, in analogy with [8], they prove Lipschitz continuity of the solution of parameter-dependent Poisson equations within the context of mainstream stability theory for Markov chains [20], see their Proposition 7. However, geometric ergodicity and its variants do not imply contractivity of the probability transition kernels in the space of differences of probability measures in any of the metrics used in [20]. Hence, proving stability of the inhomogeneous Markov chain given by (3) with arbitrary $\theta_n \in \mathbf{R}^d$, without further enforced restrictions on the dynamics of θ_n, as in [4], requires substantial technical efforts.

To give a brief historical assessment, it is somewhat surprising that SA in a Markovian framework, in spite of its wide range of applicability and a fairly comprehensive theory presented in [4], has attracted relatively limited attention until the surge of interest aroused by machine learning (ML). The reason for this is twofold. On the one hand, the applicability of [4] is limited by the restrictions they impose on the state space and the parameter space. In particular, the state space is assumed to be a Euclidean space, a finite set or their Cartesian product. This assumption may not hold even for SA algorithms designed for very simple problems, such as estimating a parameter from quantized linear regression with Gaussian noise, see [22, 23]. SA methods with more general state spaces are also important for ML, as more advanced ML methods have state spaces containing graphs [24], signals constituting infinite-dimensional function spaces [25], or the state space could even be a differentiable manifold [26]. Finally, we may need function spaces as state space for cyber-physical systems incorporating a physical system the dynamics of which is governed by a stochastic partial differential equation (SPDE). A prototype for this, exhibiting remarkable stability properties, is the Markov process defined by an Allen–Cahn equation, central to statistical physics, see [27]. The above limitation of [4] is bypassed in the present paper by considering Markov chains on general measurable spaces.

The applicability of [4] is further limited by the assumption that the parameter space is a subset of a Euclidean space $\mathbb{R}^d$. In contrast, SA methods serving as the basis for most RL methods [28] typically fail to fit into this framework. Namely, RL methods, such as Q-learning or TD-learning, recursively estimate the optimal value function or the value function of a control policy, which span an infinite-dimensional space for more general (infinite) state and action spaces [29]. Kernel methods for ML constitute another example, which are fundamentally nonparametric, where the model space is defined as a Reproducing Kernel Hilbert Space (RKHS) [30]. Recursive estimation of these functions needs to consider SA methods working with more general parameter spaces [11], such as separable Banach spaces covered by our paper.

A more serious impediment in the applicability of [4] is the complexity of some of their assumptions under which key technical results such as Theorem 6, Chapter 2, Part II. on p. 262, the Lipschitz continuity of the solution of a parameter-dependent Poisson equation can be established. We shall provide a detailed critical assessment of these assumptions subsequently at the end of Sect. 4. For a start, we point out only that the verification of these assumptions, if they hold at all, requires considerable additional case-by-case analysis. In contrast, our alternative Assumption 3 together with Assumptions 1 and 2 or their relaxed versions, Assumptions 5 and 6, are much more convenient to work with. We demonstrate the power of our approach on a problem arising in the design of queuing systems, unfit for the analysis of [4], in Chapter 8.

The structure of the paper is as follows: In Sect. 2, we provide a brief introduction to the stability theory for Markov chains developed in [19], with appropriate interpretations and eventual simplifications. In Sect. 3, fundamental properties of the Poisson equation, existence and uniqueness of solutions, are discussed. In Sect. 4, a simple condition imposing the Lipschitz continuity of the kernel is introduced, and its implications are discussed. The first main result of the paper is stated in Sect. 5, as Theorem 2, stating the Lipschitz continuity of the solutions of a parameter-dependent Poisson equation under reasonable conditions.

In Sect. 6, the cited results of [19] are restated under relaxed conditions, in particular, imposing conditions on some power of the kernel, P_θ^r, reflecting the demarcation between contractivity and stability of ordinary matrices. In Sect. 7, we restate the results of Sect. 5 under relaxed conditions, leading to our second main result, Theorem 4. In Sect. 8, the viability of our results is demonstrated on the modification of a classical textbook example, the design of a simple queuing system with open-loop control, a preliminary version of which has been presented in [8]. The paper is concluded with a brief discussion on potential future research directions.

Given the technical nature of the paper, pre-determined by the subject matter, we have chosen a semiclassical structure to enhance readability: Relevant concepts, novel theorems, and their structured proofs are presented in the main body of the paper, whereas the details of the proofs of lemmas and corollaries are given in the Appendix.

2 On a Theorem of Hairer and Mattingly

In this section, we provide a brief summary of an intriguing addition to mainstream stability theory for Markov Chains, developed in [19]. Consider a family of Markov chains $(X_n(\theta)),\ 0 \le n < +\infty$ with arbitrary state space $\mathbf{X}$ equipped with a σ-field $\mathcal{A}$ of events, and Θ being an open set of a normed space. Therefore, we consider the θ-dependent transition probability kernels $P_\theta(x, A)$, with $x \in \mathbf{X},\ A \in \mathcal{A}$, a shorthand for the conditional probability $P(X_1(\theta) \in A \mid X_0(\theta) = x)$. We will assume that for each $A \in \mathcal{A}$, $P_\cdot(\cdot, A)$ is (x, θ)-measurable. These assumptions can always be satisfied if $\mathbf{X}$ is a Polish space and Θ is a subset of a separable Banach space.
For any probability measure μ over $(\mathbf{X}, \mathcal{A})$ and measurable $\varphi : \mathbf{X} \to \mathbb{R}$, define

$$\begin{aligned}(P_\theta\mu)(A) &:= \int_{\mathbf{X}} P_\theta(x, A)\mu(\mathrm{d}x),\\ (P_\theta^*\varphi)(x) &:= \int_{\mathbf{X}} \varphi(y) P_\theta(x, \mathrm{d}y) = \mathbb{E}_\theta\big[\,\varphi(X_1) \mid X_0 = x\,\big],\end{aligned} \tag{8}$$

assuming the second integral exists. Hence, $(P_\theta\mu)(.)$ is the probability measure of $X_1(\theta)$ assuming that the probability measure of $X_0(\theta)$ is μ, and $(P_\theta^*\varphi)(x)$ is the conditional expectation $\mathbb{E}_\theta\big[\,\varphi(X_1) \mid X_0 = x\,\big]$. The next condition is motivated by [19], stated there for a single Markov chain.

Assumption 1 (Uniform Drift Condition for P_θ) There exists a measurable function $V : \mathbf{X} \to [0, \infty)$ and constants $\gamma \in (0, 1)$ and $K \ge 0$ such that for all $x \in \mathbf{X}$ and $\theta \in \Theta$

$$(P_\theta^* V)(x) \ \le\ \gamma V(x) + K. \tag{9}$$

$V(x)$ is called a Lyapunov function. Note that $V(x)$, γ and K are not θ-dependent. Let us take a measure μ such that

$$\mu(V) := \int_{\mathbf{X}} V(x)\mu(\mathrm{d}x) < \infty \quad \text{and} \quad \mu(\mathbf{X}) < \infty. \tag{10}$$

Then, integrating (9) with respect to μ we get for all $\theta \in \Theta$:

$$P_\theta\mu(V) \ \le\ \gamma\mu(V) + K\mu(\mathbf{X}). \tag{11}$$

Condition (10) is often expressed as $\mu(1 + \beta V) < \infty$ with some (and therefore any) $\beta > 0$. Inequality (11) extends for any signed measure η, with $|\eta|(1 + \beta V) < \infty$:

$$|P_\theta\eta|(V) \ \le\ \gamma|\eta|(V) + K|\eta|(\mathbf{X}), \tag{12}$$

for all $\theta \in \Theta$, due to the inequality $|P_\theta\eta| \le P_\theta|\eta|$. The set of signed measures η with $|\eta|(1 + \beta V) < \infty$ is denoted by $\mathcal{M}_V$.

Note that the Lyapunov function $V(\cdot)$ can be fairly general, as opposed to [4]. In particular, we may use $V(x) = e^{cx}$, known to be the right choice for queuing systems

[20, Section 16.4]. The next condition is a natural extension of Assumption 2 of [19] for a parametric family of Markov chains, which itself is a modification of a standard condition in the stability theory of Markov chains [20] requiring minorization on what is called a small set.

Assumption 2 (Local Minorization) Let $R > 2K/(1-\gamma)$, where γ and K are the constants from Assumption 1, and define the set $\mathcal{C} = \{x \in \mathbf{X} : V(x) \leq R\}$. There exist a probability measure $\bar{\mu}$ on $\mathbf{X}$ and a constant $\bar{\alpha} \in (0, 1)$ such that, for all $\theta \in \Theta$, all $x \in \mathcal{C}$, and all measurable A,

$$P_\theta(x, A) \geq \bar{\alpha}\bar{\mu}(A). \tag{13}$$

Remark 1 (Interpretation of R) If there exists an invariant measure μ_θ^* such that $\int_{\mathbf{X}} V(x)\mu_\theta^*(\mathrm{d}x) < \infty$, then integrating both sides of inequality (9), we infer that

$$\int_{\mathbf{X}} V(x)\mu_\theta^*(\mathrm{d}x) \leq \frac{K}{1-\gamma}. \tag{14}$$

Thus, the parameter R in Assumption 2 must exceed twice the mean of V w.r.t. any of the invariant measures.

Remark 2 (Constant shifts of V) We can and will assume that $\inf_x V(x) = 0$ without loss of generality. In fact, if a function V satisfies Assumptions 1 and 2, then $V' := V - \inf_x V(x)$ also satisfies Assumptions 1 and 2 with the same contraction coefficient γ, the same minorization domain $\mathcal{C}$, and the same $\overline{\alpha}$, with appropriately chosen parameters K' and R'.

A key technique used in [19] for the stability analysis of Markov processes is the so-called weighted total variation distance between two probability measures:

Definition 1 Let μ_1 and μ_2 be two probability measures on $\mathbf{X}$ and $\beta > 0$. Define the weighted total variation distance as

$$\rho_\beta(\mu_1, \mu_2) := \int_{\mathbf{X}} (1 + \beta V(x))|\mu_1 - \mu_2|(\mathrm{d}x), \tag{15}$$

where $|\mu_1 - \mu_2|$ is the total variation measure of $(\mu_1 - \mu_2)$.

For $\beta = 0$, the weighted total variation distance would become just the standard total variation distance $|\mu_1 - \mu_2|_{\mathrm{TV}}$.

Alternatively, writing $\eta = \mu_1 - \mu_2$, we can define

$$\rho_\beta(\eta) := \int_{\mathbf{X}} (1 + \beta V(x))|\eta|(\mathrm{d}x). \tag{16}$$

The definition extends to any signed measure η with $\int_{\mathbf{X}}(1 + \beta V(x))|\eta|(\mathrm{d}x) < \infty$, leading to what we call the weighted total variation norm of η.

An equivalent definition of ρ_β can be given by allowing more general weighting functions $\varphi\colon \mathbf{X} \to \mathbb{R}$ replacing $1 + \beta V(x)$. Specifically, let us introduce the norm:

Definition 2 For any measurable function $\varphi \colon \mathbf{X} \to \mathbb{R}$, set

$$\|\varphi\|_\beta := \sup_x \frac{|\varphi(x)|}{1 + \beta V(x)}. \tag{17}$$

The linear space spanned by the functions such that $\|\varphi\|_\beta < \infty$ will be denoted by $\mathcal{L}_V$. Note that $\mathcal{L}_V$ is neither affected by constant shift of V, nor the choice of β; moreover, $\mathcal{L}_V$ with the norm $\|\cdot\|_\beta$ becomes a Banach space for any $\beta > 0$.

Given the norm $\|\cdot\|_\beta$, an equivalent definition of the weighted total variation norm can be obtained as follows:

$$\rho_\beta(\eta) := \sup_{\varphi: \|\varphi\|_\beta \leq 1} \int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x). \tag{18}$$

Weighted total variation norms have been also used in the classic book [20], introduced in Chapter 14. Conditions for geometric convergence of $P^n(x, .)$ to the unique invariant measure, interpreted via weighted total variation norms, are given in Theorem 16.0.1, see Remark 4 for details. However, the smart choice of the weighting factor β, ensuring the contractivity of P_θ^*, showing up in Proposition 2, and the simplification of the conditions required are the main novelty of [19].

To capture the smoothing effect of P_θ^* acting on $\mathcal{L}_V$, define a measure of oscillation for functions $\varphi \in \mathcal{L}_V$ as follows:

Definition 3 For any function $\varphi \in \mathcal{L}_V$, set

$$|||\varphi|||_\beta = \min_{c \in \mathbb{R}} \|\varphi + c\|_\beta. \tag{19}$$

It is readily seen that $\|\varphi + c\|_\beta$ is continuous in c, and unbounded when c tends to $\pm\infty$, and hence, the right-hand side of (19) is well defined. Obviously, $|||\varphi|||_\beta \leq \|\varphi\|_\beta$.

It is easily seen that $|||\cdot|||_\beta$ is a semi-norm on $\mathcal{L}_V$ and $|||\varphi|||_\beta = 0$ if and only if φ is a constant function. Letting $\mathbb{R}_X$ denote the linear vector space of constant functions on $\mathbf{X}$, it follows that $|||\cdot|||_\beta$ is a norm on the linear factor space $\mathcal{L}_{V,0} := \mathcal{L}_V/\mathbb{R}_X$. It is also easily seen that $\mathcal{L}_{V,0}$ becomes a Banach space with the norm $|||\cdot|||_\beta$. In what follows, $\mathcal{L}_{V,0}$ will denote the latter Banach space with a specific, fixed choice of β to be described in Proposition 2. We note that the above definition of $|||\varphi|||_\beta$ is a simplification of what is given in [19].

A useful linear subspace of the dual space $\mathcal{L}_{V,0}^*$ is obtained by considering the linear space of signed measures η such that

$$\int_{\mathbf{X}} (1 + \beta V(x))|\eta|(\mathrm{d}x) < \infty, \quad \text{and} \quad \eta(\mathbf{X}) = 0, \tag{20}$$

which will be denoted by $\mathcal{M}_V^0$. It is easily seen that

$$\varphi \mapsto \int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x) \tag{21}$$

is a continuous linear functional the dual norm of which is

$$\sigma_\beta(\eta) := \sup_{\varphi:|||\varphi|||_\beta \le 1} \int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x). \tag{22}$$

Proposition 1 *(cf. [19]) For any $\eta \in \mathcal{M}_V^0$ and any $\beta > 0$*

$$\sigma_\beta(\eta) = \rho_\beta(\eta). \tag{23}$$

The dual approach in defining the same norm proved to be and will prove to be extremely useful.

