J. Appl. Probab. 1–14 (2024)
*doi:*10.1017/jpr.2024.70

ON THE STRONG STABILITY OF ERGODIC ITERATIONS

LÁSZLÓ GYÖRFI,* *Budapest University of Technology and Economics*
ATTILA LOVAS,** *HUN-REN Alfréd Rényi Institute of Mathematics and Budapest University of Technology and Economics*
MIKLÓS RÁSONYI,*** *HUN-REN Alfréd Rényi Institute of Mathematics*

Abstract

We revisit processes generated by iterated random functions driven by a stationary and ergodic sequence. Such a process is called strongly stable if a random initialization exists for which the process is stationary and ergodic, and for any other initialization the difference of the two processes converges to zero almost surely. Under some mild conditions on the corresponding recursive map, without any condition on the driving sequence we show the strong stability of iterations. Several applications are surveyed such as generalized autoregression and queuing. Furthermore, new results are deduced for Langevin-type iterations with dependent noise and for multitype branching processes.

Keywords: Iterated random functions; strong stability; stationary and ergodic sequences; recursive maps; generalized autoregression

2020 Mathematics Subject Classification: Primary 60G10
Secondary 60K37; 60K25; 60J80

1. Introduction

We are studying stochastic processes defined by iterating random functions. For a measurable function $F:\mathbb{R}^d\times\mathbb{R}^k\to\mathbb{R}^d$, consider the following iteration: set $X_n=X_n(v)$ such that $X_0=v$, with a vector $v\in\mathbb{R}^d$, and let

$$X_{n+1}=F(X_n, Z_{n+1}), \tag{1}$$

where the driving sequence $\{Z_i\}_1^\infty$ is a stochastic process with values in $\mathbb{R}^k$.

In the standard setup, $\{Z_i\}_1^\infty$ is independent and identically distributed (i.i.d.) and so $\{X_i(v)\}_0^\infty$ is a homogeneous Markov process [17, 31]. Furthermore, (1) is called a *forward iteration*. If, in (1), $Z_1,\ldots,Z_n$ is replaced by $Z_n,\ldots,Z_1$, then the resulting iteration $\tilde{X}_n$ is called the *backward iteration* [40, 41]. Clearly, X_n and $\tilde{X}_n$ have the same distribution for each n.

Received 18 April 2023; accepted 12 June 2024.
* Postal address: Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary. Email: gyorfi@cs.bme.hu
** Postal address: Department of Analysis and Operations Research, Faculty of Natural Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary. Email: lovas@renyi.hu
*** Postal address: HUN-REN Alfréd Rényi Institute of Mathematics, Reáltanoda street 13-15, H-1053, Budapest. Email: rasonyi@renyi.hu

In the present paper the main role is played by another type of iteration, called *negative iteration*, defined as follows. For a $k \leq 0$, let the random double array $X_n^{(k)} = X_n^{(k)}(0)$, $k \leq n$, be defined such that $X_k^{(k)} = 0$ and

$$X_{n+1}^{(k)} = F\big(X_n^{(k)}, Z_{n+1}\big), \qquad n \geq k, \tag{2}$$

i.e. the iteration starts at negative time k with initial vector 0. This iteration scheme is also often used and appears, e.g., in [9, 10, 16, 17, 19, 20, 25, 26, 31].

Under mild conditions, [17] proved that the backward iteration $\tilde{X}_n$ is almost surely (a.s.) convergent to a random vector V with a distribution ν, which implies that the forward iteration X_n has the limit distribution ν. As in the standard setup of Markov chains, if X_0 has distribution ν (and it is independent from the driving sequence $\{Z_i\}_1^\infty$), then X_n will be a stationary Markov process.

More general schemes have also been considered, where $\{Z_i\}_1^\infty$ is merely stationary and ergodic [16, 19, 31]. In [9, 10, 20] such processes are treated under the name 'stochastically recursive sequences'. We remark for later use that, by the Doob–Kolmogorov theorem, a stationary sequence $\{Z_i\}_1^\infty$ can always be completed to a sequence $\{Z_i\}_{-\infty}^\infty$, defined on the whole integer lattice $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. We assume henceforth that this completion has been carried out.

In [13] the stationary process $\{X_i'\}_{-\infty}^\infty$ was called a *weak solution* of the iteration if there exists a $\{Z_i'\}_{-\infty}^\infty$ such that (X_i', Z_i') satisfies the recursion (1), and $\{Z_i\}_{-\infty}^\infty$ and $\{Z_i'\}_{-\infty}^\infty$ have the same distribution. $\{X_i\}_{-\infty}^\infty$ is called a *strong solution* if it is stationary and (X_i, Z_i) satisfies the recursion (1).

In this paper, we study the strong solutions. Using the almost sure limit of the negative iteration we construct such solutions under mild conditions. Actually, we investigate the following novel concept of *strong stability*.

Definition 1. The class of random processes $\{X_n(v), v \in \mathbb{R}^d\}$ is called strongly stable if:

(i) there exists a random vector V^* such that $\{X_i(V^*)\}_0^\infty$ is stationary and ergodic;

(ii) for any random vector V, $X_n(V) - X_n(V^*) \to 0$ a.s.

Note that in this definition the random initial vector V may depend on the entire trajectory of $\{Z_i\}_1^\infty$. As a result, the concept of strong stability may seem overly demanding. Furthermore, for integer-valued processes it follows from (ii) that there is a random index τ such that, for all $n > \tau$, we have $X_n(V) = X_n(V^*)$. In other words, $\{X_i(V)\}_0^\infty$ is *forward coupled* with $\{X_i(V^*)\}_0^\infty$. This stronger notion of stability was introduced in [35] and also discussed in [21]. Traditional proofs establishing the existence of a unique limiting distribution for Markov chains on Polish spaces under Doeblin's minorization condition involve representing transitions through iterated i.i.d. random maps and a coupling argument [5]. In this way, it can be shown that the iteration starting from any possibly random initial value is forward coupled with its stationary counterpart.

The aim of this paper is to show the strong stability of $\{X_n(v), v \in \mathbb{R}^d\}$ in great generality for several relevant models and important applications. As in [16], under some mild conditions on the function F we show that the almost sure limiting process $X_n^* = \lim_{k \to -\infty} X_n^{(k)}$ exists and is stationary and ergodic; [13] had similar results in the particular case of monotonic F, see (10).

Our main results are stated and proved in Section 2. Generalized autoregressions, queuing systems, and generalized Langevin dynamics are surveyed in Sections 3, 4, and 5, respectively.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Section 6 treats multi-type generalized Galton–Watson processes. Finally, Section 7 discusses open problems.

