Mathematical Control and Related Fields

doi:10.3934/mcrf.2023008

MOMENT STABILITY OF STOCHASTIC PROCESSES WITH APPLICATIONS TO CONTROL SYSTEMS

Arnab Ganguly*[1] and Debasish Chatterjee[2]

[1]Department of Mathematics, Louisiana State University, USA

[2]Systems & Control Engineering, Indian Institute of Technology Bombay
Mumbai 400076, India

(Communicated by Vivek Shripad Borkar)

Abstract. We establish new conditions for obtaining uniform bounds on the moments of discrete-time stochastic processes. Our results require a weak negative drift criterion along with a state-dependent restriction on the centered conditional moments of the process. They, in particular, generalize the main result of [22] which requires a constant bound on the averaged one-step jumps of the process. The state-dependent feature of our results make them suitable for a large class of multiplicative-noise processes. Under the additional assumption of the Markovian property, we prove new results on ergodicity that do not rely on a minorization condition typically needed in ergodic theorems. Several applications to iterative systems, control systems, and other dynamical systems with state-dependent multiplicative noise are included, and these illustrative examples demonstrate the wide applicability of our results.

1. **Introduction.** We study stability properties of a general class of discrete-time stochastic systems. Assessment of stability of dynamical systems is an important research area which has been studied extensively over the years. For example, in control theory a primary objective is to design suitable control policies which will ensure appropriate stability properties (e.g., bounded variance) of the underlying controlled system. There are various notions of stability of a system. In mathematics, stability often refers to equilibrium stability, which, for deterministic dynamical systems, is mainly concerned with qualitative behaviors of the trajectories of the system that start near the equilibrium point. For the stochastic counterpart, in Markovian setting it usually involves study of existence of invariant distributions and associated convergence and ergodic properties. A comprehensive source of results on different ergodicity properties for discrete-time Markov chains using

2020 *Mathematics Subject Classification.* Primary: 60J05, 60J20, 60F17, 93D05, 93E15; Secondary: 93E15.

Key words and phrases. Moment bounds, stability, ergodicity, Markov processes, control systems.

Research of A. Ganguly is supported in part by NSF DMS - 1855788 and Louisiana Board of Regents through the Board of Regents Support Fund (contract number: LEQSF(2016-19)-RD-A-04).

*Corresponding author: Arnab Ganguly.

Foster-Lyapunov functions is [18] (also see the references therein). Several extensions of such results have since then been explored quite extensively in the literature (for example, see [16, 17]). Another important book in this area is [9], which uses expected occupation measures of the chain for identifying conditions for stability.

The primary objective of the paper is to study moment stability, which concerns itself with uniform bounds on moments of a general stochastic process X_n or, more generally, on expectations of the form $\mathbb{E}(V(X_n))$ for a given function V. This is slightly different from the usual notions of stability in the Markovian setting as mentioned in the previous paragraph, but they are not unrelated. Indeed, if the process $\{X_n\}$ has certain underlying Lyapunov structure a strong form of Markovian stability holds which in particular implies moment stability. The result, which is based on Foster-Lyapunov criterion, can be described as follows. Given a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ taking values in a Polish space $\mathcal{S}$ with a transition probability kernel $\mathcal{P}$, suppose there exists a non-negative measurable function $u : \mathcal{S} \to [0, \infty)$, called a Foster-Lyapunov function, such that the process $\{u(X_n)\}_{n\in\mathbb{N}}$ possesses has the following negative drift condition: for some constant $b \geqslant 0,\ \theta > 0$, a set $A \subset \mathcal{S}$, and a function $V : \mathcal{S} \to [0, \infty)$

$$\mathbb{E}\left[u(X_{n+1}) - u(X_n)|X_n = x\right] \equiv \int_{\mathcal{S}} \mathcal{P}(x, dy)u(y) - u(x) \leqslant -\theta V(x) + b\mathbb{1}_{\{x\in A\}}. \quad (1)$$

If the set A is petite, (which, roughly speaking, are the sets that have the property that any set B is 'equally accessible' from any point inside the petite set - for definition and more details, see [15, 18]), the process $\{X_n\}$ has a unique invariant distribution π and also $\pi(V) = \int_{\mathcal{S}} \pi(dx)V(x) < \infty$. Moreover, under aperiodicity, it can be concluded that the chain is Harris ergodic, that is,

$$\|\mathcal{P}^n(x, \cdot) - \pi\|_V \to 0, \qquad \text{as } n \to \infty,$$

where $\|\cdot\|_V$ is the V-norm (see, the definition at the end of introduction) [18, Chapter 14]. In particular, one has $\mathbb{E}[V(X_n)] \to \pi(V)$ as $n \to \infty$ (which of course implies boundedness of $\mathbb{E}[V(X_n)]$). Thus for a Markov process $\{X_n\}$, one way to get a uniform bound on $V(X_n)$ is to find a Foster-Lyapunov function u such that (1) holds.