For the sake of clarity, we briefly recapitulate the argument leading to (23). Obviously, the set $\{\varphi : \|\varphi\|_\beta \le 1\}$ is a subset of $\{\varphi : |||\varphi|||_\beta \le 1\}$, hence for any signed measure $\eta \in \mathcal{M}_V^0$:

$$\sup_{\varphi:|||\varphi|||_\beta \le 1} \int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x) \ge \sup_{\varphi:\|\varphi\|_\beta \le 1} \int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x). \tag{24}$$

On the other hand, for any fixed φ such that $|||\varphi|||_\beta \le 1$ there exists a c such that $\|\varphi + c\|_\beta \le 1$. But

$$\int_{\mathbf{X}} \varphi(x)\eta(\mathrm{d}x) = \int_{\mathbf{X}} (\varphi(x) + c)\eta(\mathrm{d}x). \tag{25}$$

Hence, a strict inequality in (24) cannot occur. □

The norm $\sigma_\beta(.)$ defined for signed measures $\eta \in \mathcal{M}_V^0$ naturally leads to the following definition of a metric:

Definition 4 Let μ_1, μ_2 be two signed measures on $\mathbf{X}$ with $\int_{\mathbf{X}}(1+\beta V(x))|\mu_i|(\mathrm{d}x) < \infty$, for $i = 1, 2$, and moreover, $\mu_1(\mathbf{X}) = \mu_2(\mathbf{X})$. Then, we define the distance

$$\sigma_\beta(\mu_1, \mu_2) := \sup_{\varphi:|||\varphi|||_\beta \le 1} \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x). \tag{26}$$

It is readily seen that $\sigma_\beta(\mu_1, \mu_2)$ is a metric in the space of probability measures. A simple corollary of Proposition 1 is

Corollary 1 *Let μ_1, μ_2 be two signed measures on $\mathbf{X}$ as in Definition 4. Then,*

$$\sigma_\beta(\mu_1, \mu_2) = \rho_\beta(\mu_1, \mu_2). \tag{27}$$

Now, we summarize the main results of [19]. It is well known that the kernels P_θ acting on probability measures are non-expansive in the total variation distance:

$$|P_\theta\mu_1 - P_\theta\mu_2|_{\mathrm{TV}} \le |\mu_1 - \mu_2|_{\mathrm{TV}}. \tag{28}$$

A major contribution of [19] is the result stating that the kernels P_θ are actually contractions in weighted total variation distance by choosing $\beta < \overline{\alpha}/K$, under simple conditions, see Proposition 3. The path to proving this result is to establish first that the operator P_θ^* acting on the Banach space $\mathcal{L}_{V,0}$ is a contraction, as stated in [19, Theorem 3.1]:

Proposition 2 *Under Assumptions 1 and 2, there exist $\beta > 0$ and $\alpha \in (0, 1)$ such that for all θ and $\varphi \in \mathcal{L}_V$*

$$\left|\left|\left| P_\theta^* \varphi \right|\right|\right|_\beta \le \alpha |||\varphi|||_\beta. \tag{29}$$

The pair (β, α) can be chosen as follows:

$$\beta = \alpha_0/K \qquad \text{and} \qquad \alpha = (1 - (\bar{\alpha} - \alpha_0)) \vee (2 + R\beta\gamma_0)/(2 + R\beta),$$

where $\alpha_0 \in (0, \bar{\alpha})$ and $\gamma_0 \in (\gamma + 2K/R, 1)$.

Remark 3 Although there is a freedom in choosing α_0 and γ_0, the provable contraction coefficient α, ensured by Proposition 2, can be easily shown to satisfy $\alpha > \gamma$, i.e., not surprisingly, α is strictly larger than the contraction coefficient γ in the drift condition, Assumption 1.

Proposition 2 can be restated as saying that P_θ^* is a contraction on the Banach space $\mathcal{L}_{V,0}$. But then its adjoint operator P_θ, having the same norm, is also a contraction. Thus, we get the what is essentially stated in [19, Theorem 1.3]:

Proposition 3 *Under the assumptions of Proposition 2, there exist $\beta > 0$ and $\alpha \in (0, 1)$, such that for all θ, and any signed measure $\eta \in \mathcal{M}_V^0$, we have*

$$\sigma_\beta(P_\theta \eta) \le \alpha \sigma_\beta(\eta). \tag{30}$$

Alternatively, let μ_1, μ_2 be two signed measures on $\mathbf{X}$ as in Definition 4. Then,

$$\sigma_\beta(P_\theta \mu_1, P_\theta \mu_2) \le \alpha \sigma_\beta(\mu_1, \mu_2). \tag{31}$$

In what follows, β and α are chosen as in Proposition 2. Using standard arguments, one can easily show the following proposition, also stated in [19, Theorem 3.2]:

Proposition 4 *Under Assumptions 1 and 2 for all θ, there is a unique probability measure μ_θ^* on $\mathbf{X}$ such that $\mu_\theta^*(V) = \int_{\mathbf{X}} V(x)\, \mu_\theta^*(\mathrm{d}x) < \infty$ and $P_\theta\, \mu_\theta^* = \mu_\theta^*$.*

Remark 4 A mirror image of Proposition 2 given within [20, Theorem 14.0.1] is geometric ergodicity, restated as

$$\sup_{\|\varphi\|_\beta \le 1} \left(\mathbf{E}\left[\varphi(X_n)\,|X_0 = x\right] - \mathbf{E}_{\mu^*}\varphi(x)\right)/(1 + \beta V(x)) \le C\alpha^n.$$

It is readily seen that $|||P^{*n}\, \varphi|||_\beta \le C\alpha^n \|\varphi\|_\beta$ follows, but this is much weaker than Proposition 2. The conditions of [20] are also much different by assuming that the Markov chain is ψ-irreducible and aperiodic. The drift condition is supplemented by a local minorization condition on a "small set" $\mathcal{C}$ defined in terms of an irreducibility measure ψ so that $\psi(\mathcal{C}) > 0$.

3 Existence and uniqueness of the solution of a Poisson equation

In what follows, we consider the Poisson equations, depending on the parameter $\theta \in \Theta$,

$$(I - P_\theta^*)u_\theta(x) = f_\theta(x) - h_\theta, \tag{32}$$

where $f_\theta : \mathbf{X} \to \mathbb{R}$ is the input data, $h_\theta = \mu_\theta^*(f_\theta)$, and $u_\theta : \mathbf{X} \to \mathbb{R}$ is the sought-after solution.

First, we prove the existence and the uniqueness (up to an additive constant) of the solution for a fixed θ, adapting standard arguments, and then, we formulate smoothness conditions on the kernel P_θ^*, and the right-hand side, f_θ. Using these conditions, we prove Lipschitz continuity w.r.t. θ in the norm $\|.\|_\beta$ of the particular solution u_θ for which $\mu_\theta^*(u_\theta) = 0$. For a start, let $\theta \in \Theta$ be fixed.

Theorem 1 *Let $\theta \in \Theta$ be fixed. Let $P = P_\theta$ be such that Assumptions 1 and 2 hold. Let $\beta > 0$ be as given in Proposition 2, and let μ^* denote the unique invariant probability measure of P. Let $f : \mathbf{X} \to \mathbb{R}$ be a measurable function such that $|||f|||_\beta < \infty$, and let $h = \mu^*(f)$. Then, the Poisson equation*

$$(I - P^*)u(x) = f(x) - h \tag{33}$$

has a unique solution $u(\cdot)$ up to an additive constant. The particular solution for which $\mu^*(u) = 0$ can be written as

$$u(x) = \sum_{n=0}^{\infty} (P^{*n} f(x) - h), \tag{34}$$

where the right-hand side is absolutely convergent, and

$$|u(x)| \le |||f|||_\beta \, K_u(1 + \beta V(x)), \tag{35}$$

for some constant $K_u > 0$ depending only on the constants appearing in Assumptions 1, 2, given by

$$K_u := \frac{1}{1-\alpha}\left(2 + \beta \frac{K}{1-\gamma}\right).$$

It also follows that $\|u\|_\beta \le K_u |||f|||_\beta < \infty$.

Proof It is immediate to check that (33) is formally satisfied by u. We show that u is well defined. First, consider any function φ such that $|||\varphi|||_\beta \le 1$. By the definition of the metric σ_β, see (26), the inequality

$$\left| \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x) \right| \le \sigma_\beta(\mu_1, \mu_2) \tag{36}$$

holds true for any pair of probability measures μ_1, μ_2, or even for any pair of signed measures μ_1, μ_2 as in Definition 4. On the other hand, any generic function φ can be rescaled by $\frac{1}{|||\varphi|||_\beta}$, so that we also have

$$\left| \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x) \right| \leq |||\varphi|||_\beta \sigma_\beta(\mu_1, \mu_2). \tag{37}$$

To estimate the n th term of the right-hand side of (34), consider the equalities

$$\begin{aligned} \frac{1}{|||f|||_\beta} |P^{*n} f(x) - \mu^*(f)| &= \frac{1}{|||f|||_\beta} |(P^n \delta_x - \mu^*)(f)| \\ &= \frac{1}{|||f|||_\beta} \left| \int_{\mathbf{X}} f(y)(P^n \delta_x - P^n \mu^*)(\mathrm{d}y) \right|. \end{aligned} \tag{38}$$

Using (37), we can bound the right-hand side by $\sigma_\beta(P^n \delta_x, P^n \mu^*)$. Now applying Proposition 3 and taking into account Corollary 1, we can further bound it by

$$\begin{aligned} \sigma_\beta(P^n \delta_x, P^n \mu^*) &\leq \alpha^n \sigma_\beta(\delta_x, \mu^*) \\ &= \alpha^n \sup_{\varphi : \|\varphi\|_\beta \leq 1} \int_{\mathbf{X}} \varphi(x)(\delta_x - \mu^*)(\mathrm{d}x). \end{aligned} \tag{39}$$

Take into account the trivial estimate

$$\int_{\mathbf{X}} \varphi(x)(\delta_x - \mu^*)(\mathrm{d}x) \leq \int_{\mathbf{X}} |\varphi(x)|(\delta_x + \mu^*)(\mathrm{d}x), \tag{40}$$

and note that $\|\varphi\|_\beta \leq 1$ implies $|\varphi(x)| \leq 1 + \beta V(x)$ for all x. Putting together (38) - (40) with the fact that $\int_{\mathbf{X}}(1 + \beta V(x))(\delta_x + \mu^*)(\mathrm{d}x) = 2 + \beta V(x) + \beta \mu^*(V)$, we conclude:

$$\frac{1}{|||f|||_\beta} |P^{*n} f(x) - \mu^*(f)| \leq \alpha^n (2 + \beta V(x) + \beta \mu^*(V)). \tag{41}$$

It follows that the series $\sum_{n=0}^{\infty}(P^{*n} f(x) - h)$ is absolutely convergent, so $u(x)$ is well defined and satisfies the desired upper bound. Indeed, $(P^* u)(x)$ can be written as

$$\int_{\mathbf{X}} P(x, \mathrm{d}y) u(y) = \int_{\mathbf{X}} P(x, \mathrm{d}y) \sum_{n=0}^{\infty} (P^{*n} f(y) - h), \tag{42}$$

where the integration and the summation can be interchanged due to the Lebesgue dominated convergence theorem, the conditions of which are ensured by (41). Thus, we get

$$(P^* u)(x) = \sum_{n=1}^{\infty} (P^{*n} f(x) - h) = u(x) - (f(x) - h), \tag{43}$$

which implies the claim. Using similar arguments, and Fubini's theorem as in (A3) of the Appendix, we get that

$$\int_{\mathbf{X}} u(x)\mu^*(\mathrm{d}x) = 0. \tag{44}$$

To prove uniqueness, assume that there are two solutions u_1 and u_2, and define $\Delta u = u_2 - u_1$. Then, $(I - P^*)\Delta u = 0$, implying $P^*\Delta u = \Delta u$, from which $|||P^*\Delta u|||_\beta = |||\Delta u|||_\beta$. But, by Proposition 2, $|||P^*\Delta u|||_\beta \le \alpha|||\Delta u|||_\beta$, and hence $|||\Delta u|||_\beta = 0$. Hence, Δu is a constant.

Summing the inequalities (41) over n and using (14), we get

$$|u(x)| \le \frac{|||f|||_\beta}{1-\alpha}\left(2 + \beta V(x) + \beta\frac{K}{1-\gamma}\right), \tag{45}$$

from which the claim of the theorem follows after trivial arithmetics. □

4 Lipschitz continuity of the Kernel

Now, we consider a parametric family of kernels (P_θ). A critical point in the discussion is to define appropriate smoothness conditions for them in the context of [19].

Assumption 3 (Lipschitz Continuity of P_θ) There exists a constant L_P such that for every $\theta, \theta' \in \Theta$ and all $x \in \mathbf{X}$:

$$\sigma_\beta(P_\theta\delta_x, P_{\theta'}\delta_x) \ \le\ L_P|\theta - \theta'|(1 + \beta V(x)). \tag{46}$$

This assumption can be rewritten in the equivalent form: For any $f \in \mathcal{L}_\mathcal{V}$, we have, with $\beta > 0$ as in Proposition 2,

$$|P_\theta\delta_x(f) - P_{\theta'}\delta_x(f)| \le L_P\|f\|_\beta|\theta - \theta'|(1 + \beta V(x)).$$

An inequality analogous to (46) with a general measure μ replacing δ_x is established in the following lemma:

Lemma 1 *Suppose that Assumption 3 holds and that Assumption 1 holds without requiring $\gamma < 1$. Let μ be a measure with $\mu(1 + \beta V) < \infty$. Then, for all $\theta, \theta' \in \Theta$:*

$$\sigma_\beta(P_\theta\mu, P_{\theta'}\mu) \ \le\ L_P|\theta - \theta'|\mu(1 + \beta V). \tag{47}$$

The starting point of the proof is the observation that Assumption 3 implies that for all φ such that $\|\varphi\|_\beta \le 1$, implying also $|||\varphi|||_\beta \le 1$, we have

$$\int_{\mathbf{X}} \varphi(y)\left(P_\theta(x, \mathrm{d}y) - P_{\theta'}(x, \mathrm{d}y)\right) \le L_P|\theta - \theta'|(1 + \beta V(x)). \tag{48}$$

Integrating this inequality with respect to $\mu(\mathrm{d}x)$, the right-hand side of (48) becomes the right-hand side of (47). For the integral of the left-hand side, we apply Fubini's theorem to get the claim of the lemma. Details will be given in the Appendix.

The relaxed version of Assumption 1 not requiring $\gamma < 1$ will be referred to as *uniform one-step growth condition*. It is analogous to Assumption A'.5, (i') on page 290 in [4] and will play a dominant role in Sect. 6. It is readily seen that the corollaries of Assumption 1 given as inequalities (11) and (12) remain valid.