2. Iterated ergodic function systems

Defining $F_n(x) := F(x, Z_n)$, $n \in \mathbb{Z}$, we can write

$$X_n^{(k)}(v) = F_n \circ \cdots \circ F_{k+1}(v), \quad k \le 0, \;\; n \ge k,$$

where the empty composition is defined as the identity function.

In the following, $|\cdot|$ refers to the standard Euclidean norm on $\mathbb{R}^d$. For a function $g : \mathbb{R}^d \to \mathbb{R}^d$, set

$$\|g\| = \sup_{x \neq y} \frac{|g(x) - g(y)|}{|x - y|}.$$

The following theorem is an extension of [17, Theorem 5.1] and [31, Theorem 6.2]. It is contained in [19, Theorem 3] (except for proving (ii)). We provide a proof for completeness.

Theorem 1. *Assume that* $\{Z_i\}_{-\infty}^{\infty}$ *is a stationary and ergodic sequence. Suppose that*

(i) $\mathbb{E}\{(\log \|F_1\|)^+\} < \infty$, *and*

(ii) for some n,

$$\mathbb{E}\{\log \|F_n \circ \cdots \circ F_1\|\} < 0. \tag{3}$$

Then the class $\{X_n(v),\, v \in \mathbb{R}^d\}$ *is strongly stable.*

Notice that (3) is a sort of long-run contraction condition here.

Proof. For the stationary and ergodic process $\mathbf{Z} = \{Z_i\}_{-\infty}^{\infty}$, let f_n, $n = 1, 2, \ldots$, be vector-valued functions such that $f_i(T^i\mathbf{Z}) = X_i(0)$, where T stands for the shift transformation. Let's calculate $f_n(\mathbf{Z})$. If the process $\{X_n^{(k)}\}$ is defined in (2), then $f_n(\mathbf{Z}) = X_0^{(-n)}(0)$, i.e. $X_0^{(-n)}(0)$ is the value of the process at time 0, when the process started at negative time $-n$ with the 0 vector. We show that, under the conditions (i) and (ii),

$$X_0^{(-n)}(0) \text{ is a.s. convergent to a random vector } V^*. \tag{4}$$

It will be clear that $V^* = f(\mathbf{Z})$ for some suitable functional f, so we will in fact show that

$$f_n(\mathbf{Z}) \to f(\mathbf{Z}) \text{ a.s.} \tag{5}$$

As for (4), we show that this sequence is a.s. a Cauchy sequence, i.e. even

$$\sum_{n=1}^{\infty} \left| X_0^{(-n)}(0) - X_0^{(-n-1)}(0) \right| < \infty$$

holds a.s. Notice that iterating (2) yields

$$X_{-n}^{(-n-1)}(0) = F(X_{-n-1}^{(-n-1)}(0), Z_{-n}) = F(0, Z_{-n}) = X_{-n}^{(-n)}(F(0, Z_{-n})),$$

So

$$\begin{aligned} X_0^{(-n)}(0) - X_0^{(-n-1)}(0) &= X_0^{(-n)}(0) - X_0^{(-n)}(F(0, Z_{-n})) \\ &= F_0 \circ \cdots \circ F_{-n+1}(0 - F(0, Z_{-n})). \end{aligned}$$

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Thus, $\left|X_0^{(-n)}(0) - X_0^{(-n-1)}(0)\right| \le \|F_0 \circ \cdots \circ F_{-n+1}\| \cdot |F(0, Z_{-n})|$. We will show that the ergodicity of $\{Z_i\}_{-\infty}^{\infty}$ together with (i) and (ii) implies

$$\sum_{n=1}^{\infty} \|F_0 \circ \cdots \circ F_{-n+1}\| \cdot |F(0, Z_{-n})| < \infty \tag{6}$$

a.s., and so (4) is verified.

In the following, the key ingredient is [19, Proposition 2], which is an extension of Fűrstenberg and Kesten [22]. To prove (6), note that by [19, Proposition 2], the condition (3) implies that the sequence

$$E_n := \frac{1}{n}\mathbb{E}\{\log \|F_0 \circ \cdots \circ F_{-n+1}\|\} \to E, \quad n \to \infty,$$

with $E < 0$ such that

$$\frac{1}{n} \log \|F_0 \circ \cdots \circ F_{-n+1}\| \to E \tag{7}$$

a.s. Note that E is called the Lyapunov exponent. Next, we argue as in [31, Proposition 6.1]. We have $\|F_0 \circ \cdots \circ F_{-n+1}\| = \exp\{n(1/n) \log \|F_0 \circ \cdots \circ F_{-n+1}\|\}$. The limit in (7) implies that there are random integer n_0 and $a > 0$ such that, for all $n \ge n_0$,

$$\frac{1}{n} \log \|F_0 \circ \cdots \circ F_{-n+1}\| \le -a < 0.$$

Thus,

$$\sum_{n=n_0}^{\infty} \|F_0 \circ \cdots \circ F_{-n+1}\| \cdot |F(0, Z_{-n})| \le \sum_{n=n_0}^{\infty} |F(0, Z_{-n})|\mathrm{e}^{-na}.$$

Since $\lambda \ln^+ |F(0, Z_{-1})|$ is integrable for all $\lambda > 0$, it follows that

$$\sum_{n=1}^{\infty} \mathbb{P}(\lambda \ln |F(0, Z_{-1})| > n) < \infty.$$

Applying this observation for $\lambda := 2/\alpha$, the Borel–Cantelli lemma implies that $|F(0, Z_{-1})| < \mathrm{e}^{na/2}$ holds except for finitely many n a.s., which implies that $\sum_{n=n_0}^{\infty} |F(0, Z_{-n})|\mathrm{e}^{-na} < \infty$ a.s.

Because $X_i(V^*) = f(T^i\mathbf{Z})$, $\{X_i(V^*)\}_0^{\infty}$ is stationary and ergodic so Definition 1(i) is proved. Furthermore,

$$\|X_n(V) - X_n(V^*)\| \le \|F_n \circ \cdots \circ F_1\| \cdot |V - V^*| \to 0$$

a.s., as before. Thus, Definition 1(ii) is verified, too. □

Theorem 1 applies, in particular, under the one-step contraction condition (8) in the following result.