The objective of the first part of the paper is to explore scenarios where a strong negative drift condition like (1) does not hold or at least such a Lyapunov function is not easy to find for a specific V. We do note that the required conditions in our results are formulated in terms of the target function V itself. One pleasing aspect of this feature is that search for a suitable Lyapunov function u is not required for applying these results.

Our main result, Theorem 2.2, deals with the general regime where the state process $\{X_n\}$ is a general stochastic process and not necessarily Markovian. While past studies on stability mostly concern homogeneous Markov processes, the literature in the case of more general processes including non-homogeneous Markov processes and processes with long-range dependence, is rather sparse. One important work in this direction, which partially influenced our work and which we describe in more detail later in the Introduction and in Section 2, is [22].

The starting point in Theorem 2.2 is a weak negative drift like condition:

$$\mathbb{E}\left(V(X_{n+1}) - V(X_n)|\mathcal{F}_n\right) \leqslant -A, \quad X_n \notin \mathcal{D}, \quad (2)$$

which, if $\{X_n\}$ is a homogeneous Markov chain, is of course equivalent to $\mathcal{P}V(x) - V(x) \leqslant -A$ for x outside $\mathcal{D}$. As is well known (and may be observed by comparing (2) with (1)), even in the Markovian setting, the results of [18, Chapter 14] do not imply $\sup_n \mathbb{E}(V(X_n)) < \infty$. In fact, the condition (2) is not enough to guarantee such an assertion even in a deterministic setting. For example, consider the sequence $\{x_n\}$ on $\mathbb{N}$ defined by

$$x_{n+1} = \begin{cases} x_n - 1 & \text{if } x_n > 1, \\ n+1 & \text{if } x_n = 1. \end{cases}$$

Clearly, $\sup_{n\geqslant 1} x_n = \infty$ even though the negative drift condition is satisfied for $\mathcal{D} = \{1\}$. But we show in Theorem 2.2 that under a state-dependent restriction on the conditional moments of $V(X_{n+1})$ given $\mathcal{F}_n$ (see Assumption 2.1 for details), the desired uniform moment bound can be achieved. Note that the above sequence $\{x_{n+1}\}$ fails (2.1-c) of Assumption 2.1 but satisfies the other two conditions.

In the (homogeneous) Markovian framework, Theorem 2.2 leads to new results (c.f., Theorem 3.2 and its variant, Theorem 3.4) on Harris ergodicity of Markov chains which will be useful in situations where the Foster-Lyapunov drift criterion in the form of (1) cannot be verified. It is important to note that Theorem 3.2 does *not* require $\mathcal{D}$ to be petite or any minorization conditions or prior checking of aperiodicity of the chain.

As mentioned, Theorem 2.2 is influenced in part by a result of Pemantle and Rosenthal [22] which established a uniform bound on $\mathbb{E}(V^r(X_n))$ under (2) and the additional assumption of a *constant bound* on conditional p-th moment of one-step jumps of the process given $\mathcal{F}_n$, that is, $\mathbb{E}\left[|V(X_{n+1}) - V(X_n)|^p|\mathcal{F}_n\right]$. However, for a large class of stochastic systems the latter requirement of a uniform bound on conditional moments of jump sizes cannot be fulfilled and weaker conditions are needed. In particular, our work is motivated by some problems on stability of a class of stochastic systems with *multiplicative* noise where such conditions on one-step jumps are typically state-dependent and consequently fail to be bounded by a constant. This includes the general model considered in Section 4.1 where the result of [22] cannot be applied.

Our results generalize those of [22] in two important directions:

- we employ a different "metric" to control the one step jumps, and
- we permit such jumps to be bounded by a suitable *state dependent* function.

Specifically, instead of $\mathbb{E}\left[|V(X_{n+1}) - V(X_n)|^p|\mathcal{F}_n\right]$, we control the *centered* conditional p-th moment of $V(X_{n+1})$, that is, $\mathbb{E}\left[|V(X_{n+1}) - \mathbb{E}(V(X_{n+1})|\mathcal{F}_n)|^p \,\Big|\mathcal{F}_n\right]$, in a state-dependent fashion. The latter quantity can be viewed as a distance between the actual position $V(X_{n+1})$ at time $n+1$, and the expected position at that time given the past information, $\mathbb{E}(V(X_{n+1})|\mathcal{F}_n)$; in contrast, [22] uses the distance between actual positions at times $n+1$ and n. These extensions require a different approach involving different auxiliary estimates. The advantages of this new 'jump metric' and the state dependency feature have been discussed in detail after the the proof of Theorem 2.2. Together, they significantly expand the scope of stability results by encompassing a larger class of stochastic systems than hitherto possible.