Lemma 1 readily extends to signed measures:

Lemma 2 *Let the conditions of Lemma 1 be satisfied, and let η be a signed measure such that $|\eta|(1+\beta V) < \infty$. Then:*

$$\sigma_\beta(P_\theta \eta, P_{\theta'}\eta) \le L_P|\theta-\theta'||\eta|(1+\beta V). \tag{49}$$

The proof is based on using the Hahn–Jordan decomposition $\eta = \eta^+ - \eta^-$, where η^+ and η^- are nonnegative measures with disjoint supports. Details will be given in the Appendix. The previous results culminate in what follows, stating the Lipschitz continuity of P_θ^n when acting on signed measures:

Lemma 3 *Let (P_θ) satisfy Assumptions 1 and 2. Let $\beta > 0$ be such that Proposition 3 holds. Let Assumption 3, requiring the Lipschitz continuity of (P_θ), hold with the above β. Then for any signed measure η with $|\eta|(1+\beta V) < \infty$ and $\theta, \theta' \in \Theta$:*

$$\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta) \le L_P|\theta-\theta'| \cdot C_P|\eta|(1+\beta V), \tag{50}$$

where C_P is independent of θ, θ' and η, and is given by

$$C_P := \frac{1}{1-\alpha}\left(1+\beta\frac{K}{1-\gamma}\right) \vee \frac{1}{\alpha-\gamma}. \tag{51}$$

The proof is based on the telescopic sum decomposition:

$$(P_\theta^n - P_{\theta'}^n) = \sum_{k=0}^{n-1}(P_\theta^{n-k}P_{\theta'}^k - P_\theta^{n-k-1}P_{\theta'}^{k+1}). \tag{52}$$

For the k-th term, we use the contraction property of P_θ^{n-k-1}, Proposition 3, and estimate $\sigma_\beta(P_\theta P_{\theta'}^k \eta - P_{\theta'}P_{\theta'}^k\eta)$ from above using the Lipschitz continuity of the kernels P_θ as defined in Assumption 3. We will also need the observation that iterating the drift condition given in the form (12) we get

$$|P_\theta^k \eta|(V) \le \gamma^k|\eta|(V) + \frac{K}{1-\gamma}|\eta|(\mathbf{X}), \tag{53}$$

restated in (A9) with details given in the Appendix.

Applying Lemma 3 for $\eta = \delta_x$, we get:

$$\sigma_\beta(P_\theta^n\delta_x, P_{\theta'}^n\delta_x) \le L_P|\theta-\theta'| \cdot C_P(1+\beta V(x)). \tag{54}$$

Letting $n \to \infty$, and taking into account Proposition 3, we get the Lipschitz continuity of the invariant measure w.r.t. θ:

Corollary 2 *Under the assumptions of Lemma 3, we get*

$$\sigma_\beta(\mu^*_\theta, \mu^*_{\theta'}) \leq L_P|\theta - \theta'|\cdot C_P. \tag{55}$$

The constant C_P can be replaced by the smaller constant

$$C'_P := \frac{1}{1-\alpha}\left(1 + \beta \frac{K}{1-\gamma}\right). \tag{56}$$

Details of the proof will be given in the Appendix.

A surprising variant of the above lemma is the following:

Lemma 4 *Under the conditions of Lemma 3 for every signed measure $\eta \in \mathcal{M}^0_V$, implying $\eta(\mathbf{X}) = 0$, and $\theta, \theta' \in \Theta$, we have*

$$\sigma_\beta(P^n_\theta \eta, P^n_{\theta'}\eta) \leq L_P|\theta - \theta'|n\alpha^{n-1}\,|\eta|(1 + \beta V). \tag{57}$$

Equivalently, we can write: For any $f \in \mathcal{L}_\mathcal{V}$, we have

$$|P^n_\theta \eta\,(f) - P^n_{\theta'}\eta\,(f)| \leq L_P\|f\|_\beta\,|\theta - \theta'|n\alpha^{n-1}\,|\eta|(1 + \beta V).$$

The starting point of the proof is the same telescopic decomposition (52), applied to η. The novelty w.r.t. to the proof of Lemma 3 is the observation that $\eta(\mathbf{X}) = 0$ implies that $P^k_{\theta'}\eta$ converges exponentially fast to the zero measure, see Proposition 3. Details will be given in the Appendix.

A final note: a key technical result of [4] on the Lipschitz continuity of the solutions of the Poisson equation, stated as Theorem 6, p. 262, assumes smoothness of the kernels in a way which is in most aspects significantly more restrictive than our Assumption 3. In particular, we can reformulate assumptions (iii) and (iv) of [4, Theorem 6], adapted to the context and notation of this paper, as follows.

Let $\mathbf{X} := \mathbb{R}^k$ be a Euclidean space, and let $V(x) = |x|^p$ with $p > 0$. Letting $\beta = 1$ write $\|g\|_1 =: \|g\|_V$. Let $g(\cdot)$ be differentiable, and let $g'(\cdot)$ denote its gradient. Then, the modulated version of assumption (iii) would read:

$$|P^n_\theta \delta_x\,(g) - P^n_{\theta'}\delta_x\,(g)| \leq L_P(\|g\|_V + \|g'\|_V)\,|\theta - \theta'|\,(1 + V(x)),$$

for any pair $\theta,\ \theta' \in \Theta$ and any $x \in \mathbb{R}^k$. To modulate (and relax) assumption (iv), let $\eta \in \mathcal{M}^0_V$ and let the test functions $g(\cdot)$ be twice differentiable. Then, impose:

$$\left|P^n_\theta \eta\,(g) - P^n_{\theta'}\eta\,(g)\right| \leq L_p\,\rho^n\,\|g\|_{3,V}\,|\theta - \theta'|\,|\eta|(1 + V),$$

with $0 < \rho < 1$ and $\|g\|_{3,V} = (\|g\|_V + \|g'\|_V + \|g''\|_V)$. Note in passing that assumption (iv) in [4] is formulated in terms of measures of the form $\eta = \delta_{x_1} - \delta_{x_2}$, however, it can be readily extended to measures $\eta \in \mathcal{M}^0_V$ via Choquet's theorem.

The strength of these assumptions, or those of [4], is that the set of test functions are strictly smaller than $\mathcal{L}_V$, needed for our Assumption 3. On the other hand, the effect of the probability transition kernels $P^n_\theta \eta$ on the space of measures on $\mathbf{X}$ is not discussed at all in [4], and thus, the dual perspective characterizing [19] is absent. In particular, the distances of measures $P^n_\theta \eta$ and $P^n_{\theta'} \eta$, in any suitable metric, are not estimated. It is questionable if the methodology of [19] can be extended even to the case $\mathbf{X} = \mathbb{R}^d$ using a smaller class of test functions such as satisfying $\|\varphi\|_{2,\beta} := \|\varphi\|_\beta + \|\varphi'\|_\beta < \infty$. For such an extension, we may wish to redefine the metrics $\rho_\beta(\mu_1, \mu_2)$, taking into account (18), as follows. Let μ_1, μ_2 be two signed measures on $\mathbf{X}$ such that $\int_{\mathbf{X}} (1 + \beta V(x)) |\mu_i|(\mathrm{d}x) < \infty$, for $i = 1, 2$, and $\mu_1(\mathbf{X}) = \mu_2(\mathbf{X})$. Then, define the distance

$$\kappa_\beta(\mu_1, \mu_2) := \sup_{\varphi : \|\varphi\|_{2,\beta} \le 1} \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x). \tag{58}$$

To the best of our knowledge, this metric has not been explored in the stability theory of Markov chains.

More grievously, both assumptions (iii) and (iv) of Theorem 6, Chapter 2, Part II of [4], p. 262, assume some kind of a priori stability of the kernels P^n_θ, while similar inequalities are actually proven in our Lemmas 3 and 4 under practically attractive assumptions. In particular, the arguments of [4] presented in Section 2.3 for linear stochastic systems do not seem to be appropriate for queuing systems, discussed in Sect. 8 of this paper, since the mapping

$$X_{n+1} = (X_n + U_{n+1})_+,$$

defining their dynamics is not contractive in X. In addition, for the analysis of this process a Lyapunov functions of the form $V(x) = |x|^p$ with $p \ge 1$, used overall in [4], are inappropriate. What we need is an exponential Lyapunov function of the form $V(x) = e^{cx}$, known to be the right tools for studying stability of queuing systems, see [20], Section 16.4.

5 Lipschitz continuity of the solution of the Poisson equation

In this section, we revisit the family of Poisson equations, depending on the parameter $\theta \in \Theta$, defined under (32). In addition to the Lipschitz continuity of the kernel P^*_θ, given in Assumption 3, we need to formulate the Lipschitz continuity of the right-hand side, (f_θ) with $\theta \in \Theta$, as well.

Assumption 4 (Lipschitz Continuity of f_θ) We have $K_f := \sup_{\theta \in \Theta} |||f_\theta|||_\beta < \infty$, and there exists a constant L_f such that, for all θ, θ', it holds that

$$\|f_\theta - f_{\theta'}\|_\beta \le L_f |\theta - \theta'|. \tag{59}$$

The main result of this paper is that under the above conditions the particular solution u_θ, for which $\mu^*_\theta(u_\theta) = 0$, is Lipschitz continuous w.r.t. θ in the norm $\|.\|_\beta$:

Theorem 2 *Assume that the kernels* (P_θ) *satisfy Assumptions 1, 2 and 3. Let us fix* $\beta > 0$ *as given in Proposition 2. Let* (f_θ) *be a family of* $\mathbf{X} \to \mathbb{R}$ *measurable functions such that Assumption 4 holds with the above* β*. Let* μ_θ^* *denote the unique invariant probability measure of* P_θ*, and let* $h_\theta = \mu_\theta^*(f_\theta)$*. Consider the Poisson equations*

$$(I - P_\theta^*)u_\theta(x) = f_\theta(x) - h_\theta. \tag{60}$$

Then, h_θ *is Lipschitz continuous in* θ*:*

$$|h_\theta - h_{\theta'}| \leq L_h|\theta - \theta'|, \tag{61}$$

and the particular solution, given by (34) as $u_\theta(x) = \sum_{n=0}^{\infty}(P_\theta^{*n} f_\theta(x) - h_\theta)$ *is Lipschitz continuous in* θ*:*

$$|u_\theta(x) - u_{\theta'}(x)| \leq L_u|\theta - \theta'| \left(1 + \beta V(x)\right), \tag{62}$$

where L_u *is independent of* x*. Alternatively, we can write*

$$\|u_\theta - u_{\theta'}\|_\beta \leq L_u|\theta - \theta'|. \tag{63}$$

Here, the constants L_h *and* L_u *depend only on the constants appearing in Assumptions 1, 2, 3 and 4.*

Proof Consider the extended parametric family of Poisson equations, where P^* and f are independently parameterized, with the notation $h_{\theta,\psi} = \mu_\theta^*(f_\psi)$,

$$(I - P_\theta^*)u_{\theta,\psi}(x) = f_\psi(x) - h_{\theta,\psi}, \tag{64}$$

Step 1. First, we prove that $h_{\theta,\psi}$ is Lipschitz continuous in θ and ψ. Since $h_\theta = \mu_\theta^*(f_\theta) = h_{\theta,\theta}$, the Lipschitz continuity of h_θ, stated in (61), then follows. We can write

$$|h_{\theta,\psi} - h_{\theta,\psi'}| = \lim_{n\to\infty} |P_\theta^{*n} f_\psi(x) - P_\theta^{*n} f_{\psi'}(x)|, \tag{65}$$

$$|h_{\theta,\psi} - h_{\theta',\psi}| = \lim_{n\to\infty} |P_\theta^{*n} f_\psi(x) - P_{\theta'}^{*n} f_\psi(x)|. \tag{66}$$

Note that the limits of the right-hand side are finite by Assumption 4 and the drift condition Assumption 1.

We can bound the right-hand side of (65) as follows:

$$|P_\theta^{*n} f_\psi(x) - P_\theta^{*n} f_{\psi'}(x)| \leq \left(P_\theta^{*n}|f_\psi - f_{\psi'}|\right)(x)$$

$$= \left(P_\theta^n \delta_x\right)|f_\psi - f_{\psi'}| \leq \left\| f_\psi - f_{\psi'}\right\|_\beta \left(P_\theta^n \delta_x\right)(1 + \beta V). \tag{67}$$

Using the Lipschitz continuity of f, as given by Assumption 4, the right-hand side can be bounded from above by

$$L_f|\psi - \psi'|\left(P_\theta^n \delta_x\right)(1 + \beta V). \tag{68}$$

Note that $\left\|\left|P_\theta^n \delta_x - \mu_\theta^*\right|\right\|_\beta \to 0$ as $n \to \infty$, by Corollary, 1 and, hence, $\|P_\theta^n \delta_x - \mu_\theta^*\|_\beta \to 0$. By Remark 1, this allows us to upper bound the limit of (68), and consequently the l.h.s. (left-hand side) of (65), as follows:

$$L_f|\psi - \psi'|\mu_\theta^*(1+\beta V) \leq L_f|\psi - \psi'| \left(1 + \beta \frac{K}{1-\gamma}\right). \tag{69}$$

Now, the l.h.s. of (66) can be written as and bounded by

$$\left|\int_{\mathbf{X}} f_\psi(x)(\mu_\theta^* - \mu_{\theta'}^*)(\mathrm{d}x)\right| \leq \left\|\left| f_\psi \right|\right\|_\beta \sigma_\beta(\mu_\theta^*, \mu_{\theta'}^*). \tag{70}$$

Here, $\left\|\left| f_\psi \right|\right\|_\beta \leq \sup_{\psi\in\Theta} \left\|\left| f_\psi \right|\right\|_\beta = K_f < \infty$ by Assumption 4, and $\sigma_\beta(\mu_\theta^*, \mu_{\theta'}^*) \leq L_P|\theta - \theta'| \cdot C_P$ by Corollary 2.