Proposition 1. *Assume that $\{Z_i\}_{-\infty}^{\infty}$ is a stationary and ergodic sequence such that the distribution of Z_1 is denoted by μ. Suppose that $\int |F(0, z)|\mu(\mathrm{d}z) < \infty$, and $|F(x, z) - F(x', z)| \le K_z|x - x'|$ with*

$$\mathbb{E}\{\log K_{Z_1}\} < 0. \tag{8}$$

Then the class $\{X_n(v),\ v \in \mathbb{R}^d\}$ is strongly stable.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Proof. This proposition is an easy consequence of Theorem 1, since

$$\begin{aligned}\mathbb{E}\{\log \|F_n \circ \cdots \circ F_1\|\} \leq \mathbb{E}\{\log (\|F_n\| \cdots \|F_1\|)\} &= n\mathbb{E}\{\log \|F_1\|\} \\ &\leq n\mathbb{E}\{\log K_{Z_1}\} < 0. \qquad \square\end{aligned}$$

Definition 2. We say that the class of strongly stable random processes $\{X_n(v),\ v \in \mathbb{R}^d\}$ satisfies the strong law of large numbers (SLLN) if $\mathbb{E}\{V^*\}$ is well-defined and finite, and, for any $v \in \mathbb{R}^d$, $\lim_n (1/n) \sum_{i=1}^n X_i(v) = \mathbb{E}\{V^*\}$ a.s.

Remark 1. Under the conditions of Theorem 1, if V^* has a well-defined and finite expectation $\mathbb{E}\{V^*\}$ then we have the SLLN:

$$\left|\frac{1}{n}\sum_{i=1}^n X_i(v) - \mathbb{E}\{V^*\}\right| \leq \left|\frac{1}{n}\sum_{i=1}^n X_i(V^*) - \mathbb{E}\{V^*\}\right| + \frac{1}{n}\sum_{i=1}^n |X_i(v) - X_i(V^*)|.$$

By Birkhoff's ergodic theorem, the first term on the right-hand side tends to 0 a.s., while the almost sure convergence of the second term follows from Theorem 1(ii).

Remark 2. Now we discuss some conditions guaranteeing $\mathbb{E}\{|V^*|\} < \infty$. By Fatou's lemma and the triangle inequality, we can write

$$\mathbb{E}\{|V^*|\} \leq \liminf_{n\to\infty} \mathbb{E}\left\{\left|X_0^{(-n)}(0)\right|\right\} \leq \sum_{k=0}^{\infty} \mathbb{E}\left\{\left|X_0^{(-k)}(0) - X_0^{(-k+1)}(0)\right|\right\},$$

which we can estimate further and obtain

$$\mathbb{E}\{|V^*|\} \leq \sum_{k=0}^{\infty} \mathbb{E}\{\|F_0 \circ \cdots \circ F_{-k+1}\| \cdot |F(0, Z_{-k})|\}.$$

For the sake of simplicity, assume for the moment that $z \mapsto |F(0, z)|$ is bounded by some constant C. By stationarity, it is enough to investigate $\mathbb{E}\{\|F_k \circ \cdots \circ F_1\|\}$. Here, either we can prescribe a 'contractivity in the long run'-type condition like

$$\limsup_{k\to\infty} \mathbb{E}^{1/k}\{\|F_k \circ \cdots \circ F_1\|\} < 1, \tag{9}$$

and then the nth-root test gives the desired result (i.e., $\mathbb{E}\{|V^*|\} < \infty$), or, by Hölder's inequality, we can write

$$\mathbb{E}\{\|F_k \circ \cdots \circ F_1\|\} \leq \prod_{j=1}^{k} \mathbb{E}^{1/k}\{\|F_j\|^k\} = \mathbb{E}\{\|F_1\|^k\},$$

and thus we have

$$\mathbb{E}\{|V^*|\} \leq \sup_z |F(0, z)| \mathbb{E}\left\{\sum_{k=1}^{\infty} \|F_1\|^k\right\} \leq C\mathbb{E}\left\{\frac{\|F_1\|}{1 - \|F_1\|}\right\}.$$

We should assume here that $\|F_1\| < 1$ a.s.; moreover, $\mathbb{E}\{\|F_1\|/(1 - \|F_1\|)\} < \infty$ and hence this approach looks much more restrictive than requiring (9).

As pointed out in [47], the long-time contractivity condition (9) is stronger than (3) in Theorem 1 or, equivalently, (7). On the other hand, in the i.i.d. case, (9) reduces to $\mathbb{E}\{\|F_1\|\} < 1$.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

If $\mathbb{E}\{\|F_1\|\} < 1$ fails in the i.i.d. case then $E[V^*] = \infty$ can easily happen; see the example in Remark 3.

Remark 3. Concerning the SLLN, we should have to verify that $\mathbb{E}\{\sup_n |X_0^{(-n)}(0)|\} < \infty$. Note that $X_0^{(0)}(0) = 0$, and thus

$$\sup_n |X_0^{(-n)}(0)| \le \sum_{k=1}^{\infty} \left| X_0^{(-k+1)}(0) - X_0^{(-k)}(0) \right|.$$

One possibility would be to check that

$$\sum_{k=0}^{\infty} \mathbb{E}\{|X_0^{(-k+1)}(0) - X_0^{(-k)}(0)|\} \le \sum_{k=0}^{\infty} \mathbb{E}\left\{ |F(0, Z_{-k})| \prod_{j=-k+1}^{0} K_{Z_j} \right\} < \infty.$$

Here is a counterexample, however, to show that this is not true in general. Let $d = k = 1$ and $\{Z_i\}_{i \in \mathbb{Z}}$ be i.i.d. such that $\mathbb{P}\{Z_0 = \mathrm{e}^{-2}\} = \frac{2}{3}$ and $\mathbb{P}\{Z_0 = \mathrm{e}^{2}\} = \frac{1}{3}$. Put $X_0 = v > 0$ and $X_{n+1} = Z_{n+1} \cdot X_n$. Clearly, $K_{Z_0} = Z_0$ with $\mathbb{E}\{\log K_{Z_0}\} = -\frac{2}{3} < 0$ and so the conditions of Proposition 1 are satisfied. Furthermore, $\mathbb{E}\{K_{Z_0}\} = \mathbb{E}\{Z_1\} = \frac{2}{3}\mathrm{e}^{-2} + \frac{1}{3}\mathrm{e}^{2} > 1$. By independence,

$$\mathbb{E}\{X_{n+1}\} = \mathbb{E}\{Z_{n+1}\} \cdot \mathbb{E}\{X_n\} = \mathbb{E}\{Z_1\}^{n+1} \cdot v = \left(\tfrac{2}{3}\mathrm{e}^{-2} + \tfrac{1}{3}\mathrm{e}^{2}\right)^{n+1} \cdot v \to \infty$$

as $n \to \infty$. Therefore, $\mathbb{E}\{V^*\} = \infty$.

Next, the contraction condition of Theorem 1 is replaced by a monotonicity assumption. We denote by $\mathcal{F}_t$ the sigma-algebra generated by Z_j, $-\infty < j \le t$. Furthermore, we use the notation $\mathbb{R}_+^d$ for the positive orthant endowed with the usual coordinatewise partial ordering, i.e., $x \le y$ for $x, y \in \mathbb{R}^d$ when each coordinate of x is less than or equal to the corresponding coordinate of y. In what follows, $|x|_p = [|x_1|^p + \cdots + |x_d|^p]^{1/p}$ stands for the usual l_p-norm on $\mathbb{R}^d$.