This improvement in scope is amply demonstrated in Section 4, where a broad class of systems with multiplicative noise is studied and new stability results (see Proposition 4.2 and Corollary 4.4) are obtained. In particular, these results include stochastic switching systems and Markov processes of the form $X_{n+1} = H(X_n) + G(X_n)\xi_{n+1}$. The last part of this section is devoted to the important problem of stabilization of stochastic linear systems with bounded control inputs. The problem of interest here consists of finding conditions which guarantee L^2-boundedness of a stochastic linear system of the form $X_{n+1} = AX_n + Bu_n + \xi_{n+1}$ with bounded control actions. The particular problem has been studied in a previous work of the second author (see [25] and the references therein for more background on the problem), and it was demonstrated that when (A, B) is stabilizable, there exists a k-history dependent control policy that assures bounded variance of such a system provided the norm of the control is sufficiently large. This upper bound on the norm of the control appeared to be an artificial obstacle on its design, and it was conjectured in [25] that it is not required although a proof couldn't be provided. Here we show that this conjecture is indeed true (c.f. Proposition 4.7), and the artificial restriction on the control norm can be lifted largely owing to the new "metric" in Theorem 2.2. In fact, as Proposition 4.2 and Corollary 4.4 clearly indicate, this stabilization result can be easily extended to cover more general classes of stochastic control systems including ones with multiplicative noise.

The article is organized as follows. The main result on uniform moment bounds and related results are described in Section 2. Section 3 concerns itself with ergodicity of Markov chains. Section 4 discusses potential applications of our results for a large class of stochastic systems including switching systems, multiplicative Markov models, which are especially relevant to control theory.

Notation and terminology: For a probability kernel P on $\mathcal{S} \times \mathcal{S}$, and a function $f : \mathcal{S} \to [0, \infty)$, the function $Pf : \mathcal{S} \to [0, \infty)$ will be defined by $Pf(x) = \int_{\mathcal{S}} f(y)P(x, dy)$. In similar spirit, for a measure μ on $\mathcal{S}$, $\mu(f)$ will be defined by $\mu(f) = \int_{\mathcal{S}} f(x)\mu(dx)$. For a signed measure, μ, on $\mathcal{S}$. the corresponding total variation measure is denoted by $|\mu| = \mu^+ + \mu^-$, where $\mu = \mu^+ - \mu^-$ as per the Jordan decomposition. If $\mu = \nu_1 - \nu_2$, where ν_1 and ν_2 are probability measures, the total variation distance $\|\nu_1 - \nu_2\|_{TV}$ is given by

$$\|\nu_1 - \nu_2\|_{TV} = |\mu|(\mathcal{S}) = 2 \sup_{\mathcal{A} \in \mathcal{B}(\mathcal{S})} |\nu_1(\mathcal{A}) - \nu_2(\mathcal{A})|.$$

$\mathcal{B}(\mathcal{S})$ denotes the Borel sigma-algebra on $\mathcal{S}$. More generally, if $g : \mathcal{S} \to [0, \infty)$ is a measurable function, the *g-norm* of $\mu = \nu_1 - \nu_2$ is defined by $\|\mu\|_g = \sup\{|\mu(f)| : f \text{ measurable and } 0 \leqslant f \leqslant g\}$.

Throughout, we will work on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mathbb{E}$ will denote the expectation operator under $\mathbb{P}$. In context of the process $\{X_n\}$, $\mathbb{E}_x$ will denote the conditional expectation given $X_0 = x$.

2. **Uniform bounds for moments of stochastic processes.** The section presents our result on uniform bounds on functions of a general stochastic processes $\{X_n\}$ taking values in a topological space $\mathcal{S}$. The primary assumption involves a negative drift condition outside a set $\mathcal{D}$, together with a state-dependent control on the size of one-step jumps of $\{X_n\}$.