Setting $\psi = \theta$, $\psi' = \theta'$ in (65) and $\psi = \theta'$ in (66), we get by the triangle inequality (61): $|h_\theta - h_{\theta'}| \leq L_h|\theta - \theta'|$, with

$$\begin{aligned} L_h &= L_f\left(1 + \beta\frac{K}{1-\gamma}\right) + K_f L_P \frac{1}{1-\alpha}\left(1 + \beta\frac{K}{1-\gamma}\right) \\ &= \left(L_f + K_f L_P \frac{1}{1-\alpha}\right)\left(1 + \beta\frac{K}{1-\gamma}\right). \end{aligned} \tag{71}$$

With this, Step 1 is completed. Next, we consider the Lipschitz continuity of the doubly parameterized particular solution

$$u_{\theta,\psi}(x) = \sum_{n=0}^{\infty}(P_\theta^{*n} f_\psi(x) - h_{\theta,\psi}). \tag{72}$$

Step 2. We show that $u_{\theta,\psi}(x)$ is Lipschitz continuous w.r.t. ψ. Indeed, we can express $|u_{\theta,\psi}(x) - u_{\theta,\psi'}(x)|$ as

$$\begin{aligned} &\left|\sum_{n=0}^{\infty}(P_\theta^{*n}(f_\psi(x) - f_{\psi'}(x)) - (h_{\theta,\psi} - h_{\theta,\psi'})\right| \\ &= \left|\sum_{n=0}^{\infty}(P_\theta^n \delta_x - \mu_\theta^*)(f_\psi - f_{\psi'})\right|. \end{aligned} \tag{73}$$

The absolute value of n-th term can be written, using (37), as

$$|(P_\theta^n \delta_x - P_\theta^n \mu_\theta^*)(f_\psi - f_{\psi'})| \leq \sigma_\beta(P_\theta^n(\delta_x - \mu_\theta^*)) \left\|\left|(f_\psi - f_{\psi'})\right|\right\|_\beta.$$

Taking into account Propositions 3 and 1, (16), and Assumption 4, the right-hand side can be bounded from above by

$$\begin{aligned}&\alpha^n \sigma_\beta(\delta_x - \mu_\theta^*) \cdot L_f |\psi - \psi'| \\ &\quad \le \alpha^n (\delta_x + \mu_\theta^*)(1 + \beta V) \cdot L_f |\psi - \psi'|. \end{aligned} \tag{74}$$

The right-hand side is equal to and can be upper bounded by

$$\begin{aligned}&\alpha^n \left(2 + \beta V(x) + \beta \mu_\theta^*(V)\right) \cdot L_f |\psi - \psi'| \\ &\le \alpha^n \left(2 + \beta V(x) + \beta \frac{K}{1-\gamma}\right) \cdot L_f |\psi - \psi'|. \end{aligned} \tag{75}$$

Inserting this into (73), we get the upper bound

$$\begin{aligned}&|u_{\theta,\psi}(x) - u_{\theta,\psi'}(x)| \\ &\le \frac{1}{1-\alpha} \left(2 + \beta V(x) + \beta \frac{K}{1-\gamma}\right) L_f |\psi - \psi'|. \end{aligned} \tag{76}$$

Step 3. The final, critical point is to show that $u_{\theta,\psi}(x)$ is Lipschitz continuous in θ. Let us write $u_{\theta,\psi}(x) - u_{\theta',\psi}(x)$ as

$$\sum_{n=0}^{\infty} (P_\theta^{*n} f_\psi(x) - \mu_\theta^*(f_\psi) - P_{\theta'}^{*n} f_\psi(x) + \mu_{\theta'}^*(f_\psi)). \tag{77}$$

The n-th term can be written as

$$\left(P_\theta^n \delta_x - P_\theta^n \mu_\theta^* - P_{\theta'}^n \delta_x + P_{\theta'}^n \mu_{\theta'}^*\right)(f_\psi). \tag{78}$$

Let us denote the measure acting on f_ψ by Δ_n. Adding and subtracting $P_{\theta'}^n \mu_\theta^*$ within Δ_n, we can write

$$\begin{aligned}\Delta_n &= [P_\theta^n \left(\delta_x - \mu_\theta^*\right) - P_{\theta'}^n \left(\delta_x - \mu_\theta^*\right)] + [P_{\theta'}^n \left(\mu_{\theta'}^* - \mu_\theta^*\right)] \\ &=: \Delta_{n,1} + \Delta_{n,2}. \end{aligned} \tag{79}$$

With this notation (78), the n-th term of (77) can be written and bounded from above in absolute value, using (37), as

$$|\Delta_n(f_\psi)| \le \sigma_\beta(\Delta_n) |||f_\psi|||_\beta = \sigma_\beta(\Delta_{n,1} + \Delta_{n,2}) |||f_\psi|||_\beta. \tag{80}$$

To bound $\sigma_\beta(\Delta_{n,1})$, we can apply Lemma 4 with $\eta = \delta_x - \mu_\theta^*$:

$$\sigma_\beta(\Delta_{n,1}) \le L_P |\theta - \theta'| n \alpha^{n-1} |\delta_x - \mu_\theta^*| (1 + \beta V). \tag{81}$$

The right-hand side can be trivially upper bounded by

$$\begin{aligned} &L_P|\theta - \theta'|n\alpha^{n-1}\left(2 + \beta V(x) + \beta\mu_\theta^*(V)\right) \\ &\leq L_P|\theta - \theta'|n\alpha^{n-1}\left(2 + \beta V(x) + \beta\frac{K}{1-\gamma}\right). \end{aligned} \tag{82}$$

To bound $\sigma_\beta(\Delta_{n,2})$, we refer to Proposition 3, yielding

$$\sigma_\beta(\Delta_{n,2}) \leq \alpha^n \sigma_\beta(\mu_{\theta'}^* - \mu_\theta^*).$$

This can be bounded from above by Corollary 2, resulting in

$$\sigma_\beta(\Delta_{n,2}) \leq \alpha^n L_P \frac{1}{1-\alpha}\left(1 + \beta\frac{K}{1-\gamma}\right)|\theta - \theta'|. \tag{83}$$

Combining (80)–(83), we get that the n-th term of (77), reformulated as (78), can be bounded from above by

$$\begin{aligned} &L_P\, n\alpha^{n-1}\left(2 + \beta V(x) + \beta\frac{K}{1-\gamma}\right)K_f\, |\theta - \theta'| \\ &+ L_P\, \alpha^n \frac{1}{1-\alpha}\left(1 + \beta\frac{K}{1-\gamma}\right)K_f\, |\theta - \theta'|. \end{aligned} \tag{84}$$

Summation over n, in view of (77), yields the upper bound

$$\begin{aligned} &|u_{\theta,\psi}(x) - u_{\theta',\psi}(x)| \\ &\leq L_P \frac{1}{(1-\alpha)^2}\left(2 + \beta V(x) + \beta\frac{K}{1-\gamma}\right)K_f\, |\theta - \theta'| \\ &+ L_P \frac{1}{(1-\alpha)^2}\left(1 + \beta\frac{K}{1-\gamma}\right)K_f\, |\theta - \theta'|. \end{aligned} \tag{85}$$

Combining (85) and (76), the proof is complete. □

Remark 5 The Lipschitz constant L_u can be chosen as

$$L_u = \frac{1}{1-\alpha}\left(L_f + L_P\frac{2}{(1-\alpha)}K_f\right)\left(2 + \beta\frac{K}{1-\gamma}\right). \tag{86}$$

The details of this elementary calculation are omitted.

6 Relaxations of the conditions

In this section, we restate the results of [19] cited in Sect. 2 as Propositions 2–4 under relaxed conditions, obtained by additional arguments. To facilitate reading and to highlight the parallel structure, we state them as Propositions 5–7.

A key condition of Propositions 2–4 is Assumption 1, requiring the existence of a common Lyapunov function. To illustrate the delicacy of this assumption, consider the analogous scenario that a set of $n \times n$ matrices $\{A_\theta : \theta \in \Theta \subset \mathbf{R}^p\}$, such that A_θ is stable for all $\theta \in \Theta$, has a common quadratic Lyapunov function $V(x) := x^\top Q x$, where Q is a symmetric positive definite matrix. In fact we would require that for some $0 < \gamma < 1$ we have $A_\theta^\top Q A_\theta \le \gamma Q$ for all $\theta \in \Theta$ in the sense of semi-definite ordering. Hence, the matrix Q induces a metric with respect to which A_θ is a contraction, simultaneously for all θ, with the same contraction factor γ, and thus, the family of matrices $\{A_\theta : \theta \in \Theta\}$ is jointly stable.

Now, if joint stability fails to hold, but $\{A_\theta : \theta \in \Theta\}$ is a compact set of stable $n \times n$ matrices, then we can find a positive integer r such that $\|A_\theta^r\| \le \gamma_r < 1$ for all $\theta \in \Theta$, and hence, the family of matrices $\{A_\theta^r : \theta \in \Theta\}$ is jointly stable. This analogy motivates the following relaxation of the drift condition given as Assumption 1, similar to Assumption A'.5, (i) and (i') on page 290 of [4]:

Assumption 5 (Uniform Drift Condition for P_θ^r) There exists a positive integer r, a measurable function $V : \mathbf{X} \to [0, \infty)$ and constants $\gamma_r \in (0, 1)$ and $K_r \ge 0$ such that for all $\theta \in \Theta$ and $x \in \mathbf{X}$, we have

$$(P_\theta^{*r} V)(x) \le \gamma_r V(x) + K_r. \tag{87}$$

Assumption 6 (Uniform One-Step Growth Condition for P_θ) With the same measurable function $V : \mathbf{X} \to [0, \infty)$ as above, we have for all $\theta \in \Theta$ and all $x \in \mathbf{X}$:

$$(P_\theta^* V)(x) \le \gamma_1 V(x) + K_1, \tag{88}$$

where we can and will assume that $\gamma_1 > 1$ and $K_1 \ge 0$.

Integrating (88) with respect to any $\mu \in \mathcal{M}_V$, we get

$$P_\theta \mu(V) \le \gamma_1 \mu(V) + K_1 \mu(\mathbf{X}), \tag{89}$$

for all $\theta \in \Theta$, in exact analogy with (11). The following lemma is a relaxed version of Proposition 2:

Lemma 5 *The uniform one-step growth condition given above implies that for any* $\beta > 0$, *for all functions* $\varphi \in \mathcal{L}_V$ *and all* $\theta \in \Theta$, *with* $\alpha' = \gamma_1 \vee (1 + \beta K_1)$, *we have*

$$\|P_\theta^* \varphi\|_\beta \le \alpha' \|\varphi\|_\beta \quad \textit{and} \quad \left|\left|\left| P_\theta^* \varphi \right|\right|\right|_\beta \le \alpha' |||\varphi|||_\beta. \tag{90}$$

The first statement is obtained by straightforward calculations, while the second statement follows by using the definition $|||\psi|||_\beta = \min_c \|\psi + c\|_\beta$. For details, see the Appendix.

Lemma 6 *Under Assumption 6, for any* $\beta > 0$ *and* $\theta \in \Theta$, *the kernel* P_θ *is a bounded linear operator on* $\mathcal{M}_V^0$; *specifically, with* α' *as in Lemma 5, for any* $\eta \in \mathcal{M}_V^0$, *we have*

$$\sigma_\beta(P_\theta \eta) \le \alpha' \sigma_\beta(\eta). \tag{91}$$

Alternatively, let μ_1, μ_2 be two possibly signed measures on $\mathbf{X}$ as in Definition 4. Then,

$$\sigma_\beta(P_\theta \mu_1, P_\theta \mu_2) \le \alpha' \sigma_\beta(\mu_1, \mu_2). \tag{92}$$

Assumption 7 (Uniform Local Minorization for P_θ^r). Let $R_r > 2K_r/(1-\gamma_r)$, where γ_r and K_r are the constants from Assumption 5, and let $\mathcal{C}_r = \{x \in \mathbf{X} : V(x) \le R_r\}$. There exist a probability measure $\bar{\mu}_r$ and a constant $\bar{\alpha}_r \in (0, 1)$ such that for all $\theta \in \Theta$, $x \in \mathcal{C}_r$ and measurable A it holds:

$$P_\theta^r(x, A) \ge \bar{\alpha}_r \bar{\mu}_r(A). \tag{93}$$

The first main result of [19], cited as Proposition 2 in Sect. 2, can be restated as follows:

Proposition 5 *Under Assumptions 5, 6 and 7, there exist constants $\beta > 0$, $\alpha \in (0, 1)$ and $C > 0$ such that for any $\varphi \in \mathcal{L}_V$, all $\theta \in \Theta$ and all $n > 0$ we have:*

$$\left|\left|\left| P_\theta^{*n} \varphi \right|\right|\right|_\beta \le C\alpha^n |||\varphi|||_\beta. \tag{94}$$

Here, we can choose $\beta = \beta_r$ and $\alpha = \alpha_r^{1/r}$, with β_r and α_r provided by Proposition 2 applied to P_θ^r, and noting that $\alpha_r \in (0, 1)$ and $C = \alpha_r^{-1}(\alpha')^{r-1}$ with α' as in Lemma 5.

Proof Let us fix a $\theta \in \Theta$ and write $P = P_\theta$. By Proposition 2, there exist $\beta = \beta_r > 0$, and $\alpha_r \in (0, 1)$ such that $|||P^{*r}\varphi|||_\beta \le \alpha_r |||\varphi|||_\beta$, implying for any positive integer m

$$\left|\left|\left| P^{*rm} \varphi \right|\right|\right|_\beta \le \alpha_r^m |||\varphi|||_\beta. \tag{95}$$

For a general positive integer n, write $n = rm + k$ with $0 \le k \le r - 1$. Then, we get

$$\left|\left|\left| P^{*n} \varphi \right|\right|\right|_\beta \le \alpha_r^m \left|\left|\left| P^{*k} \varphi \right|\right|\right|_\beta. \tag{96}$$

To complete the proof, estimate the last term above applying the second inequality of (90) $k \le r - 1$ times to obtain

$$\left|\left|\left| P^{*n} \varphi \right|\right|\right|_\beta \le \alpha_r^m (\alpha')^{r-1} |||\varphi|||_\beta. \tag{97}$$

Now, $m = (n-k)/r > n/r - 1$, hence $\alpha_r^m < \alpha_r^{n/r}\alpha_r^{-1}$, and thus, the proposition follows. ◻

Repeating the arguments leading to Proposition 3, we get:

Proposition 6 *Under Assumptions 5, 6 and 7, there exist constants $\beta = \beta_r > 0$, $\alpha \in (0, 1)$ and $C > 0$ such that for all signed measure $\eta \in \mathcal{M}_V^0$, θ and $n \ge 0$, we have*

$$\sigma_\beta(P_\theta^n \eta) \le C\alpha^n \sigma_\beta(\eta). \tag{98}$$

The constants $\beta = \beta_r$, α *and* C *are the same as in Proposition 5. Alternatively, let* μ_1, μ_2 *be two possibly signed measures on* $\mathbf{X}$ *as in Definition 4. Then, we have*

$$\sigma_\beta(P_\theta^n \mu_1, P_\theta^n \mu_2) \le C\alpha^n \sigma_\beta(\mu_1, \mu_2). \tag{99}$$

Finally, we have the following extension of Proposition 4:

Proposition 7 *Under Assumptions 5, 6 and 7 for all* $\theta \in \Theta$*, there exists a unique probability measure* μ_θ^* *on* $\mathbf{X}$ *such that* $\mu_\theta^*(V) < \infty$ *and* $P_\theta \mu_\theta^* = \mu_\theta^*$*. Denoting the unique invariant probability measure for* P_θ^r *by* $\mu_{\theta,r}^*$*, we have* $\mu_\theta^* = \mu_{\theta,r}^*$*.*

Proof Let us fix any $\theta \in \Theta$ and write $P = P_\theta$, $\mu^* = \mu_\theta^*$ and $\mu_r^* = \mu_{\theta,r}^*$. Thus, μ_r^* is the unique invariant probability measure for P^r the existence of which is ensured by Proposition 4. Now, we show that for any $k > 0$ we have $\int_{\mathbf{X}} V \,\mathrm{d}P^k \mu_r^* := \int_{\mathbf{X}} V(x)\, P^k \mu_r^*(\mathrm{d}x) < \infty$. Indeed, write:

$$\int_{\mathbf{X}} V \,\mathrm{d}P^k \mu_r^* = \int_{\mathbf{X}} (P^k)^* V \,\mathrm{d}\mu_r^* = \int_{\mathbf{X}} (P^*)^k V \,\mathrm{d}\mu_r^*. \tag{100}$$

The r.h.s. can be upper bounded, using the definition of $\|\cdot\|_\beta$ and the first half of (90), by

$$\int_{\mathbf{X}} \|(P^*)^k V\|_\beta (1 + \beta V)\,\mathrm{d}\mu_r^* \le \int_{\mathbf{X}} (\alpha')^k (1 + \beta V)\,\mathrm{d}\mu_r^*, \tag{101}$$

which is finite since $\int_{\mathbf{X}} V(x)\,\mathrm{d}\mu_r^*(x) < \infty$. Then, the probability measure μ^* defined by

$$\mu^* := \frac{1}{r}(I + P + \ldots P^{r-1})\mu_r^* \tag{102}$$

also satisfies $\int_{\mathbf{X}} V(x)\,\mathrm{d}\mu^*(x) < \infty$, and it is readily seen to be invariant for P. Since any probability measure invariant for P is also invariant for P^r, there cannot be measures that are invariant for P besides μ_r^*, and thus, we have $\mu^* = \mu_r^*$. □

7 Analysis of the Poisson equation under relaxed conditions

In what follows, the main results of Sects. 3 and 5 will be now extended, with minor modifications, assuming the above relaxed conditions.