Proposition 2. *Assume that* $\{Z_i\}_{-\infty}^{\infty}$ *is a stationary and ergodic sequence, and* $F : \mathbb{R}_+^d \times \mathbb{R}^k \to \mathbb{R}_+^d$ *is monotonic in its first argument:*

$$F(x, z) \le F(x', z) \quad \text{if } x \le x'. \tag{10}$$

For a fixed $1 \le p < \infty$*, suppose that there exist a constant* $0 \le \rho < 1$ *and* $K > 0$ *such that*

$$\mathbb{E}[|F(x, Z_1)|_p \mid \mathcal{F}_0] \le \rho |x|_p + K \tag{11}$$

a.s. for all $x \in \mathbb{R}^d$*, and*

$$\mathbb{E}[|F(0, Z_0)|_p] < \infty. \tag{12}$$

Then, the class $\{X_n(v), v \in \mathbb{R}^d\}$ *is strongly stable. Furthermore, the SLLN is satisfied.*

Notice that (11) also implies, for all t,

$$\mathbb{E}[|F(x, Z_{t+1})|_p \mid \mathcal{F}_t] \le \rho |x|_p + K \tag{13}$$

a.s. for all $x \in \mathbb{R}_+^d$. [9, 21, 40] studied the monotonic iteration under the fairly restrictive condition that the range of the iteration is a bounded set.

Proof. This proposition extends the Foster–Lyapunov stability argument to a non-Markovian setup. We apply the notations in the proof of Theorem 1 such that we verify the condition (5) or,

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

equivalently, check (4). Since $F_n(\cdot) := F(\cdot, Z_n) : \mathbb{R}_+^d \to \mathbb{R}_+^d$, $n \in \mathbb{Z}$, are order-preserving maps and $F(0, Z_{-n}) \geq 0$,

$$\begin{aligned} X_0^{(-n-1)}(0) - X_0^{(-n)}(0) &= X_0^{(-n)}(F(0, Z_{-n})) - X_0^{(-n)}(0) \\ &= F_0 \circ \cdots \circ F_{-n+1}(F(0, Z_{-n})) - F_0 \circ \cdots \circ F_{-n+1}(0) \end{aligned}$$

is non-negative, therefore $\big(X_0^{(-n)}(0)\big)_{n\in\mathbb{N}}$ is monotonically increasing and so (4) is verified.

As for Theorem 1(i), we need that $V^* = \lim_{n\to\infty} X_0^{(-n)}(0)$ takes finite values a.s., which would follow from $\mathbb{E}\big\{\sup_n \big|X_0^{(-n)}(0)\big|_p\big\} < \infty$. It is easily seen that $x \leq y$ implies $|x|_p \leq |y|_p$ for $x, y \in \mathbb{R}_+^d$, and therefore, by the Beppo Levi theorem and the monotonicity of $X_0^{(-n)}(0)$, $n \in \mathbb{N}$, $\mathbb{E}\big\{\sup_n \big|X_0^{(-n)}(0)\big|_p\big\} = \sup_n \mathbb{E}\big\{\big|X_0^{(-n)}(0)\big|_p\big\}$. From (13) we get that

$$\mathbb{E}\big[\big|X_{i+1}^{(-n)}(0)\big|_p \mid \mathcal{F}_i\big] = \mathbb{E}\big[\big|F(X_i^{(-n)}(0), Z_{i+1})\big|_p \mid \mathcal{F}_i\big] \leq \rho\big|X_i^{(-n)}(0)\big|_p + K, \quad i \geq -n.$$

Iterating this leads to

$$\mathbb{E}\big[\big|X_0^{(-n)}(0)\big|_p\big] \leq \rho^n \mathbb{E}\big[\big|X_{-n}^{(n)}(0)\big|_p\big] + \sum_{j=0}^{n-1} K\rho^j \leq \frac{K}{1-\rho} < \infty;$$

hence, $\sup_n \mathbb{E}\big\{\big|X_0^{(-n)}(0)\big|_p\big\}$ is finite, which completes the proof. □

Remark 4. Proposition 2 holds with an analogous proof if we replace (13) by

$$\mathbb{E}[|F(x, Z_{t+1})|_p \mid \mathcal{F}_t] \leq \rho(t)|x|_p + K,$$

where $\rho(t)$, $t \in \mathbb{Z}$, is a stationary process adapted to $\mathcal{F}_t$ and $\limsup_{n\to\infty} \mathbb{E}^{1/n}\{\rho(1)\cdots\rho(n)\} < 1$.

3. Generalized autoregression

In this section, $\|A\|$ will denote the operator norm of a matrix A. Several authors investigated the iteration of matrix recursion. Set $X_n = X_n(v)$ such that $X_0 = v$, and

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad n \geq 0, \tag{14}$$

where $\{(A_n, B_n)\}$ are i.i.d., A_n is a $d \times d$ matrix, and B_n is a d-dimensional vector. In this section we study a more general case of (14) where the sequence $\{(A_n, B_n)\}$ is stationary and ergodic. The minimal sufficient condition for the existence of a stationary solution for (14) was proved in [12]. In the i.i.d. case, [11] showed that those conditions are indeed minimal.

For the stationary and ergodic case, we now reprove the sufficiency part of [11, Theorem 2.5].

Proposition 3. *Assume that* $\{(A_n, B_n)\}$ *is stationary and ergodic such that* $\mathbb{E}\{\log^+ \|A_0\|\} < \infty$ *and* $\mathbb{E}\{\log^+ |B_0|\} < \infty$. *If*

$$\mathbb{E}\{\log ||A_0 \cdots A_{-n}||\} < 0 \tag{15}$$

for some n, *then the class* $\{X_n(v), v \in \mathbb{R}^d\}$ *is strongly stable.*

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Proof. Since $\|A_n A_{n-1} \cdots A_1\|$ corresponds to $\|F_n \circ \cdots \circ F_1\|$, we can verify the conditions of Theorem 1 easily: (15) implies (ii) and the integrability conditions of Proposition 1 imply (i). □

Remark 5. Note that the argument above showed up first in [22], which proved that if $\mathbb{E}\{\log^+ \|A_0\|\} < \infty$, then $\lim_n (1/n) \log \|A_n A_{n-1} \cdots A_1\| = E$ a.s., where E stands for the Lyapunov exponent.