Assumption 2.1. *There exist measurable functions* $V : \mathcal{S} \to [0, \infty)$, $\varphi : \mathcal{S} \to [0, \infty)$, *and a measurable set* $\mathcal{D} \subset \mathcal{S}$ *such that*

(2.1-a) *for all* $n \in N$,
$$\mathbb{E}_{x_0}[V(X_{n+1}) - V(X_n) \mid \mathcal{F}_n] \leqslant -A \quad \text{on} \quad \{X_n \notin \mathcal{D}\};$$

(2.1-b) *for each* $n \in \mathbb{N}$ *and some* $p > 2$, Ξ_n, *the centered conditional* p*-th moment of* $V(X_{n+1})$ *given* $\mathcal{F}_n$, *satisfies*
$$\Xi_n \doteq \mathbb{E}_{x_0}\left[|V(X_{n+1}) - \mathbb{E}(V(X_{n+1})|\mathcal{F}_n)|^p \Big| \mathcal{F}_n\right] \leqslant \varphi(X_n),$$
where $\varphi(x) \leqslant \mathscr{C}_\varphi(1 + V^s(x))$ *for some* $0 \leqslant s < p/2 - 1$ *and some constant* $\mathscr{C}_\varphi > 0$.

(2.1-c) $\sup_{x \in \mathcal{D}} V(x) < \infty$, *and for some constant* $\bar{\mathscr{B}}^p_{\mathcal{D}}$ *(which could also depend on the starting point* x_0*)*,
$$\mathbb{E}_{x_0}\left[(\mathbb{E}[V(X_{n+1})|\mathcal{F}_n])^p \mathbb{1}_{\{X_n \in \mathcal{D}\}}\right] < \bar{\mathscr{B}}^p_{\mathcal{D}}.$$

Theorem 2.2. *Suppose that Assumption 2.1 holds for the process* $\{X_n\}_{n \geq 0}$ *with* $X_0 = x_0$. *Then*
$$\mathscr{B}_{V,r}(x_0) \doteq \sup_{n \in \mathbb{N}} \mathbb{E}_{x_0}\left[V(X_n)^r\right] < \infty$$
for any $0 \leqslant r < \varsigma(s, p)$, *where*
$$\varsigma(s,p) = \begin{cases} p\left(1 - \frac{s}{p-2}\right) - 1 & \text{for } s \in [0, (p-2)^2/2p) \cup [1 - 2/p,\ p/2 - 1), \text{ when } 2 < p < 4; \\ & \text{for all } s \in [0, p/2 - 1), \text{ when } p \geqslant 4; \\ p - 2 & \text{for } (p-2)^2/2p \leqslant s < 1 - 2/p, \text{ when } 2 < p < 4. \end{cases}$$

Remark 2.3. Note that (2.1-c) is implied by the simpler condition:$\mathbb{E}_{x_0}[V(X_{n+1})|\mathcal{F}_n] \leqslant \bar{\mathscr{B}}_{\mathcal{D}}$ on $\{X_n \in \mathcal{D}\}$ for some constant $\bar{\mathscr{B}}_D$.

Remark 2.4 (Comparative analysis)**.** At this stage it is instructive to compare Theorem 2.2 with [22, Theorem 1] and precisely note some of the improvements the former offers.

- The first significant extension is that Theorem 2.2 allows the jump sizes in (2.1-b) to be state dependent; observe that [22, Theorem 1] requires
$$\mathbb{E}_{x_0}\left[|V(X_{n+1}) - V(X_n)|^p|\mathcal{F}_n\right] \leqslant B, \tag{†}$$
for some constant $B > 0$. The resulting benefits are obvious since the dependence on the states allows the result in particular to be applicable to large class of multiplicative systems of the form
$$X_{n+1} = H(X_n) + G(X_n)\xi_{n+1},$$
which [22, Theorem 1] cannot cover in general.
- The second notable distinction is in the 'metric' used in (2.1-b) in controlling jump sizes: while [22, Theorem 1] involves $\mathbb{E}_{x_0}\left[|V(X_{n+1}) - V(X_n)|^p|\mathcal{F}_n\right]$, our result only requires the centered conditional p-th moments of $V(X_{n+1})$ given $\mathcal{F}_n$, namely, $\mathbb{E}_{x_0}\left[\left|V(X_{n+1}) - \mathbb{E}[V(X_{n+1})|\mathcal{F}_n]\right|^p \Big| \mathcal{F}_n\right]$, to be controlled (in a state-dependent fashion). Of course, the latter leads to weaker hypothesis since
$$\mathbb{E}_{x_0}\left[\left|V(X_{n+1}) - \mathbb{E}[V(X_{n+1})|\mathcal{F}_n]\right|^p \Big| \mathcal{F}_n\right] \leqslant 2^p \mathbb{E}_{x_0}\left[|V(X_{n+1}) - V(X_n)|^p|\mathcal{F}_n\right].$$
It is important to emphasize the advantage of our weaker hypothesis because the condition in (†) precludes it from being applicable to some additive models. To

illustrate this with a simple example, consider a $[0,\infty)$-valued process $\{X_n\}$ given by