Theorem 3 *Let Assumptions 5, 6 and 7 hold. Let* $\beta = \beta_r > 0$ *be as given in Proposition 5. Let* $f : \mathbf{X} \to \mathbb{R}$ *be a measurable function such that* $|||f|||_\beta < \infty$*. Let* $P = P_\theta$ *for some fixed* θ*, and let* μ^* *denote the unique invariant probability measure of* P*, and let* $h = \mu^*(f)$*. Then, the Poisson equation*

$$(I - P^*)u(x) = f(x) - h \tag{103}$$

has a unique solution $u(\cdot)$ up to additive constants. The particular solution for which $\mu^*(u) = 0$ can be written as

$$u(x) = \sum_{n=0}^{\infty} (P^{*n} f(x) - h), \tag{104}$$

where the right-hand side is absolutely convergent, and

$$|u(x)| \leq |||f|||_\beta K_{r,u}(1 + \beta V(x)) \tag{105}$$

for some constant $K_{r,u} > 0$ depending only on the constants appearing in Assumptions 5, 6 and 7. It also follows:

$$\|u\|_\beta \leq K_{r,u} |||f|||_\beta < \infty. \tag{106}$$

Proof Consider the Poisson equation

$$(I - P^{*r})v(x) = f(x) - h, \tag{107}$$

where $h = \mu^*(f)$, recalling that $\mu_r^* = \mu^*$. In view of Theorem 1, it has a unique solution, up to an additive constant. The particular solution with $\mu^*(v) = 0$ can be written as

$$v(x) = \sum_{n=0}^{\infty} (P^{*nr} f(x) - h), \tag{108}$$

which is well defined, the r.h.s. is absolutely convergent, and

$$|v(x)| \leq |||f|||_\beta K_{r,v}(1 + \beta V(x)), \tag{109}$$

implying the inequality

$$\|v\|_\beta \leq K_{r,v} |||f|||_\beta , \tag{110}$$

where $K_{r,v}$ is given by

$$K_{r,v} := \frac{1}{1-\alpha_r} \max\left(2 + \beta \frac{K_r}{1-\gamma_r}, \beta\right). \tag{111}$$

The solution of (107) is related to that of (103) by noting that

$$(I - P^{*r}) = (I - P^*)(I + P^* + \ldots + P^{*(r-1)}). \tag{112}$$

It follows that

$$u(x) := (I + P^* + \ldots + P^{*(r-1)})v(x) \tag{113}$$

is a solution of (103).

To get an upper bound for $|u(x)|$, write

$$u(x) = (I + P + \ldots + P^{r-1})\, \delta_x(v). \tag{114}$$

Taking into account the upper bound for $|v(x)|$ given in (109), it is seen that it is sufficient to derive upper bounds for $(P^k \delta_x)(V) = P^{*k} V(x)$ for $k = 1, \ldots, r-1$.

Now, in view of the one-step growth condition and inequality (89), we have $P\mu(V) \leq \gamma_1 \mu(V) + K_1$, for any probability measure μ with $\mu(V) < \infty$. By repeated application of this inequality, we obtain the upper bound for $(P^k \mu)(V)$:

$$\gamma_1^k \mu(V) + \sum_{\ell=0}^{k-1} \gamma_1^\ell K_1 \leq \gamma_1^k \mu(V) + \frac{\gamma_1^k K_1}{\gamma_1 - 1}. \tag{115}$$

Using $\mu = \delta_x$ and summing over k from 0 to $r-1$, we get

$$(I + P + \ldots + P^{r-1})\delta_x(V) \leq \sum_{k=0}^{r-1} \gamma_1^k \left(V(x) + \frac{K_1}{\gamma_1 - 1} \right).$$

The right-hand side is bounded from above by

$$\frac{\gamma_1^r}{\gamma_1 - 1} \left(V(x) + \frac{K_1}{\gamma_1 - 1} \right). \tag{116}$$

Combining these inequalities with (109) and (113), we get

$$\begin{aligned} |u(x)| =& |(I + P + \ldots + P^{r-1})\delta_x(v)| \leq \|\!|f|\!\|_\beta \, K_{r,v} \times \\ &\times \left(r + \beta \frac{\gamma_1^r}{\gamma_1 - 1} \left(V(x) + \frac{K_1}{\gamma_1 - 1} \right) \right), \end{aligned} \tag{117}$$

implying the upper bound of the form given in (105).

As for *uniqueness*, assume that there are two solutions $u_1, u_2 \in \mathcal{L}_V$, and let $\Delta u = u_2 - u_1$. Then, $(I - P^*)\Delta u(x) = 0$ for all x, and hence, $P^* \Delta u = \Delta u$. Iterating this $r-1$ times, we get $P^{*r} \Delta u = \Delta u$, and thus, by Theorem 1, Δu is a constant function, completing the proof. □

The extension of Theorem 2, on the Lipschitz continuity of $u_\theta(\cdot)$, seems to be straightforward. However, we should point out that we have to assume the Lipschitz continuity of the *one-step* kernels (P_θ), as given in Assumption 3.

Theorem 4 *Assume that the kernels (P_θ) satisfy Assumptions 5, 6 and 7. In addition, assume that the family of one-step kernels (P_θ) is Lipschitz continuous is the sense of Assumption 3. Let us fix $\beta = \beta_r > 0$ as given in Proposition 5. Let (f_θ) be a family of $\mathbf{X} \to \mathbb{R}$ measurable functions such that Assumption 4 holds with the above β. Let μ_θ^* denote the unique invariant probability measure of P_θ, and let $h_\theta = \mu_\theta^*(f_\theta)$. Consider the Poisson equations*

$$(I - P_\theta^*) u_\theta(x) = f_\theta(x) - h_\theta. \tag{118}$$

Then, h_θ is Lipschitz continuous in θ:

$$|h_\theta - h_{\theta'}| \le L_{r,h}|\theta - \theta'|, \tag{119}$$

and the particular solution, given in Theorem 3 by (104) as $u_\theta(x) = \sum_{n=0}^{\infty}(P_\theta^{*n} f_\theta(x) - h_\theta)$, is Lipschitz continuous in θ:

$$|u_\theta(x) - u_{\theta'}(x)| \le L_{r,u}|\theta - \theta'|(1 + \beta V(x)) \tag{120}$$

where $L_{r,u}$ is independent of x. Alternatively, we can write

$$\|u_\theta - u_{\theta'}\|_\beta \le L_{r,u}|\theta - \theta'|. \tag{121}$$

Here, the constants $L_{r,h}$ and $L_{r,u}$ depend only on the constants appearing in Assumptions 3, 4, 5, 6, and 7.

For the proof, we need a simple variant of Lemma 3 providing an at most exponentially growing bound for $\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n)$:

Lemma 7 *Let (P_θ) satisfy the uniform one-step growth condition, Assumption 6. In addition, let Assumption 3, requiring the Lipschitz continuity of (P_θ), hold with some $\beta > 0$. Then for any signed measure η with $|\eta|(1 + \beta V) < \infty$ and $\theta, \theta' \in \Theta$, and for any $\alpha'' > \alpha' := \max(1 + \beta K_1, \gamma_1)$*

$$\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta) \le L''_P|\theta - \theta'|(\alpha'')^n |\eta| (1 + \beta V) \tag{122}$$

for all $n > 0$, where L''_P depends only on the constants appearing in the conditions of the lemma and on α''.

The proof follows the proof of Lemma 3; however, to estimate $|P_{\theta'}^k \eta|(V)$, we need a modification of (115), with $|\eta|$ replacing μ, yielding a restatement of (A21):

$$|P_{\theta'}^k \eta|(V) \le \gamma_1^k \left(|\eta|(V) + K_1|\eta|(\mathbf{X})/(\gamma_1 - 1)\right). \tag{123}$$

Details will be given in the Appendix.

Proof of Theorem 4 First, note that $\mu_\theta^* = \mu_{\theta,r}^*$ implies that

$$h_\theta = \mu_\theta^*(f_\theta) = \mu_{\theta,r}^*(f_\theta). \tag{124}$$

Applying Theorem 2 for the Poisson equation

$$(I - P_\theta^{*r})v_\theta(x) = f_\theta(x) - h_\theta, \tag{125}$$

noting that P_θ^{*r} satisfies the relevant conditions in view of Lemma 3, we conclude that h_θ is Lipschitz continuous in θ:

$$|h_\theta - h_{\theta'}| \le L_{r,h}|\theta - \theta'|, \tag{126}$$

where $L_{r,h}$ is given, according to (71), by

$$L_{r,h} = \left(L_f + K_f L_{P^r} \frac{1}{1-\alpha_r}\right)\left(1 + \beta_r \frac{K_r}{1-\gamma_r}\right), \tag{127}$$

where L_{P^r} is the Lipschitz constant for (P_θ^r), ensured by Lemma 3, and β_r and α_r are chosen as in Proposition 5.

In order to prove the second part of Theorem 4, note that, in view of Theorem 2, the particular solution given by

$$v_\theta(x) = \sum_{n=0}^{\infty} P_\theta^{*nr}(f_\theta(x) - h_\theta)$$

is Lipschitz continuous w.r.t. θ, and

$$|v_\theta(x) - v_{\theta'}(x)| \leq L_{r,v}|\theta - \theta'|(1 + \beta V(x)), \tag{128}$$

where $L_{r,v}$ is defined in (86) in the role of L_u, according to

$$L_{r,v} = \frac{1}{1-\alpha_r}\left(L_f + L_{P^r} \frac{2}{(1-\alpha_r)} K_f\right)\left(2 + \beta_r \frac{K_r}{1-\gamma_r}\right).$$

Hence, the particular solution of the Poisson equation (118):

$$u_\theta(x) := (I + P_\theta^* + \ldots + P_\theta^{*(r-1)})v_\theta(x) \tag{129}$$

is also Lipschitz continuous in θ. Indeed, for $1 \leq m \leq r-1$

$$\begin{aligned} &P_\theta^{*m} v_\theta - P_{\theta'}^{*m} v_{\theta'} \\ &= P_\theta^{*m}(v_\theta - v_{\theta'}) + (P_\theta^{*m} - P_{\theta'}^{*m})v_{\theta'}, \end{aligned} \tag{130}$$

and the first term on the r.h.s. is bounded from above as

$$|P_\theta^{*m}(v_\theta - v_{\theta'})(x)| \leq L_{r,v}|\theta - \theta'|P_\theta^{*m}(1 + \beta V)(x),$$

for all x. Applying (123) with $\eta = \delta_x$, we get the upper bound

$$|P_\theta^{*m}(v_\theta - v_{\theta'})| \leq L_{r,v}|\theta - \theta'|\gamma_1^m \left(V(x) + \frac{K_1}{\gamma_1 - 1}\right). \tag{131}$$

For the second term on the right-hand side of (130), first we note that $|||v_{\theta'}|||_\beta < \infty$ by (110) in the proof of Theorem 3:

$$|||v_{\theta'}|||_\beta \leq \|v_{\theta'}\|_\beta \leq K_{r,v}|||f_{\theta'}|||_\beta \leq K_{r,v}K_f \tag{132}$$

for some constant $K_{r,v} > 0$ depending only on the constants appearing in Assumptions 5, 6 and 7. Now, we can write

$$\begin{aligned}|(P_\theta^{*m} - P_{\theta'}^{*m})v_{\theta'}(x)| &= |(P_\theta^m - P_{\theta'}^m)\delta_x(v_{\theta'})| \\ &\leq \sigma_\beta\big((P_\theta^m - P_{\theta'}^m)\delta_x\big)\,|||v_{\theta'}|||_\beta \\ &= \sigma_\beta(P_\theta^m\delta_x, P_{\theta'}^m\delta_x)\,|||v_{\theta'}|||_\beta. \end{aligned} \tag{133}$$

Applying Lemma 7, we get that the r.h.s. is bounded by

$$L_P''|\theta - \theta'|(\alpha'')^m\,(1 + \beta V(x)) \cdot K_{r,v}|||f_{\theta'}|||_\beta \tag{134}$$

where $K_{r,v}$ is defined by (111). Recall that $\sup_{\theta'} |||f_{\theta'}|||_\beta = K_f < \infty$ by Assumption 4. Taking into account the representation of $u_\theta(x)$ given in (129), and the decomposition given in (130), and adding the upper bounds (131) and (134) for $m = 0, \ldots, r-1$, we get the second claim. □

8 Design of queues

In this section, the applicability of our results will be demonstrated on the modification of a classical textbook example, a simple queuing system. For the sake of simplicity, we restrict attention to a single arrival process handled by a single server, where both the arrival and service process may be subject to control. For now, we consider a system with open-loop control, a multivariate version of which may be a realistic model for a client assignment systems. Extension of our approach to queuing systems allowing feedback, such as Call Admission Control, see, e.g., Chapter 11 of [31], may be the subject of future research.