In the rest of this section we recall results about a similar iteration which are based on negative iterations (though they cannot be treated by our results in this paper). Let us consider the stochastic gradient method for least-squares regression, when there are given observation sequences of random, symmetric, and positive semi-definite $d \times d$ matrices A_n and random d-dimensional vectors V_n such that $A = \mathbb{E}(A_n)$ and $V = \mathbb{E}(V_n)$ $(n = 0, \pm 1, \pm 2, \ldots)$. If A^{-1} exists then the aim is to estimate $\vartheta = A^{-1}V$. For this reason, a stochastic gradient algorithm with constant gain is introduced. Set $X_0 = v$, and

$$X_{n+1} = X_n - \lambda(A_{n+1}X_n - V_{n+1}), \quad n \geq 0, \tag{16}$$

followed by an averaging, $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$.

If λ depends on n, then $\bar{X}_n$ is called an averaged stochastic approximation introduced in [39, 45, 46]. If the sequence $\{(A_n, V_n)\}_{-\infty}^{\infty}$ is i.i.d., then they proved the optimal rate of convergence of $\bar{X}_n$ to ϑ.

The case of dependent $\{(A_n, V_n)\}_{-\infty}^{\infty}$ was studied in [26]. Assume that the sequence $\{(A_n, V_n)\}_{-\infty}^{\infty}$ is stationary and ergodic such that $\mathbb{E}\|A_n\| < \infty$, $\mathbb{E}|V_n| < \infty$, and A is positive definite. Then there is a $\lambda_0 > 0$ such that, for all $0 < \lambda < \lambda_0$, there exists a stationary and ergodic process $\{X_n^*\}_{-\infty}^{\infty}$ satisfying the recursion (16) and $\lim_n (X_n - X_n^*) = 0$ a.s. Moreover, $\lim_n \bar{X}_n = \vartheta + \delta_\lambda$ a.s. with an asymptotic bias vector δ_λ. In [26] there is a 3-dependent example of $\{(A_n, V_n)\}_{-\infty}^{\infty}$, where $\delta_\lambda \neq 0$. Furthermore, under a suitable mixing condition on $\{(A_n, V_n)\}_{-\infty}^{\infty}$, $|\delta_\lambda|$ is of order $\sqrt{\lambda}$.

4. Lindley process

We recall some results from queuing theory. They do not follow from arguments of the present paper but they are also based on negative iterations and hence provide one more illustration of the usefulness of this technique.

For $d = k = 1$, consider the following iteration. Set $X_n = X_n(v)$ such that $X_0 = v \geq 0$, and $X_{n+1} = (X_n + Z_{n+1})^+$. The next proposition is an extension of the concept of strong stability. Let $\{X_i'\}_0^{\infty}$ be a stationary and ergodic sequence. Recall that $\{X_i\}_0^{\infty}$ is *forward coupled* with $\{X_i'\}_0^{\infty}$ if there is a random index τ such that, for all $n > \tau$, $X_n' = X_n$.

Proposition 4. *Assume that $\{Z_i\}_{-\infty}^{\infty}$ is a stationary and ergodic sequence with $\mathbb{E}\{Z_1\} < 0$. Put $V^* = \sup_{j \leq 0} (Z_j + \cdots + Z_0)^+$ and $X_n' = X_n(V^*)$. Then $\{X_i'\}_0^{\infty}$ is stationary and ergodic, and $\{X_i\}_0^{\infty}$ is forward coupled with $\{X_i'\}_0^{\infty}$.*

As an application of Proposition 4 consider the extension of the G/G/1 queuing model. Let X_n be the waiting time of the nth arrival, S_n be the service time of the nth arrival, and T_{n+1} be the inter-arrival time between the $(n+1)$th and nth arrivals. Then, we get the recursion $X_{n+1} = (X_n - T_{n+1} + S_n)^+$.

The generalized G/G/1, where either the arrival times, or the service times, or both are not memoryless, was studied in [25, 37]; see also [2, 3, 9, 13, 23].

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Proposition 4 implies that if $\{Z_i\} := \{S_{i-1} - T_i\}_{-\infty}^{\infty}$ is a stationary and ergodic sequence with $\mathbb{E}\{S_0\} < \mathbb{E}\{T_1\}$, then $\{X_i\}_0^{\infty}$ is forward coupled with a stationary and ergodic $\{X_i'\}_0^{\infty}$.

5. The Langevin iteration

For a measurable function $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$, the Langevin iteration is defined as follows. Set $X_n = X_n(v)$ such that $X_0 = v$, and

$$X_{n+1} = X_n - \lambda H(X_n, Y_n) + \sqrt{2\lambda} N_{n+1}, \quad n \geq 0, \tag{17}$$

where $\{Y_i\}_{-\infty}^{\infty}$ and $\{N_i\}_{-\infty}^{\infty}$ are random sequences. In the literature of stochastic approximation $\lambda > 0$ is called the step size, while in machine learning it is called the learning rate.

The simplest case is where the sequences N_n and Y_n are independent, Y_n is i.i.d., and N_n is i.i.d. standard d-dimensional Gaussian. This algorithm was introduced in [50] and later analyzed by a large corpus of literature which we cannot review here. It is called 'stochastic gradient Langevin dynamics' and it can be used for sampling from high-dimensional, not necessarily log-concave, distributions and for finding the global minimum of high-dimensional functionals. In this context, Y_n represents a data sequence (obtained, for instance, by averaging a big dataset over randomly chosen minibatches), and N_n is artificially added noise to guarantee that the process does not get stuck near local minima.

The case where the data sequence Y_n may be a dependent stationary process (but N_n is still i.i.d. standard Gaussian) has been treated less extensively: see [4, 15] for the convex and [14] for the non-convex settings.

Another stream of literature, starting from [30], concentrated on stochastic differential equations driven by colored Gaussian noise. The discrete-time case of difference equations was treated in [49]. This setting corresponds to the case where in (17), Y_n is constant and N_n is a dependent Gaussian sequence.

We know of no studies so far that allowed *both* Y_n and N_n to be only stationary. We manage to establish strong stability in this case, under reasonable assumptions.

Defining $F(x, z) = x - \lambda H(x, y) + \sqrt{2\lambda} u$, $z = (y, u)$, and $Z_{i+1} := (Y_i, N_{i+1})$, Proposition 1 implies the following result.