$$X_{n+1} = X_n/2 + \xi_{n+1}, \quad X_0 \geqslant 0,$$

where ξ_n are $[0,\infty)$-valued random variables with $\mu_p = \sup_n \mathbb{E}(\xi_n^p) < \infty$ for $p > 2$. Since $X_{n+1} - X_n = -X_n/2 + \xi_{n+1}$, clearly the negative drift condition (c.f (2.1-a)) holds with $V(x) = |x|$. But for the jump sizes we can only have

$$\mathbb{E}_{x_0}\left[|X_{n+1} - X_n|^p | \mathcal{F}_n\right] = O(X_n^p).$$

This means that [22, Theorem 1] cannot be used to get $\sup_n \mathbb{E}_x(X_n) < \infty$ for this simple additive system — a fact which easily follows from an elementary iteration argument (note, $\mathbb{E}_x(X_n) \xrightarrow[n\to\infty]{} 2\mu_1$). On the other hand, Theorem 2.2 clearly covers such cases as

$$\mathbb{E}_{x_0}\left[|X_{n+1} - E\left(X_{n+1}|\mathcal{F}_n\right)|^p \middle| \mathcal{F}_n \right] \leqslant \bar{\mu}_p, \quad \bar{\mu}_p = \sup_n \mathbb{E}|\xi_n - \mathbb{E}(\xi_n)|^p.$$

- It should actually be noted that had Theorem 2.2 simply controlled the jump sizes by imposing the more restrictive condition $\mathbb{E}\left[|X_{n+1} - X_n|^p | \mathcal{F}_n\right] \leqslant \varphi(X_n)$, the state-dependency feature would not be enough to salvage the moment bound of the above additive system (because of the requirement $\varphi(x) = O(V^s(x))$ for $s < p/2 - 1$). It is interesting to note that the results of [18] based on Foster-Lyapunov drift conditions also cannot directly be used in this simple example since $\{X_n\}$ is not necessarily Markov because the ξ_n are not assumed to be i.i.d. To summarize, the weaker jump metric coupled with the state dependency feature makes Theorem 2.2 a rather powerful tool in understanding stability for a broad class of stochastic systems. Some important results in this direction for switching systems have been discussed in the application section.

2.1. **Proof of Theorem 2.2 and its preparation.** The proof of Theorem 2.2, which is presented at the end of this section, combines Proposition 2.6 and Proposition 2.8. Proposition 2.6 first establishes a weaker version of the above assertion by showing that $\sup_{n\in\mathbb{N}} \mathbb{E}_{x_0}\left[V(X_n)^r\right] < \infty$ for all $r < p/2 - 1$. However, an extension of the result from there to all $r < \varsigma(s,p)$ (notice that $\varsigma(s,p) \geqslant p/2 - 1$) requires a substantial amount of extra work and is achieved through Proposition 2.8.

The following lemma will be used in various necessary estimates.

Lemma 2.5. *Let M_n be a martingale relative to the filtration $\{\mathcal{F}_n\}$,*

$$\gamma_n \doteq \mathbb{E}\left[|M_{n+1} - M_n|^p \;\middle|\, \mathcal{F}_n\right], \quad n \geqslant 0 \tag{3}$$

Θ a non-negative random variable, and $b > 0$ a constant. Then for any $p \geqslant 1$, $0 \leqslant r \leqslant p$

$$\mathbb{E}\left[(|M_n - M_k| + \Theta)^r \mathbb{1}_{\{|M_n - M_k| + \Theta) \geqslant b\}} | \mathcal{F}_k\right] \leqslant 2^{p-1}\left(\bar{c}_p (n-k)^{\frac{p}{2}-1} \sum_{m=k}^{n-1} \mathbb{E}\left[\gamma_m|\mathcal{F}_k\right] + \mathbb{E}\left[|\Theta|^p|\mathcal{F}_k\right]\right) b^{r-p},$$

where $\bar{c}_p$ is the upper Burkholder's constant.

Proof. Note that by the discrete version of Burkholder-Davis-Gundy inequality (e.g., see [24]), there exists $\bar{c}_p > 0$ such that

$$\mathbb{E}\big[|M_n - M_k|^p \bigm| \mathcal{F}_k\big] \leqslant \bar{c}_p \mathbb{E}\left[\left(\sum_{m=k}^{n-1}\big[|M_{m+1}-M_m|^2\big]\right)^{p/2} \Bigg| \mathcal{F}_k\right].$$

Hölder's inequality and (3) yields

$$\begin{aligned}\mathbb{E}\big[|M_n - M_k|^p \bigm| \mathcal{F}_k\big] &\leqslant \bar{c}_p (n-k)^{\frac{p}{2}-1}\sum_{m=k}^{n-1}\mathbb{E}\left[|M_{m+1}-M_m|^p\,|\mathcal{F}_k\right]\\ &\leqslant \bar{c}_p (n-k)^{\frac{p}{2}-1}\sum_{m=k}^{n-1}\mathbb{E}\left[\gamma_m|\mathcal{F}_k\right]. \qquad (4)\end{aligned}$$