Let us first describe the dynamics of a queue without control. Let the arrival process be identified by a simple point process $\tau_n,\ n \geq 0$ with $\tau_0 = 0$, and let T_{n+1} be the (finite) time elapsed between the arrivals of customers n and $n+1$, i.e., $T_{n+1} = \tau_{n+1} - \tau_n$. Let S_n be the service time of customer n. It is assumed that the sequences (T_n) and (S_n), $n \in \mathbb{N}$, are i.i.d. sequences of $\mathbb{R}_+$-valued random variables, respectively, independent of each other. We define $x^+ := \max\{x, 0\}$, $x \in \mathbb{R}$. The waiting time of the n-th customer will be denoted by W_n. It is readily seen that it satisfies the recursion, with $W_0 = 0$ as initial value, $n \in \mathbb{N}$ and with $U_{n+1} = S_n - T_{n+1}$:

$$W_{n+1} = (W_n + U_{n+1})^+. \tag{135}$$

In the case of controlled queues, both the service time and the arrival time may depend on a control parameter θ. Let $\Theta \subset \mathbb{R}^k$ be a connected, open set as above, and let $D \subset \mathbb{R}^k$ be a compact set such that $D \subset \Theta$. The choice of the control parameter θ may determine the law of the service times and that of the arrival times. Compactness of D is a technical condition needed for the verification of Assumption 5. The dynamics of the queue is described, with $W_{\theta,0} = 0$, and $U_{\theta,n+1} := S_{\theta,n} - T_{\theta,n+1}$, by

$$W_{\theta,n+1} = (W_{\theta,n} + U_{\theta,n+1})^+. \tag{136}$$

If the initial condition is $W_{\theta,0} = x$, then the waiting time at n will be denoted by $W_{\theta,n}(x)$. To guarantee stability of the queue, we have to assume

$$\mathbb{E}[\, U_{\theta,1} \,] < 0, \tag{137}$$

for all $\theta \in D$. This is a standard condition implying stability of the queue for any fixed θ in a variety of interpretations, see, e.g., [32, 33] and [34].

The validity of the drift condition, given as Assumption 1, with no parameter dependence, has been established under appropriate technical conditions, using the Lyapunov function $V(x) := e^{\chi x}$, $x \in \mathbb{R}_+$, with $\chi > 0$ small enough, in Section 16.4 of [20]. A uniform version of this result will be established below. In fact, we will show that under reasonable additional conditions on a controlled queue Assumptions 1, 3, 7 are satisfied with $V(x) := e^{\chi x}$, $x \in \mathbb{R}_+$, with $\chi > 0$ small enough, when θ is restricted to compact set $D \subset \Theta$. Thus, the results of the paper, in particular Theorems 7, 3, 4, imply the existence, uniqueness, and Lipschitz continuity of the solution $u_\theta(x)$ of the parameter-dependent Poisson equation

$$u_\theta(x) - \mathbb{E}_\theta[\, W_{\theta,1} \mid W_0 = x \,] \;=\; f_\theta(x) - h_\theta, \tag{138}$$

where the normalizing constant h_θ is the expectation of $f_\theta(x)$ under the (unique) invariant measure, when θ is restricted to an open set $\Theta' \subset D$ in place of Θ.

The conditions below will be given in terms of the r.v. $U_{\theta,1}$, thus ensuring the generality of our results. Specific conditions in terms of $S_{\theta,0}$ and $T_{\theta,1}$ will be given at the end of the section. To guarantee stability of the system (136), we stipulate:

Assumption 8 We have $\mathbb{E}[\, U_{\theta,1} \,] < 0$, for all $\theta \in D$.

This is a standard condition implying stability of the queue for any fixed θ in a variety of interpretations, see, e.g., [32, 33] and [34]. A further standard condition in queuing theory, and also in the area of risk processes [35], is the existence of a finite positive exponential moment of $S_{\theta,n} - T_{\theta,n+1}$, or equivalently that of $U_{\theta,n}$. A uniform version of this condition in terms of $U_{\theta,1}$ is given below:

Assumption 9 We have $\sup_{\theta\in D} \mathbb{E}[\, \exp(\eta\, U_{\theta,1}) \,] < \infty$, for some $\eta > 0$.

Observe that Assumption 9 is automatically satisfied if $\sup_{\theta\in D} \mathbb{E}[\, \exp(\eta\, S_{\theta,0}) \,] < \infty$. Finally, we will need the following continuity condition for $U_{\theta,1}$:

Assumption 10 The probability distribution of $U_{\theta,1}$ is weakly continuous in θ for $\theta \in D$, i.e., $\mathbb{E}[\, f(U_{\theta,1}) \,]$ is continuous in θ for all bounded continuous functions f.

We note in passing that these three assumptions imply that the stability condition, Assumption 8, is satisfied uniformly in θ for $\theta \in D$:

$$\sup_{\theta\in D} \mathbb{E}[\, U_{\theta,1} \,] < 0. \tag{139}$$

Uniform Drift Condition. The validity of the uniform drift condition, given as Assumption 1, with no parameter dependence, has been established using the Lyapunov function $V(x) := e^{\chi x}$, $x \in \mathbb{R}_+$, with χ small enough, see, e.g., Section 16.4

of [20]. (We should note that the use of exponential moments is also a standard tool in the theory of risk processes, see [35].) For the sake of completeness and further reference, we restate this result and provide its proof in the Appendix.

Lemma 8 *Let us assume that U is an $\mathbb{R}$-valued random variable such that $\mathbb{E}\, U < 0$, and for some $\eta > 0$ we have $\mathbb{E}\,[\, e^{\eta U}\,] < \infty$. Then, there exist $0 < \chi_0 < \eta$ such that for all $0 < \chi \le \chi_0$ there exists $0 < \gamma < 1$, such that*

$$\mathbb{E}\,[\, e^{\chi (x+U)^+}\,] \le \gamma e^{\chi x} + 1. \tag{140}$$

An extension of Lemma 8, ensuring the uniform validity of the drift condition for exponential Lyapunov functions, is stated below.

Lemma 9 *Let $U(\theta) := U_{\theta,1},\ \theta \in D$, be a family of $\mathbb{R}$-valued random variables, satisfying Assumptions 8, 9 and 10. Then, there exist $0 < \chi_0 < \eta$ such that for all $0 < \chi \le \chi_0$ there exists $0 < \gamma < 1$, such that for all $\theta \in D$*

$$\mathbb{E}\,[\, e^{\chi (x+U(\theta))^+}\,] \le \gamma e^{\chi x} + K. \tag{141}$$

Remark 6 Taking into account the proof of Lemma 8, it is clear that all we need to prove Lemma 9 is that a uniform version of (A25) is valid, i.e., that there exists a $0 < \chi_0 < \eta$ and some $\varepsilon > 0$ such that for all $0 < \chi \le \chi_0$ and for all $\theta \in D$

$$\frac{\mathbb{E}\,[\, e^{\chi U(\theta))}\,] - 1}{\chi} \le -\varepsilon < 0. \tag{142}$$

Let us define the family of functions $g(\chi, \theta) := \mathbb{E}\,[\, e^{\chi U(\theta)}\,]$. By Assumption 9, it is readily seen that the random variables $e^{\chi U(\theta)}$ are uniformly integrable for $\chi < \eta$, and, therefore, by Assumption 10 it follows that $g(\chi, \theta)$ is continuous in θ. The desired claim (142) now follows from the following lemma (proved in the Appendix) formulated in the context of convex analysis:

Lemma 10 *Let $g(\chi, \theta)$ be a family of convex functions in the variable χ with $0 \le \chi < \eta$ and $\theta \in D \subset \mathbb{R}^k$ with D being a compact set, such that*

- *$g(0, \theta) = 1$ for all $\theta \in D$,*
- *for all fixed $\theta \in D$ we have*

$$\inf_{0<\chi<\eta} \frac{g(\chi, \theta) - 1}{\chi} < 0, \tag{143}$$

- *for all fixed $0 \le \chi < \eta$ the function $g(\chi, .)$ is continuous in θ.*

Then, there exists $\chi_0 > 0$ such that for $0 < \chi \le \chi_0$ we have

$$\sup_{\theta\in D} \frac{g(\chi, \theta) - 1}{\chi} < 0. \tag{144}$$

It follows that $\sup_{\theta\in D} g(\chi, \theta) < 1$ for $0 < \chi \le \chi_0$.

Remark 7 It also readily follows that

$$\sup_{\theta \in D} \inf_{0<\chi<\eta} \frac{g(\chi,\theta)-1}{\chi} = \inf_{0<\chi<\eta} \sup_{\theta \in D} \frac{g(\chi,\theta)-1}{\chi} < 0.$$

Local Minorization for $P_\theta^r(x, .)$. As for the local minorization condition it will be verified in a form slightly stronger than our Assumption 7, by showing that for any fixed $R > 0$ there exists some integer $r \geq 1$, which may depend on R, such that

$$\inf_{\theta \in D} \inf_{0 \leq x \leq R} P_\theta^r(x, \{0\}) > 0. \tag{145}$$

Indeed, the above inequality implies Assumption 7. To see this, take an arbitrary R_r, as in Assumption 7, satisfying $R_r > 2K_r/(1-\gamma_r)$. Then, the set

$$\mathcal{C}_r = \{x \in \mathbf{X} : V(x) \leq R_r\} = \{x \in \mathbf{X} : e^{\chi x} \leq R_r\} \tag{146}$$

is of the form $\{0 \leq x \leq R\}$ with some R. Letting $\bar{\mu}_r$ denote the probability measure assigning unit mass to 0, inequality (145) implies $P_\theta^r(x, A) \geq \bar{\alpha}_r \bar{\mu}_r(A)$ with some $\bar{\alpha}_r \in (0, 1)$ for all $\theta \in \Theta$ and $x \in \mathcal{C}_r$, as postulated by (93).

To prove (145), we will use arguments familiar in queuing theory, but to establish uniform bounds extra care is needed. Consider first a fixed $\theta \in D$,, and let $R > 0$ be any fixed real number, defining the bounded $[0, R]$. Let $0 \leq x \leq R$, and consider the r-step transition probability $P_\theta^r(x, \{0\}) = P(W_{\theta,r}(x) = 0)$ for some integer $r \geq 1$.

Lemma 11 *There is $\epsilon > 0$ such that*

$$\upsilon := \inf_{\theta \in D} P(U_{\theta,1} < -\epsilon) > 0. \tag{147}$$

Corollary 3 *For each $R > 0$, there is $r \in \mathbb{N}$ such that*

$$\inf_{\theta \in D} \inf_{0 \leq x \leq R} P(W_{\theta,r}(x) = 0) > 0. \tag{148}$$

A nice additional result is the following: Let $W_{\theta,s}^*$ denote the stationary solution of the queue dynamics given by

$$W_{\theta,s}^* := \max_{-\infty \leq k \leq s} (U_{\theta,k} + \ldots + U_{\theta,s})^+. \tag{149}$$

Then,

$$\inf_{\theta \in D} \inf_{0 \leq x \leq R} P(W_{\theta,r}(x) = 0) \geq C_r \inf_{\theta \in D} P(W_{\theta,r}^* = 0), \tag{150}$$

where $C_r \to 1$ exponentially fast as $r \to \infty$. Moreover,

$$\inf_{\theta \in D} P(W_{\theta,r}^* = 0) > 0. \tag{151}$$

Lipschitz Continuity of P_θ. In order to verify Assumption 3, we will need to strengthen our assumptions on $S_{\theta,0}$ and $T_{\theta,1}$. We may consider various scenarios, briefly discussed below, both of them implying a common condition on $U_{\theta,1}$ as follows:

Assumption 11 The probability distribution function of $U_{\theta,1}$ has a density function for all $\theta \in D$, denoted by $\zeta_\theta(\cdot)$ There exist $\eta'' > 0$ and $C'' > 0$ such that for all $\theta,\ \theta' \in \Theta$ and all $x \in \mathbb{R}$, it holds that

$$|\zeta_\theta(x) - \zeta_{\theta'}(x)| \le C''\, e^{-\eta''|x|}|\,\theta - \theta'|. \tag{152}$$

Assumption 11 can be conveniently verified by imposing the following assumption on $S_{\theta,0}$ and $T_{\theta,1} := T_1$, when $T_{\theta,1}$ is assumed to be independent of θ : The probability distribution functions of $S_{\theta,0}$ have a density function for all $\theta \in D$, denoted by $\xi_\theta(\cdot)$, and there exist $C', \eta' > 0$ such that for all $\theta, \theta' \in \Theta$ and all $x \in \mathbb{R}_+$, it holds that $|\xi_\theta(x) - \xi_{\theta'}(x)| \le C'e^{-\eta' x}|\theta - \theta'|$. Moreover, the probability distribution function of T_1 has a density function denoted by $\kappa(\cdot)$ such that for all $x \in \mathbb{R}_+$ it holds that $\kappa(x) \le C'e^{-\eta' x}$.

The first part of the above auxiliary assumption can be conveniently checked by requiring the existence of a density function $\xi_\theta(\cdot)$ such that the mapping $(\theta, x) \to \xi_\theta(x) \in \mathbb{R}_+$ is measurable, and for each fixed x continuously differentiable in $\theta \in \Theta$, moreover there exist $C', \eta' > 0$ such that for all $\theta \in \Theta$ and all $x \in \mathbb{R}_+$ it holds that $\|\frac{\partial}{\partial\theta}\xi_\theta(x)\| \le C'e^{-\eta' x}$. The proof is readily obtained by the mean-value theorem. Incidentally, it also follows that the law of $S_{\theta,0}$ is continuous in total variation, and, a fortiori, also weakly, implying Assumption 10.

A condition reciprocal to the above is obtained by interchanging the role of S_0 and T_1, i.e., assuming that S_0 does not depend on θ, while $T_1 = T_{\theta,1}$ does. We note that in this case the requirement that the density function of S_0, denoted by $\xi(\cdot)$, satisfies $\xi(x) \le C'e^{-\eta' x}$ for all $x \in \mathbb{R}_+$ implies Assumption 9 with any $\eta < \eta'$.

In order to verify Assumption 3, let us consider the transition probabilities $P_\theta(x, A)$. For any Borel set $A \subset \mathbb{R}_+,\ 0 \notin A$, we can write

$$\begin{aligned} P_\theta(x, A) &= P(W_{\theta,1}(x) \in A) = \mathbb{E}\,[\mathbf{1}_A(x + U_{\theta,1})] \\ &= \int_{-\infty}^{\infty} \mathbf{1}_A(x+z)\,\zeta_\theta(z)\,\mathrm{d}z. \end{aligned} \tag{153}$$

State 0, being an atom for $P_\theta(x, .)$, is reached with probability

$$\begin{aligned} P_\theta(x, \{0\}) = P(U_{\theta,1} \le -x) = \int_{-\infty}^{-x} \zeta_\theta(z)\,\mathrm{d}z = \\ \int_{-\infty}^{\infty} \mathbf{1}_{(-\infty,0]}\,(x+z)\,\zeta_\theta(z)\,\mathrm{d}z. \end{aligned} \tag{154}$$

Lemma 12 *Let Assumption 11 hold, and let* $V(x) := e^{\chi x},\ x \in \mathbb{R}_+$, *with* $\chi < \eta''$. *Then, for all* $\phi \in \mathcal{L}_V$ *with* $||\phi||_\beta \le 1$ *and all* $\theta, \theta' \in \Theta$ *we have with some* $L > 0$

$$\int_{\mathbb{R}_+} \phi(y)(P_\theta(x, \mathrm{d}y) - P_{\theta'}(x, \mathrm{d}y)) \le L(1 + \beta V(x))|\theta - \theta'|.$$

9 Discussion

We have revisited a key technical issue in the theory of stochastic approximation in a Markovian framework, developed in [4]: the Lipschitz continuity of the solution of the associated Poisson equation w.r.t. the parameter θ, characterizing the system dynamics. A set of simple conditions have been formulated under which the desired Lipschitz continuity can be established, significantly simplifying the relevant conditions and results of [4]. We demonstrated the utility of a powerful, off-beat result on the stability of Markov chains in [19], proving that the transition kernels are contractions in the space of differences of probability measures in a suitable metric.