Corollary 1. *Assume that the sequence* $\{(Y_i, N_{i+1})\}_{-\infty}^{\infty}$ *is stationary and ergodic, and that, for a* $\lambda_0 > 0$ *and for all* $0 < \lambda < \lambda_0$, $\mathbb{E}|-\lambda H(0, Y_1) + \sqrt{2\lambda} N_2| < \infty$ *and*

$$|x - \lambda H(x, z) - (x' - \lambda H(x', z))| \leq K_z |x - x'|$$

with (8). *Then, for all* $0 < \lambda < \lambda_0$, *the class* $\{X_n(v), v \in \mathbb{R}^d\}$ *is strongly stable.*

Remark 6. Next, we prove the convergence of the iterative scheme (17) assuming only that H satisfies

$$\langle \partial_1 H(x, y) v, v \rangle \geq m(y) |v|^2 \quad \text{and} \quad |\partial_1 H(x, y)| \leq M(y) \tag{18}$$

with measurable $m, M : \mathbb{R}^m \to [0, \infty)$. (This is a parametric form of the usual *strong convexity condition*. We can replace it by a so-called *dissipativity* condition and hence extend the analysis beyond the convex case. However, this direction requires a different technology.) We introduce $g(t) = H(tx' + (1-t)x, y)$, and thus we can write

$$|F(x, y) - F(x', y)|^2 = |x - x'|^2 - 2\lambda \langle x - x', g(0) - g(1) \rangle + \lambda^2 |g(0) - g(1)|^2,$$

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

where $g(0) - g(1) = \int_0^1 \partial_1 H(tx' + (1-t)x, y)(x - x')\,\mathrm{d}t$. Using (18), we estimate $|g(0) - g(1)| \le M_y|x - x'|$ and $\langle x - x', g(0) - g(1)\rangle \ge m_y|x - x'|^2$, and arrive at

$$|F(x, y) - F(x', y)| \le (1 + \lambda^2 M_y^2 - 2\lambda m_y)^{1/2}|x - x'|.$$

In (18), without loss of generality, we can assume that $\mathbb{E}(m_{Z_1}) < \infty$. Furthermore, requiring $\mathbb{E}(M_y^2) < \infty$, we can set λ such that the conditions of Proposition 1, i.e. (8), are satisfied. It is also not restrictive to assume that $\mathbb{E}(m_{Z_1})^2 < \mathbb{E}(M_{Z_1}^2)$. Since $K_{Z_1} = (1 + \lambda^2 M_{Z_1}^2 - 2\lambda m_{Z_1})^{1/2}$, Jensen's inequality implies that

$$\mathbb{E}(K_{Z_1}) \le (1 + \lambda^2\mathbb{E}(M_{Z_1}^2) - 2\lambda\mathbb{E}(m_{Z_1}))^{1/2} < 1$$

whenever $\lambda < 2\mathbb{E}(m_{Z_1})/\mathbb{E}(M_{Z_1}^2)$.

Remark 7. Strong convexity is a usual assumption in the stochastic gradient Langevin dynamics literature [4, 15, 18]. Remark 6 shows that our results cover this case. It would, however, be nice to weaken this condition to dissipativity (see [14, 36]). It seems that such a generalization requires much more advanced techniques; see, e.g., [49].

Remark 8. Note that the iteration (16) in Section 3 is a special case of the Langevin iteration (17) such that $H(X_n, A_{n+1}) = A_{n+1}X_n$, with the difference that for this, only convexity holds and not strong convexity. In a least-squares regression setup it is important that A_n is assumed positive semi-definite only and not necessarily positive definite.

6. Generalized multi-type Galton–Watson process

We follow the notation of [33]. A d-type Galton–Watson branching process with immigration (GWI process) $\mathbf{X}_n = (X_{n,1}, \ldots, X_{n,d})$, $n \in \mathbb{Z}_+$, is defined as

$$\begin{cases} \mathbf{X}_n = \sum_{j=1}^{X_{n-1,1}} \mathbf{A}_{n,j,1} + \cdots + \sum_{j=1}^{X_{n-1,d}} \mathbf{A}_{n,j,d} + \mathbf{B}_n, & n \ge 1, \\ \mathbf{X}_0 = v, \end{cases}$$

where $v \in \mathbb{N}^d$ and $\{\mathbf{A}_{n,j,i}, \mathbf{B}_n : n, j \in \mathbb{N},\ i \in \{1, \ldots, d\}\}$ are random vectors with non-negative integer coordinates. Here, $X_{n,i}$ is the number of i-type individuals in the nth generation of a population, $\mathbf{A}_{n,j,i}$ is the vector of the number of offspring produced by the jth individual of type i belonging to the $(n-1)$th generation, and $\mathbf{B}_n$ is the vector of the number of immigrants.

Let $\mathbf{C}_n := \{\mathbf{A}_{n,j,i} : j \in \mathbb{N},\ i \in \{1, \ldots, d\}\}$. In the standard setup the families of random variables $\{\mathbf{C}_n : n \in \mathbb{N}\}$ and $\{\mathbf{B}_n : n \in \mathbb{N}\}$ are independent and $(\mathbf{C}_n, \mathbf{B}_n)$, $n \in \mathbb{N}$, is a sequence of independent vectors. The process $\{\mathbf{X}_n : n \in \mathbb{N}\}$ is called homogeneous when $(\mathbf{C}_n, \mathbf{B}_n)$, $n \in \mathbb{N}$, are identically distributed, otherwise it is inhomogeneous. In this section we study the generalization of the homogeneous case, when $\{Z_n := (\mathbf{C}_n, \mathbf{B}_n) : n \in \mathbb{N}\}$ is a stationary and ergodic process. As before, we extend this stationary process to the timeline $\mathbb{Z}$. We furthermore assume for each i that $\mathbf{A}_{1,j,i}$ has, for each $j \in \mathbb{N}$, the same conditional law with respect to $\mathcal{F}_0$.

Note that the state space of Z_n is $\mathbb{R}^{\mathbb{N}}$. It can easily be checked that all the arguments of our paper also apply for such state spaces.

Homogeneous multi-type GWI processes were introduced and studied in [42, 43]. In [42], a necessary and sufficient condition is given for the existence of a stationary distribution in the subcritical case. A complete answer is given by [32]. Also, [38] gives a sufficient condition for the existence of a stationary distribution, and in a special case shows that the limiting distribution is a multivariate Poisson process with independent components.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

Branching process models are extensively used in various parts of the natural sciences, in biology, epidemiology, physics, and computer science, among other subjects. In particular, multi-type GWI processes were used to determine the asymptotic mean and covariance matrix of deleterious and mutant genes in a stationary population in [24], and in a non-stationary population in [34]. Another rapidly developing area where multi-type GWI processes can be applied is the theory of polling systems: [44] pointed out that a large variety of polling models can be described as a multi-type GWI process. [1, 6, 7, 44, 48] investigated several communication protocols applied in info-communication networks with differentiated services. There are different quality of service requirements, e.g. some of them are delay sensitive (telephone, online video, etc.), while others tolerate some delay (email, internet, downloading files, etc.). Thus, the services are grouped into service classes such that each class has its own transmission protocol, like priority queuing. In the papers mentioned above the d-type Galton–Watson process has been used, where the process was defined either by the sizes of the active user populations of the d service classes, or by the length of the d priority queues. For the general theory and applications of multi-type Galton–Watson processes we refer to [29, 38].