Observe that for a random variable Y_n, we have for $0 \leqslant r \leqslant p$,

$$\mathbb{E}\left[|Y_n|^r \mathbb{1}_{\{|Y_n|\geqslant b\}}|\mathcal{F}_k\right] \leqslant \mathbb{E}\left[|Y_n|^p/|Y_n|^{p-r}\mathbb{1}_{\{|Y_n|\geqslant b\}}|\mathcal{F}_k\right] \;\leqslant\; \mathbb{E}\left[|Y_n|^p|\mathcal{F}_k\right]/b^{p-r}.$$

Taking $Y_n = |M_n - M_k| + \Theta$ (with $n \geqslant k$) we get

$$\begin{aligned}&\mathbb{E}\left[|M_n - M_k| + \Theta|^r \mathbb{1}_{\{||M_n-M_k|+\Theta|\geqslant b\}}|\mathcal{F}_k\right]\\ &\qquad\leqslant 2^{p-1}\left(\mathbb{E}\left[|M_n-M_k|^p|\mathcal{F}_k\right] + \mathbb{E}\left[\Theta^p|\mathcal{F}_k\right]\right)/b^{p-r},\end{aligned}$$

and the assertion follows from (4). ◻

We now prove the two propositions which form the backbone of our main result, Theorem 2.2.

Proposition 2.6. *Suppose that Assumption 2.1 holds. Then for any $0 \leqslant r < p/2-1$,*

$$\mathscr{B}_{V,r}(x_0) \doteq \sup_{n\in\mathbb{N}} \mathbb{E}_{x_0}\big[V(X_n)^r\big] < \infty,$$

Proof. Observe that it is enough to prove the result for $r \in (s, p/2-1)$, and we pick such an r. Writing $\varphi(x) = \varphi(x)\mathbb{1}_{\{|V(x)|\leqslant M\}} + (\varphi(x)/V^r(x))V^r(x)\mathbb{1}_{\{|V(x)|>M\}}$, in view of the growth assumption on φ (c.f (2.1-b)), we conclude that for every $\varepsilon > 0$, there exists a constant $\mathscr{C}_1(\varepsilon)$ such that $\varphi(x) \leqslant \mathscr{C}_1(\varepsilon) + \varepsilon V^r(x)$. It is not difficult to see from (2.1-b) that one can take $\mathscr{C}_1(\varepsilon) = (2\mathscr{C}_\varphi/\varepsilon)^{r/s-1}$ (although it should be noted that this is not the best possible value of $\mathscr{C}_1(\varepsilon)$).

Define $\mathscr{M}_0 = 0$ and

$$\mathscr{M}_n = \sum_{j=0}^{n-1} V(X_{j+1}) - \mathbb{E}_{x_0}[V(X_{j+1})|\mathcal{F}_j], \quad n \geqslant 1.$$

Then $\mathscr{M}_n$ is a martingale. Fix $N \in \mathbb{N}$, and define the last time $\{X_k\}$ is in $\mathcal{D}$ by

$$\eta \equiv \max\{k \leqslant N \mid X_k \in \mathcal{D}\}.$$

Notice that $\{\eta = k\} = \{X_k \in \mathcal{D}\} \cap \bigcap_{j>k}^N \{X_j \notin \mathcal{D}\}$. On $\{\eta = k\}$, for $k < n \leqslant N$ we ahve

$$\begin{aligned}\mathscr{M}_n - \mathscr{M}_k =& V(X_n) - V(X_k) - \sum_{j=k}^{n-1}\left(\mathbb{E}_{x_0}[V(X_{j+1})|\mathcal{F}_j] - V(X_j)\right)\\ &\geqslant V(X_n) - V(X_k) - \left(\mathbb{E}_{x_0}[V(X_{k+1})|\mathcal{F}_k] - V(X_k)\right) + A(n-k-1)\end{aligned}$$

$$\equiv V(X_n) + A(n-k-1) - \mathbb{E}_{x_0}[V(X_{k+1})|\mathcal{F}_k]. \tag{5}$$

It follows that on $\{\eta = k\}$,

$$V(X_N)^r \leqslant (|\mathscr{M}_N - \mathscr{M}_k| + \xi_k)^r \quad \text{and} \quad A(N-k-1) \leqslant |\mathscr{M}_N - \mathscr{M}_k| + \xi_k,$$

where $\xi_k = \mathbb{E}_{x_0}[V(X_{k+1})|\mathcal{F}_k]\mathbb{1}_{\{X_k \in \mathcal{D}\}}$.