The uniform drift condition, Assumption 1, and its relaxation, restating it for some power of the kernel, $P_\theta^r(x, A)$, see Assumption 5, is akin to condition (A'.5)(i), p. 290 of [4], but allowing more general Lyapunov functions $V(\cdot)$. As for the condition on the Lipschitz continuity of the kernel, our Assumption 3 is in most aspects significantly less restrictive than the corresponding conditions of Theorem 6, p. 262 of [4]. Noting some overlaps between the proofs of the latter result of [4] and that of our Theorem 2, a question for future research is whether our Assumption 3 can be relaxed by using a smaller class of test functions. To complete the loop, the technology presented here will be applied for the ODE analysis of recursive estimators along the lines of [4].

The viability of our results has been demonstrated on the modification of a textbook example of a queuing system, allowing open-loop control. The extension of our analysis to complex networks, with several servers and/or customers allowing feedback control is an attractive and challenging problem.

The setup of our paper is suitable for the analysis of certain cyber-physical systems, incorporating systems described by stochastic partial differential equations (SPDEs). A prototype for Markov processes arising in such context is given in the lecture notes [27]. Finally, recent intense interest in online machine learning will definitely inspire further applications, especially in various stochastic gradient methods and reinforcement learning where the Markovian setup is the standard choice.

Appendix A Proofs

Proof of Corollary 1 Indeed, since $\{\varphi : \|\varphi\|_\beta \leq 1\} \subseteq \{\varphi : |||\varphi|||_\beta \leq 1\}$, we get $\rho_\beta(\mu_1, \mu_2) \leq \sigma_\beta(\mu_1, \mu_2)$. On the other hand, take $\varepsilon > 0$ and let φ be such that $|||\varphi|||_\beta \leq 1$ and

$$\int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x) \geq \sigma_\beta(\mu_1, \mu_2) - \varepsilon. \tag{A1}$$

By Definition 3, there is a constant c such that $|||\varphi|||_\beta = \|\varphi + c\|_\beta$. Thus, $\|\varphi + c\|_\beta \leq 1$, and hence,

$$\begin{aligned} \rho_\beta(\mu_1, \mu_2) &\geq \int_{\mathbf{X}} (\varphi(x) + c)(\mu_1 - \mu_2)(\mathrm{d}x) \\ &= \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x) \\ &\geq (\sigma_\beta(\mu_1, \mu_2) - \varepsilon). \end{aligned} \tag{A2}$$

Since ε is arbitrary, we get that $\rho_\beta(\mu_1, \mu_2) \geq \sigma_\beta(\mu_1, \mu_2)$. Combining with the opposite inequality, we get the claim. □

Proof of Lemma 1 For the integral of the l.h.s. of (48), we apply Fubini's theorem to get

$$\int_{\mathbf{X}} \left(\int_{\mathbf{X}} \varphi(y) P_\theta(x, \mathrm{d}y) \right) \mu(\mathrm{d}x) = \int_{\mathbf{X}} \varphi(y) \eta(\mathrm{d}y), \tag{A3}$$

where the measure $\eta = P_\theta \mu$ is defined as usual by $\eta(A) = \int_{\mathbf{X}} P_\theta(x, A)\mu(\mathrm{d}x)$. The measure η is finite, since $\mu(\mathbf{X}) < \infty$. The application of Fubini's theorem is justified since

$$\begin{aligned} \iint |\varphi(y)| P_\theta(x, \mathrm{d}y)\mu(\mathrm{d}x) &\leq \iint (1 + \beta V(y)) P_\theta(x, \mathrm{d}y)\mu(\mathrm{d}x) \\ &\leq \int (1 + \beta(\gamma V(x) + K))\mu(\mathrm{d}x), \end{aligned}$$

and the right-hand side is finite. Using the same argument for θ', altogether we obtain for the integral of (48)

$$\int_{\mathbf{X}} \varphi(y) \left(P_\theta \mu(\mathrm{d}y) - P_{\theta'} \mu(\mathrm{d}y) \right) \leq L_P |\theta - \theta'| \mu(1 + \beta V).$$

Since φ is arbitrary subject to $\|\varphi\|_\beta \leq 1$, we conclude that $\rho_\beta(P_\theta \mu, P_{\theta'} \mu) = \sigma_\beta(P_\theta \mu, P_{\theta'} \mu)$ is bounded by the right-hand side of (48), and we get the statement of the Lemma. □

Proof of Lemma 2 Consider the Hahn–Jordan decomposition $\eta = \eta^+ - \eta^-$, as recalled after the lemma itself. Then,

$$\sigma_\beta((P_\theta - P_{\theta'})\eta) \leq \sigma_\beta((P_\theta - P_{\theta'})\eta^+) + \sigma_\beta((P_\theta - P_{\theta'})\eta^-).$$

Using Lemma 1 for both terms, we get the upper bound:

$$L_P |\theta - \theta'| \eta^+(1 + \beta V) + L_P |\theta - \theta'| \eta^-(1 + \beta V). \tag{A4}$$

Noting that $\eta^+ + \eta^- = |\eta|$, the lemma follows.

□

Proof of Lemma 3 We can estimate $\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta)$ from above, using a kind of telescopic sequence of triangular inequalities, leading to the upper bound

$$\begin{aligned} &\sum_{k=0}^{n-1} \sigma_\beta(P_\theta^{n-k} P_{\theta'}^k \eta, P_\theta^{n-k-1} P_{\theta'}^{k+1} \eta) \\ &= \sum_{k=0}^{n-1} \sigma_\beta(P_\theta^{n-k-1} P_\theta P_{\theta'}^k \eta, P_\theta^{n-k-1} P_{\theta'} P_{\theta'}^k \eta). \end{aligned} \tag{A5}$$

Note that the measures $P_\theta P^k_{\theta'}\eta$ and $P_{\theta'} P^k_{\theta'}\eta$ satisfy $P_\theta P^k_{\theta'}\eta(\mathbf{X}) = P_{\theta'} P^k_{\theta'}\eta(\mathbf{X})$, and hence, their $\sigma_\beta(.,.)$ distance is well defined, see Definition 4.

Using the contraction property of the kernels P_θ^{n-k-1}, see Proposition 3, we obtain the upper bound

$$\sum_{k=0}^{n-1} \alpha^{n-k-1} \sigma_\beta(P_\theta P^k_{\theta'}\eta, P_{\theta'} P^k_{\theta'}\eta). \tag{A6}$$

For the k-th term, we estimate $\sigma_\beta(P_\theta P^k_{\theta'}\eta - P_{\theta'} P^k_{\theta'}\eta)$ from above applying Lemma 2 with $P^k_{\theta'}\eta$ taking the role of η to get the following upper bound for (A6):

$$L_P|\theta - \theta'| \sum_{k=0}^{n-1} \alpha^{n-k-1} |P^k_{\theta'}\eta|(1 + \beta V). \tag{A7}$$

Note that by the consequence of the drift condition given in inequality (12) we can bound $|P^k_\theta \eta|(V)$ for a general θ by

$$|P^k_\theta \eta|(V) \leq \gamma |P^{k-1}_\theta \eta|(V) + K|P^{k-1}_\theta \eta|(\mathbf{X}). \tag{A8}$$

Noting that $|P^{k-1}_\theta \eta|(\mathbf{X}) \leq |\eta|(\mathbf{X})$, and iterating the above inequality, we get

$$\begin{aligned} |P^k_\theta \eta|(V) &\leq \gamma^2 |P^{k-2}_\theta \eta|(V) + \gamma K|\eta|(\mathbf{X}) + K|\eta|(\mathbf{X}) \\ &\dots \\ &\leq \gamma^k |\eta|(V) + \sum_{\ell=0}^{k-1} \gamma^\ell K|\eta|(\mathbf{X}) \\ &\leq \gamma^k |\eta|(V) + \frac{K}{1-\gamma}|\eta|(\mathbf{X}). \end{aligned} \tag{A9}$$

By plugging (A9) into the sum in (A7), we get the upper bound

$$\sum_{k=0}^{n-1} \alpha^{n-k-1} \left(|\eta|(\mathbf{X}) + \beta \left(\gamma^k |\eta|(V) + \frac{K}{1-\gamma}|\eta|(\mathbf{X}) \right) \right).$$

We can write the latter expression as

$$\beta\alpha^{n-1} \sum_{k=0}^{n-1} \left(\frac{\gamma}{\alpha}\right)^k |\eta|(V) + \left(1 + \beta\frac{K}{1-\gamma}\right) \sum_{k=0}^{n-1} \alpha^k |\eta|(\mathbf{X}). \tag{A10}$$

Summarizing the inequalities (A5) to (A10), taking into account $\alpha > \gamma$ (see Remark 3), and bounding the geometric sums in (A10) with their limit values, we get the upper bound

$$\frac{1}{1-\alpha}\left(1 + \beta\frac{K}{1-\gamma}\right) \vee \frac{\alpha^n}{\alpha - \gamma}, \tag{A11}$$

from which the claim follows by setting $n = 0$. □

Proof of Corollary 2 Note that for any initial probability measure $\mu \in \mathcal{M}_V$, we have by the triangle inequality

$$\sigma_\beta(\mu^*_\theta, \mu^*_{\theta'}) \leq \sigma_\beta(\mu^*_\theta, P^n_\theta \mu) + \sigma_\beta(P^n_\theta \mu, P^n_{\theta'} \mu) + \sigma_\beta(P^n_{\theta'} \mu, \mu^*_{\theta'}).$$

Letting $n \to \infty$, the first and the last terms on the r.h.s. converge to zero by Proposition 3. Taking $\mu = \delta_x$, the middle term is upper bounded, for any n, in view of Lemma 3 by

$$L_P |\theta - \theta'| \cdot C_P(1 + \beta V(x)). \tag{A12}$$

Note that C_P can be replaced by what is given in (A11). Recalling that $\inf_x V(x) = 0$ by Remark 2, and letting $n \to \infty$, the claim follows with the constant C'_P stated in the corollary. □

Proof of Lemma 4 The starting point of the proof is the inequality, obtained by combining (A5)–(A6), applicable also for signed measures such that $|\eta|(1 + \beta V) < \infty$:

$$\sigma_\beta(P^n_\theta \eta, P^n_{\theta'} \eta) \leq \sum_{k=0}^{n-1} \alpha^{n-k-1} \sigma_\beta(P_\theta P^k_{\theta'} \eta, P^{k+1}_{\theta'} \eta). \tag{A13}$$

A key point is the observation that since $\eta(\mathbf{X}) = 0$, $P^k_{\theta'} \eta$ converges exponentially fast to the zero measure, see Proposition 3. To estimate the kth term of (A13), we apply Lemma 2 and Proposition 1, (23),

$$\begin{aligned} \sigma_\beta((P_\theta - P_{\theta'}) P^k_{\theta'} \eta) &\leq L_P |\theta - \theta'| |P^k_{\theta'} \eta|(1 + \beta V) \\ &= L_P |\theta - \theta'| \sigma_\beta(P^k_{\theta'} \eta). \end{aligned} \tag{A14}$$

Now applying Proposition 3 and Proposition 1, (23), again, we get the upper bound:

$$L_P |\theta - \theta'| \alpha^k \sigma_\beta(\eta) = L_P |\theta - \theta'| \alpha^k |\eta|(1 + \beta V). \tag{A15}$$

Inserting this into (A13), we get the desired upper bound. □

Proof of Lemma 5 To simplify the notations, we write $P_\theta = P$. We have $|\varphi(x)| \leq \|\varphi\|_\beta (1 + \beta V(x))$ from which we get

$$\begin{aligned} |P^* \varphi(x)| \leq P^* |\varphi|(x) &\leq \|\varphi\|_\beta (1 + P^* \beta V(x)) \\ &\leq \|\varphi\|_\beta (1 + \beta(\gamma_1 V(x) + K_1)), \end{aligned} \tag{A16}$$

by Assumption 6. The last term on the right-hand side is majorized by $\alpha'(1 + \beta V(x))$ with $\alpha' = \gamma_1 \vee (1 + \beta K_1)$, proving the first half of (90).

To prove the second half of (90), recall that for any $\psi \in \mathcal{L}_V$ we have $|||\psi|||_\beta = \min_c \|\psi + c\|_\beta$. Hence, for any constant c we have

$$|||P^* \varphi|||_\beta = |||P^* \varphi + c|||_\beta \leq \|P^* \varphi + c\|_\beta = \|P^*(\varphi + c)\|_\beta.$$

Apply the first inequality of (90) with $\varphi + c$ replacing φ :

$$\|P^*(\varphi + c)\|_\beta \le \alpha' \|\varphi + c\|_\beta. \tag{A17}$$

Choosing c so that $\|\varphi + c\|_\beta = |||\varphi|||_\beta$ yields the claim. □

Proof of Lemma 7 The proof is obtained by a simple modification of the proof of Lemma 3. Estimate $\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta)$ using a sequence of triangular inequalities to get

$$\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta) \le \sum_{k=0}^{n-1} \sigma_\beta(P_\theta^{n-k} P_{\theta'}^k \eta, P_\theta^{n-k-1} P_{\theta'}^{k+1} \eta).$$

Consider the kth term and apply Lemma 6, or (92), repeatedly $n-k-1$ times setting $\eta_1 = P_\theta P_{\theta'}^k \eta$ and $\eta_2 = P_{\theta'}^{k+1} \eta$:

$$\sigma_\beta(P_\theta^{n-k-1} \eta_1, P_\theta^{n-k-1} \eta_2) \le (\alpha')^{n-k-1} \sigma_\beta(\eta_1, \eta_2). \tag{A18}$$

Note that the conditions of Lemma 6 are satisfied for η_1, η_2: Obviously $\eta_1(\mathbf{X}) = \eta_2(\mathbf{X}) = \eta(\mathbf{X}) < \infty$ and $|\eta_i|(V) < \infty$, for $i = 1, 2$ due to the repeated application of the one-step growth condition. Combining the last two inequalities, we get:

$$\sigma_\beta(P_\theta^n \eta, P_{\theta'}^n \eta) \le \sum_{k=0}^{n-1} (\alpha')^{n-k-1} \sigma_\beta(P_\theta P_{\theta'}^k \eta, P_{\theta'}^{k+1} \eta). \tag{A19}$$