Define the random row vectors $m_i := \mathbb{E}[\mathbf{A}_{1,1;i} \mid \mathcal{F}_0]$, $i = 1, \ldots, d$, where $\mathcal{F}_t$ is the sigma-algebra generated by Z_j, $-\infty < j \leq t$. Note that, by our assumptions, m_i has the same law as $\mathbb{E}[\mathbf{A}_{n+1,j;i} \mid \mathcal{F}_n]$ for each n and j. For $x \in \mathbb{R}^d$ we will use the ℓ_1-norm $|x|_1 := \sum_{j=1}^{d} |x_i|$, where x_i is the ith coordinate of x.

Proposition 5. *If*

$$\max_{1\leq i\leq d} |m_i|_1 \leq \varrho \tag{19}$$

almost surely for some constant $\varrho < 1$ and $\mathbb{E}\{|\mathbf{B}_1|_1\} < \infty$, then $\mathbf{X}_t = \mathbf{X}_t(v)$ is strongly stable and the SLLN holds.

Proof. Define $F(x, z) := \sum_{j=1}^{d} \sum_{i=1}^{x_j} z_{i;j} + z_0$, where $z_{i;j} \in \mathbb{N}^d$, $i \in \mathbb{N}$, $1 \leq j \leq d$, $z_0 \in \mathbb{N}^d$. Note that this iteration is monotone. As already defined, the stationary process will be $Z_n = ((\mathbf{A}_{n,i;j})_{1\leq j\leq d, i\in\mathbb{N}}, \mathbf{B}_n)$.

Let us check (11) with $p = 1$:

$$\begin{aligned}
\mathbb{E}[|F(x, Z_1)|_1 \mid \mathcal{F}_0] &\leq \sum_{j=1}^{d} \sum_{i=1}^{x_j} \mathbb{E}[|\mathbf{A}_{1,i;j}|_1 \mid \mathcal{F}_0] + \mathbb{E}[|\mathbf{B}_1|_1] \\
&\leq \sum_{j=1}^{d} x_j\varrho + \mathbb{E}[|\mathbf{B}_1|_1] = \varrho|x|_1 + \mathbb{E}[|\mathbf{B}_1|_1].
\end{aligned}$$

Note also that $\mathbb{E}[|F(0, Z_1)|] = \mathbb{E}[|\mathbf{B}_1|_1] < \infty$, as required by (12). We may conclude the required result from Proposition 2. □

Remark 9. In the case where the sequence $\mathbf{A}_{n,\cdot;\cdot}$, $n \in \mathbb{N}$, is i.i.d., we have $m_i = \mathbb{E}[\mathbf{A}_{1,1;i} \mid \mathcal{F}_0] = \mathbb{E}[\mathbf{A}_{1,1;i}]$. In that case, the standard assumption is that the matrix M composed from row vectors m_i satisfies

$$\varrho(M) < 1, \tag{20}$$

where $\varrho(M)$ denotes the spectral radius [33]. Our (19) is stronger than (20). In arguments for i.i.d. $\mathbf{A}_{n,\cdot;\cdot}$, the general case (20) can be easily reduced to (19). However, it is not clear to us how to do so in the current, non-independent, setting. Perhaps the techniques of [16, Theorem 4] could be adapted.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

7. Open problems

The study of iterated random functions driven by stationary and ergodic sequences, particularly their strong stability, opens several intriguing avenues for future research. While we have established foundational results and explored various applications, there remain numerous questions that merit further investigation. Addressing these open problems could not only deepen our understanding of the underlying theoretical framework but also expand its applicability to broader contexts. Below, we outline several key areas where additional research could be highly beneficial, identifying specific challenges and potential directions for future work.

For instance, Theorem 1 implies weak convergence of the laws of the process X_n. A particularly intriguing question is under what conditions we could strengthen this to, e.g., convergence in total variation. Furthermore, it is also not uninteresting whether we could relax the contractivity condition (3) to some kind of dissipativity condition. This was done in [49] only for a particular class of Gaussian infinite moving average processes Z_n, using heavy technicalities. As for the strong law of large numbers, the condition deduced in Remark 2 to ensure $\mathbb{E}\{|V^*|\} < \infty$ also looks too restrictive. Can this assumption be weakened in a reasonable way?

The Lindley iteration is addressed in [17, Theorem 4.1] when $\{Z_i\}_{-\infty}^{\infty}$ is i.i.d., noting that the condition $\mathbb{E}\{Z_1\} < 0$ can be weakened to

$$\sum_{j=-\infty}^{0} \frac{\mathbb{P}\{Z_j + \cdots + Z_0 > 0\}}{j} < \infty.$$

We hypothesize that this observation is valid for the ergodic case as well.

In Section 5, we studied Langevin iteration driven by stationary and ergodic noise under the strong convexity assumption. It would be desirable to replace this with a dissipativity condition, thereby extending the analysis beyond the convex case.

The limit distribution of the inhomogeneous Galton–Watson processes, which means that the driving processes are independent but not identically distributed, was studied in [27, 28]. A challenging research problem is how to weaken the independence. Finally, we conjecture that the condition (19) we imposed on the expected number of offspring in the Galton–Watson process could certainly be relaxed with more sophisticated arguments.

Acknowledgements

We would like to express our sincere gratitude to the anonymous reviewer for the thorough and insightful evaluation of our paper. The careful examination not only identified errors that needed correction but also provided valuable suggestions that significantly enhanced the overall quality of the paper. We would also like to extend our thanks to the editorial team of the *Journal of Applied Probability* for facilitating this constructive review process. The thoughtful feedback received has undoubtedly strengthened the clarity and rigor of our work.

Funding information

Attila Lovas and Miklós Rásonyi were supported by the National Research, Development and Innovation Office within the framework of the Thematic Excellence Program 2021; National Research subprogramme ‘Artificial intelligence, large networks, data security: mathematical foundation and applications’, and also by the grant K 143529.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Altman, E. and Fiems, D. (2012). Branching processes and their generalization applied to wireless networking. In *Paradigms for Biologically-Inspired Autonomic Networks and Services. The BIONETS Project eBook*, eds Eitan Altman, Paolo Dini and Daniele Miorandi. pp. 54–66.

[2] Asmussen, S. (2003). *Applied Probability and Queues*. Springer, New York.

[3] Baccelli, F. and Brémaud, P. (2003). *Elements of Queueing Theory*. Springer, New York.

[4] Barkhagen, M., Chau, N. H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021). On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case. *Bernoulli* **27**, 1–33.