On $\{\eta = -\infty\}$, which corresponds to the case that the process starting outside $\mathcal{D}$ never enters $\mathcal{D}$ by time N, we have

$$V(X_N)^r \leqslant (|\mathscr{M}_N - \mathscr{M}_0| + V(x_0))^r \quad \text{and} \quad AN \leqslant |\mathscr{M}_N - \mathscr{M}_0| + V(x_0).$$

Thus for $k \leqslant N-2$,

$$\begin{aligned}
\mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=k\}}] &\leqslant \mathbb{E}_{x_0}\left[(|\mathscr{M}_N - \mathscr{M}_k| + \xi_k)^r 1_{\{\eta=k\}}\right] \\
&\leqslant \mathbb{E}_{x_0}\left[(|\mathscr{M}_N - \mathscr{M}_k| + \xi_k)^r 1_{\{|\mathscr{M}_N - \mathscr{M}_k| + \xi_k \geqslant A(N-k-1)\}}\right] \\
&\leqslant 2^{p-1}\left(\bar{c}_p (N-k)^{\frac{p}{2}-1} \sum_{m=k}^{N-1} \mathbb{E}_{x_0}\left[\varphi(X_m)\right] + \mathbb{E}_x\left[\xi_k^p\right]\right) / (N-k-1)^{p-r} \\
&\leqslant 2^{p-1}\left(\bar{c}_p (N-k)^{r-1-p/2} \sum_{m=k}^{N-1} \mathbb{E}_{x_0}\left[\varphi(X_m)\right] + \bar{\mathscr{B}}_{\mathcal{D}}^p (N-k)^{r-p}\right),
\end{aligned}$$

where we used (2.1-c) and Lemma 2.5 along with the observation that

$$\mathbb{E}_{x_0}\left[|M_{n+1} - M_n|^p\right] = \mathbb{E}_{x_0}\left[|V(X_{n+1}) - \mathbb{E}[V(X_{n+1})|\mathcal{F}_n]|^p\right] \leqslant \mathbb{E}_{x_0}[\varphi(X_n)].$$

Similarly, on $\{\eta = -\infty\}$,

$$\begin{aligned}
\mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=-\infty\}}] &\leqslant \mathbb{E}_{x_0}\left[(|\mathscr{M}_N| + V(x_0))^r 1_{\{|\mathscr{M}_N| + V(x_0) \geqslant AN\}}\right] \\
&\leqslant 2^{p-1}\left(\bar{c}_p N^{\frac{p}{2}-1} \sum_{m=0}^{N-1} \mathbb{E}_{x_0}\left[\varphi(X_m)\right] + V(x_0)^p\right) / N^{p-r} \\
&\leqslant 2^{p-1}\left(\bar{c}_p N^{r-1-\frac{p}{2}} \sum_{m=0}^{N-1} \mathbb{E}_{x_0}\left[\varphi(X_m)\right] + V(x_0)^p N^{r-p}\right).
\end{aligned}$$

Next, note that because of (2.1-b)

$$\mathbb{E}_{x_0}[V^p(X_N)|\mathcal{F}_{N-1}]\mathbb{1}_{\{X_{N-1} \in \mathcal{D}\}} \leqslant 2^{p-1}\left(\mathbb{E}_{x_0}[V(X_N)|\mathcal{F}_{N-1}]^p \mathbb{1}_{\{X_{N-1} \in \mathcal{D}\}} + \sup_{x \in \mathcal{D}} \varphi(x)\right),$$

which by (2.1-c) of course implies that for any $q \leqslant p$,

$$\mathbb{E}_{x_0}[V(X_N)^q 1_{\{\eta=N-1\}}] \leqslant \mathbb{E}_{x_0}[V^q(X_N)\mathbb{1}_{\{X_{N-1} \in \mathcal{D}\}}] \leqslant \mathscr{C}_0^{q,p} \equiv \left(2^{p-1}\left(\bar{\mathscr{B}}_{\mathcal{D}}^p + \sup_{x \in \mathcal{D}} \varphi(x)\right)\right)^{q/p}.$$

Lastly,

$$\mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=N\}}] \leqslant \mathbb{E}_{x_0}[V(X_N)^r 1_{\{X_N \in \mathcal{D}\}}] \leqslant \sup_{z \in \mathcal{D}} V^r(z),$$