Consider the k-th term, and recall the Lipschitz continuity of (P_θ), Assumption 3, implying Lemma 2. Applying the latter for the signed measure $P_{\theta'}^k \eta$, we get the upper bound

$$\sum_{k=0}^{n-1} (\alpha')^{n-k-1} L_P |\theta - \theta'| \cdot |P_{\theta'}^k \eta|(1 + \beta V). \tag{A20}$$

To estimate $|P_{\theta'}^k \eta|(V)$, we use (115), restated as

$$|P_{\theta'}^k \eta|(V) \le \gamma_1^k \left(|\eta|(V) + K_1 |\eta|(\mathbf{X})/(\gamma_1 - 1) \right). \tag{A21}$$

By plugging this into (A20), we get the upper bound

$$L_P |\theta - \theta'| \sum_{k=0}^{n-1} (\alpha')^{n-k-1} \left(1 + \beta \gamma_1^k \left(|\eta|(V) + \frac{K_1 |\eta|(\mathbf{X})}{\gamma_1 - 1} \right) \right).$$

The first term in the above sum is bounded from above by $(\alpha')^n/(\alpha' - 1)$. The second term can be written as

$$\sum_{k=0}^{n-1} (\alpha')^{n-k-1} \gamma_1^k \beta \left(|\eta|(V) + \frac{K_1 |\eta|(\mathbf{X})}{\gamma_1 - 1} \right). \tag{A22}$$

Recall that $\gamma_1 \leq \alpha'$, hence a simplified upper bound is

$$n(\alpha')^{n-1}\beta\left(|\eta|(V) + \frac{K_1|\eta|(\mathbf{X})}{\gamma_1 - 1}\right), \tag{A23}$$

and $n(\alpha')^{n-1}$ can be bounded from above by $C(\alpha'')^{n-1}$ for any $\alpha'' > \alpha'$, where C depends only on α', and α''. Summarizing the inequalities (A19) to (A21), and the arguments that follow, we get the claim. □

Proof of Lemma 8 Since the function $g(\chi) := \mathbb{E}[e^{\chi U}]$ is convex in χ, the finite difference quotients are monotone non-increasing for $\chi \downarrow 0$ with negative limit:

$$\lim_{\chi\downarrow 0} \frac{\mathbb{E}[e^{\chi U}] - 1}{\chi} \downarrow \mathbb{E}\, U < 0. \tag{A24}$$

Hence, there exists a $0 < \chi_0 < \eta$ and some $\varepsilon > 0$ such that for all $0 < \chi \leq \chi_0$

$$\sup_{0<\chi\leq\chi_0} \frac{\mathbb{E}[e^{\chi U}] - 1}{\chi} \leq -\varepsilon < 0. \tag{A25}$$

It follows that $\mathbb{E}[e^{\chi U}] \leq 1 - \chi\varepsilon$, and thus we get

$$\begin{aligned} &\mathbb{E}[e^{\chi(x+U)^+}] \\ =&\mathbb{E}[\mathbf{1}_{\{x+U\geq 0\}}\, e^{\chi(x+U)^+}] + \mathbb{E}[\mathbf{1}_{\{x+U<0\}}\, e^{\chi(x+U)^+}] \\ \leq&\mathbb{E}[e^{\chi(x+U)}] + 1 \leq e^{\chi x}(1-\chi\varepsilon) + 1. \end{aligned} \tag{A26}$$

□

Proof of Lemma 10 Let us define the function

$$g(\chi) := \sup_{\theta\in D} g(\chi, \theta), \tag{A27}$$

for $0 \leq \chi < \eta$. Obviously, $g(\cdot)$ is convex and $g(0) = 1$. The claim of the lemma can be then restated as saying that there exists $\chi_0 > 0$ such that for $0 < \chi \leq \chi_0$ we have

$$\frac{g(\chi) - 1}{\chi} < 0 \quad \text{or} \quad \inf_{0<\chi<\chi_0} \frac{g(\chi) - 1}{\chi} < 0. \tag{A28}$$

Assume that the claim is not true, and let $\chi_n \downarrow 0$ be a monotone sequence such that

$$\frac{g(\chi_n) - 1}{\chi_n} \geq 0 \quad \text{or} \quad g(\chi_n) \geq 1. \tag{A29}$$

Let $\theta_n \in D$ be such that

$$g(\chi_n) = \sup_{\theta\in D} g(\chi_n, \theta) = \max_{\theta\in D} g(\chi_n, \theta) = g(\chi_n, \theta_n).$$

Due to the compactness of D, we can assume that $\theta_n \in D$ has a limit in D, say $\lim \theta_n = \theta^* \in D$. Consider now the function $g(\cdot, \theta^*)$ and choose a χ_0 such that

$$\frac{g(\chi_0, \theta^*) - 1}{\chi_0} < 0 \quad \text{or} \quad g(\chi_0, \theta^*) =: 1 - c < 1. \tag{A30}$$

The continuity of $g(\chi_0, .)$ in θ implies $g(\chi_0, \theta_n) \leq 1 - c/2 < 1$ for sufficiently large n. On the other hand, the convexity of the function $g(., \theta_n)$, and $g(0, \theta_n) = 1$ and $g(\chi_n, \theta_n) \geq 1$ imply that for $\chi_0 > \chi_n$ we have $g(\chi_0, \theta_n) \geq 1$, a contradiction, proving the claim. □

Proof of Lemma 11 By Lemma 10, it follows that for sufficiently small χ we have $A := \sup_{\theta \in D} \mathbb{E}[e^{\chi U_{\theta,1}}] < 1$, and hence,

$$\begin{aligned} &\sup_{\theta \in D} P(U_{\theta,1} \geq -\epsilon) = \sup_{\theta \in D} P(e^{\chi U_{\theta,1}} \geq e^{-\chi\epsilon}) \leq \\ &\sup_{\theta \in D} \mathbb{E}[e^{\chi U_{\theta,1}}] e^{\chi\epsilon} \leq A e^{\chi\epsilon}, \end{aligned} \tag{A31}$$

which is strictly smaller than 1 for ϵ small enough. Therefore, we indeed get the claim that, $\upsilon := \inf_{\theta \in D} P(U_{\theta,1} < -\epsilon) > 0$, as stated. □

Proof of Corollary 3 Indeed, let ϵ, υ be as in Lemma 11 and choose r so large that $r\epsilon > R$. Then, for all $\theta \in D$ and $0 \leq x \leq R$, $P(W_{\theta,r}(x) = 0)$ is bounded from below by

$$\begin{aligned} &P(W_{\theta,k}(x) \leq (x - k\epsilon)^+, \ \forall k = 1, \ldots, r) \geq \\ &P(U_{\theta,k} < -\epsilon, \ \forall k = 1, \ldots, r) \geq \upsilon^r. \end{aligned} \tag{A32}$$

□

Proof of Lemma 12 We can write

$$\int_0^\infty \phi(y)\, P_\theta(x, \mathrm{d}y) = \int_{-x}^\infty \phi(x+z)\, \zeta_\theta(z)\, \mathrm{d}z + \phi(0) \cdot P_\theta(x, \{0\}).$$

First, for the regular part (first term on the r.h.s.) we have

$$\begin{aligned} &\int_{-x}^\infty \phi(x+z)\, (\zeta_\theta(z) - \zeta_{\theta'}(z))\, \mathrm{d}z \\ &\qquad \leq \int_{-x}^\infty (1 + \beta V(x+z)) |\theta - \theta'|\, C'' e^{-\eta''|z|}\, \mathrm{d}z \\ &\qquad \leq C''|\theta - \theta'| \int_0^\infty (1 + \beta e^{\chi(x+z)})\, e^{-\eta'' z}\, \mathrm{d}z \\ &\qquad\quad + C''|\theta - \theta'| \int_{-\infty}^0 (1 + \beta e^{\chi(x+z)})\, e^{-\eta''|z|}\, \mathrm{d}z \\ &\qquad \leq C''|\theta - \theta'| \left(\frac{1}{\eta''} + \beta e^{\chi x} \frac{1}{\eta'' - \chi} \right) \end{aligned}$$

$$+ \; C''|\theta - \theta'| \left(\frac{1}{\eta''} + \beta e^{\chi x} \frac{1}{\eta'' + \chi} \right)$$
$$\leq \; C'''|\theta - \theta'|(1 + \beta V(x)), \tag{A33}$$

with some constant C'''. On the other hand, for the atomic component (second term on the r.h.s.), we get

$$\phi(0) \int_{-\infty}^{-x} (\zeta_\theta(z) - \zeta_{\theta'}(z)) \, \mathrm{d}z \leq$$
$$\phi(0) \int_{-\infty}^{-x} |\theta - \theta'| \, C'' e^{-\eta''|z|} \, \mathrm{d}z \leq$$
$$\phi(0) \; C''|\theta - \theta'| \, \frac{1}{\eta''}. \tag{A34}$$

Inequalities (A33) and (A34) imply the claim of the lemma. □

Acknowledgements A. Carè and B. Cs. Csáji were (partially) supported by the European Commission through the H2020 project Centre of Excellence in Production Informatics and Control (EPIC, 739592). B. Cs. Csáji was supported by the ADVANCED project, no. 153 390, of the NRDI (National Research, Development and Innovation) Office of Hungary. L. Gerencsér was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF−2.3.1-21-2022-00002). M. Rásonyi and B. Gerencsér were supported by NRDI grant KKP 137490. M. Rásonyi was also supported by the NRDI grant K 143529. B. Gerencsér was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, as well. The authors thank Máté Gerencsér for his comments on the potential of Markov processes defined by SPDEs.

Author Contributions All authors contributed significantly to the development of the mathematical theory as well as to the writing and editing of the manuscript.

Funding Open access funding provided by HUN-REN Institute for Computer Science and Control.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Robbins H, Monro S (1951) A stochastic approximation method. Ann Math Stat 22(3):400–407
2. Derevitskii DP, Fradkov AL (1974) Application of the theory of Markov-processes to the analysis of the dynamics of adaptation algorithms. Autom Remote Control 2:39–48

3. Derevitskii DP, Fradkov AL (1981) Applied theory of discrete adaptive control systems. Nauka, Moscow
4. Benveniste A, Métivier M, Priouret P (1990) Adaptive algorithms and stochastic approximations, 2nd edn. Springer, Berlin
5. Watkins CJ, Dayan P (1992) Q-learning. Mach Learn 8(3):279–292
6. Sutton RS (1988) Learning to predict by the methods of temporal differences. Mach Learn 3(1):9–44
7. Ljung L, Söderström T (1983) Theory and practice of recursive identification. MIT Press, Cambridge
8. Carè A, Csáji B Cs, Gerencsér B, Gerencsér L, Rásonyi M (2019) Parameter-dependent Poisson equations: tools for stochastic approximation in a Markovian framework. In: IEEE 58th conference on decision and control (CDC), pp. 2259–2264 (2019). https://doi.org/10.1109/CDC40024.2019.9029383
9. Carè A, Csáji B Cs, Gerencsér B, Gerencsér L, Rásonyi M (2019) Poisson equations, Lipschitz continuity and controlled queues. arXiv preprint arXiv:1906.09464
10. Cutkosky A, Orabona F (2018) Black-box reductions for parameter-free online learning in Banach spaces. In: Conference on learning theory (COLT), pp. 1493–1529. PMLR
11. Tamás A, Csáji B Cs (2024) Recursive estimation of conditional kernel mean embeddings. J Mach Learn Res 25(264):1–35
12. Liu SD, Chen S, Zhang S (2025) The ODE method for stochastic approximation and reinforcement learning with Markovian noise. J Mach Learn Res 26(24):1–76
13. Gerencsér L, Michaletzky Gy, Bokor J, Polcz P (2024) Notes on input design: from multi-sine design to data-driven procedures. IEEE Control Syst Lett 8:1943–1948
14. Ljung L (1977) Analysis of recursive stochastic algorithms. IEEE Trans Autom Control 22(4):551–575
15. Gerencsér L (1992) Rate of convergence of recursive estimators. SIAM J Control Optim 30(5):1200–1227
16. Borkar V, Chen S, Devraj A, Kontoyiannis I, Meyn S (2025) The ODE method for asymptotic statistics in stochastic approximation and reinforcement learning. Ann Appl Probab 35(2):936–982
17. Kushner HJ, Yin GG (2003) Stochastic approximation and recursive algorithms and applications. Springer, New York
18. Schweitzer PJ (1968) Perturbation theory and finite Markov chains. J Appl Probab 5(2):401–413
19. Hairer M, Mattingly JC (2011) Yet another look at Harris' ergodic theorem for Markov chains. In: Seminar on stochastic analysis, random fields and applications VI, pp. 109–117. Springer
20. Meyn SP, Tweedie RL (2012) Markov chains and stochastic stability. Springer, London
21. Borkar V, Chen S, Devraj A, Kontoyiannis I, Meyn SP (2024) The ODE method for asymptotic statistics in stochastic approximation and reinforcement learning. arXiv preprint arXiv:2110.14427v6
22. Finesso L, Gerencsér L, Kmecs I (1999) Estimation of parameters from quantized noisy observations. In: 1999 European control conference (ECC), pp. 1648–1653 IEEE
23. Finesso L, Gerencsér L, Kmecs I (1999) A randomized EM-algorithm for estimating quantized linear Gaussian regression. In: Proceedings of the 38th IEEE conference on decision and control (Cat. No. 99CH36304), vol. 5, pp. 5100–5101. IEEE
24. Xia F, Sun K, Yu S, Aziz A, Wan L, Pan S, Liu H (2021) Graph learning: a survey. IEEE Trans Artif Intell 2(2):109–127
25. Ferraty F, Vieu P (2006) Nonparametric functional data analysis: theory and practice, vol 76. Springer, New York
26. Lin T, Zha H (2008) Riemannian manifold learning. IEEE Trans Pattern Anal Mach Intell 30(5):796–809
27. Berglund N (2022) An introduction to singular stochastic PDEs: Allen-Cahn equations. Metastability and regularity structures. EMS Press, Berlin
28. Bertsekas D, Tsitsiklis JN (1996) Neuro-dynamic Programming. Athena Scientific, Belmont
29. Si J, Barto AG, Powell WB, Wunsch D (2004) Handbook of learning and approximate dynamic programming, vol 2. John Wiley & Sons, Hoboken, NJ
30. Berlinet A, Thomas-Agnan C (2011) Reproducing Kernel Hilbert spaces in probability and statistics. Springer, New York
31. Lakatos L, Szeidl L, Telek M (2019) Introduction to queueing systems with telecommunication applications. Springer, Cham
32. Loynes RM (1962) The stability of a queue with non-independent inter-arrival and service times. Math Proc Cambridge Philos Soc 58:497–520
33. Borovkov AA (1998) Egodicity and stability of stochastic processes. Wiley, New York
34. Diaconis P, Freedman D (1999) Iterated random functions. SIAM Rev 41(1):45–76

35. Sparre Andersen E (1957) On the collective theory of risk in the case of contagion between the claims. In: Transactions of the 15th international congress of actuaries, pp. 219–229

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.