[5] Bhattacharya, R. and Waymire, E. C. (2002). An approach to the existence of unique invariant probabilities for Markov processes. In *Limit Theorems in Probability and Statistics* Vol. I, eds I. Berkes, E. Csaki and M. Csorgo. Janos Bolyai Mathematical Society, Budapest, pp. 181–200.

[6] Boon, M. A. A. (2012). A polling model with reneging at polling instants. *Ann. Operat. Res.* **198**, 5–23.

[7] Boon, M. A. A., Adan, I. J. B. F. and Boxma, O. J. (2010). A polling model with multiple priority levels. *Performance Evaluation* **67**, 468–484.

[8] Borovkov, A. A. (1976). *Stochastic Processes in Queueing Theory*. Springer, New York.

[9] Borovkov, A. A. (1998). *Ergodicity and Stability of Stochastic Processes*. John Wiley, Chichester.

[10] Borovkov, A. A. and Foss, S. G. (1993). Stochastically recursive sequences and their generalizations. *Trudy Inst. Mat. SO RAN* **20**, 32–103.

[11] Bougerol, P. and Picard, N. (1992). Strict stationarity on generalized autoregressive processes, *Ann. Prob.* **20**, 1714–1730.

[12] Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Prob.* **18**, 211–220.

[13] Brandt, A., Franken, P. and Lisek, B. (1990). *Stationary Stochastic Models*. John Wiley, New York.

[14] Chau, H. N., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021). On stochastic gradient Langevin dynamics with dependent data streams: The fully non-convex case. *SIAM J. Math. Data Sci.* **3**, 959–986.

[15] Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Statist. Soc. B* **79**, 651–676.

[16] Debaly, Z. M. and Truquet, L. (2021). Iterations of dependent random maps and exogeneity in nonlinear dynamics. *Econometric Theory* **37**, 1135–1172.

[17] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 45–76.

[18] Durmus, A. and Moulines, É. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli* **25**, 2854–2882.

[19] Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stoch. Process. Appl.* **34**, 39–47.

[20] Foss, S. G. and Konstantopoulos, P. T. (2003). Extended renovation theory and limit theorems for stochastic ordered graphs. *Markov Process. Relat. Fields* **9**, 413–468.

[21] Foss, S. G. and Tweedie, R. L. (1998). Perfect simulation and backward coupling. *Stoch. Models* **14**, 187–203.

[22] Fűrstenberg, H. and Kesten, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31**, 457–469.

[23] Ganesh, A., O'Connell, N. and Wischick, D. (2004). *Big Queues*. Springer, Berlin.

[24] Gladstien, K. and Lange, K. (1978). Equilibrium distributions for deleterious genes in large stationary populations. *Theoret. Pop. Biol.* **14**, 322–328.

[25] Györfi, L. and Morvai, G. (2002). Queueing for ergodic arrivals and services. In *Limit Theorems in Probability and Statistics*, eds I. Berkes, E. Csáki, and M. Csörgö. Janos Bolyai Mathematical Society, Budapest, pp. 127–141.

[26] Györfi, L. and Walk, H. (1996). On the averaged stochastic approximation for linear regression. *SIAM J. Control Optimization* **34**, 31–61.

[27] Györfi, L., Ispány, M., Kevei, P. and Pap, Gy. (2015). Asymptotic behavior of a multi-type nearly critical Galton–Watson processes with immigration. *Prob. Theory Appl.* **59**, 590–610.

[28] Györfi, L., Ispány, M., Pap, Gy. and Varga, K. (2007). Poisson limit of an inhomogeneous nearly critical INAR(1) model, *Acta Sci. Math. (Szeged)* 73, 803–829.

[29] Haccou, P., Jagers, P. and Vatutin, V. A. (2007). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.

https://doi.org/10.1017/jpr.2024.70 Published online by Cambridge University Press

[30] HAIRER, M. (2005). Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Prob.* **32**, 703–758.
[31] IOSIFESCU, M. (2009). Iterated function sytems. A critical survey. *Math. Reports* **11**, 181–229.
[32] KAPLAN, N. (1973). The multitype Galton–Watson process with immigration. *Ann. Prob.* **1**, 947–953.
[33] KEVEI, P. AND WIANDT, P. (2021). Moments of the stationary distribution of subcritical multitype Galton–Watson processes with immigration. *Statist. Prob. Lett.* **173**, 109067.
[34] LANGE, K. AND FAN, R. (1997). Branching process models for mutant genes in nonstationary populations. *Theoret. Pop. Biol.* **51**, 118–133.
[35] LINDVALL, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.
[36] LOVAS, A. AND RÁSONYI, M. (2021). Markov chains in random environment with applications in queueing theory and machine learning. *Stoch. Process. Appl.* **137**, 294–326.
[37] LOYNES, R. M. (1962). The stability of a queue with non-independent inter-arrival and service times. *Proc. Camb. Phil. Soc*. 58, 497–520.
[38] MODE, C. J. (1971). *Multitype Branching Processes. Theory and Applications*. Elsevier, New York.
[39] POLYAK, B. T. (1990). New stochastic approximation type procedures. *Avtomat. Telemekh.* 51, 98–107 (in Russian). Translated in *Automat. Remote Control* 51, 937–946.
[40] PROPP, J. AND WILSON, D. (1996). Exact sampling with coupled Markov chains. *Random Structures Algorithms* **9**, 223–252.
[41] PROPP, J. AND WILSON, D. (1998). How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms* **27**, 170–217.
[42] QUINE, M. P. (1970). The multi-type Galton–Watson process with immigration. *J. Appl. Prob.* **7**, 411–422.
[43] QUINE, M. P. (1972). The multi-type Galton–Watson process with ϱ near to 1. *Adv. Appl. Prob.* **4**, 429–452.
[44] RESING, J. A. C. (1993). Polling systems and multitype branching processes. *Queueing Systems* **13**, 409–426.
[45] RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process. Tech. rep. 781, School of Operat. Res. and Ind. Eng., Cornell University, Ithaca, NY.
[46] RUPPERT, D. (1991). Stochastic approximation. In *Handbook of Sequential Analysis*, eds B. K. Ghosh and P. K. Sen. Marcel Dekker, New York, pp. 503–529.
[47] TRUQUET, L. (2021). Ergodic properties of some Markov chains models in random environments. Preprint, arXiv:2108.06211.
[48] VAN DER MEI, R. D. (2007). Towards a unifying theory on branching-type polling systems in heavy traffic. *Queueing Systems* **57**, 29–46.
[49] VARVENNE, M. (2019). Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory. *Bernoulli* **25**, 3234–3275.
[50] WELLING, M. AND TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proc. 28th Int. Conf. Machine Learning*, eds L. Getoor and T. Scheffer. ACM, New York, pp. 681–688.