Thus,

$$\begin{aligned}
\mathbb{E}_{x_0}[V(X_N)^r] &\leqslant \sum_{k=0}^{N} \mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=k\}}] + \mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=-\infty\}}] \\
&\leqslant \sum_{k=0}^{N-2} \mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=k\}}] + \mathscr{C}_0^{q,p} + \sup_{z \in \mathcal{D}} V^r(z) + \mathbb{E}_{x_0}[V(X_N)^r 1_{\{\eta=-\infty\}}] \\
&\leqslant 2^{p-1}\bar{c}_p \sum_{k=0}^{N-2} (N-k)^{r-1-p/2} \sum_{m=k}^{N-1} \mathbb{E}_{x_0}\left[\varphi(X_m)\right] + 2^{p-1}\bar{\mathscr{B}}_{\mathcal{D}}^p \sum_{k=0}^{N-2} (N-k)^{r-p}
\end{aligned}$$

$$
\begin{aligned}
&+ \mathscr{C}_0^{q,p} + \sup_{z\in\mathcal{D}} V^r(z) + 2^{p-1}\bar{c}_p N^{r-1-p/2}\sum_{m=0}^{N-1}\mathbb{E}_{x_0}\left[\varphi(X_m)\right] \\
&+ 2^{p-1}V^p(x_0)N^{r-p} \\
\leqslant\ & \mathscr{C}_0^{q,p} + \sup_{z\in\mathcal{D}} V^r(z) + 2^{p-1}\big(\bar{\mathscr{B}}_{\mathcal{D}}^p + V^p(x_0)\zeta(p-r)\big) \\
&+ 2^p\bar{c}_p\sum_{k=0}^{N-2}(N-k)^{r-1-p/2}\sum_{m=k}^{N-1}\mathbb{E}_{x_0}\left[\varphi(X_m)\right] \\
\leqslant\ & \mathscr{C}_0^{q,p} + \sup_{z\in\mathcal{D}} V^r(z) + 2^{p-1}\big(\bar{\mathscr{B}}_{\mathcal{D}}^p + V^p(x_0)\zeta(p-r)\big) \\
&+ 2^p\bar{c}_p\sum_{k=0}^{N-2}(N-k)^{r-1-p/2}\sum_{m=k}^{N-1}\mathbb{E}_{x_0}\left[\mathscr{C}_1(\varepsilon)+\varepsilon V^r(X_m)\right] \\
\leqslant\ & \mathscr{C}_2(\varepsilon,x_0) + 2^p\bar{c}_p\varepsilon\sum_{m=0}^{N-1}\beta_m^N\mathbb{E}_{x_0}\left[V^r(X_m)\right],
\end{aligned}
$$

where, with ζ denoting the Riemann-zeta function,

$$
\begin{aligned}
\mathscr{C}_2(\varepsilon,x_0) \equiv\ & \mathscr{C}_0^{q,p} + \sup_{z\in\mathcal{D}} V^r(z) + 2^{p-1}(\bar{\mathscr{B}}_{\mathcal{D}}^p + V^p(x_0)\zeta(p-r) + 2^p\bar{c}_p\mathscr{C}_1(\varepsilon)\zeta(p/2-r) \\
=& \left(2^{p-1}\left(\bar{\mathscr{B}}_{\mathcal{D}}^p + \sup_{x\in\mathcal{D}}\varphi(x)\right)\right)^{q/p} + \sup_{z\in\mathcal{D}} V^r(z) + 2^{p-1}\big(\bar{\mathscr{B}}_{\mathcal{D}}^p + V^p(x_0)\zeta(p-r)\big) \\
&+ 2^p\bar{c}_p\left(2\mathscr{C}_\varphi/\varepsilon\right)^{r/s-1}\zeta(p/2-r).
\end{aligned} \tag{6}
$$

The choice of ε appearing above will be specified shortly, and $\beta_m^N = \sum_{k=0}^m (N-k)^{r-1-p/2}$. Iterating the inequality leads to

$$
\begin{aligned}
\mathbb{E}_{x_0}\left[V(X_N)^r\right] \leqslant \mathscr{C}_2(\varepsilon,x_0)\Bigg(& 1 + 2^p\bar{c}_p\varepsilon\sum_{l_1=0}^{N-1}\beta_{l_1}^N + (2^p\bar{c}_p\varepsilon)^2\sum_{l_1=0}^{N-1}\beta_{l_1}^N\sum_{l_2=0}^{l_1-1}\beta_{l_2}^{l_1} + \cdots \\
&\cdots + (2^p\bar{c}_p\varepsilon)^{N-1}\beta_{N-1}^N\beta_{N-2}^{N-1}\ldots\beta_1^2\beta_0^1\Bigg)\left(V^r(x_0)\vee 1\right).
\end{aligned} \tag{7}
$$

Notice that for any $k>0$, since $r<p/2-1$,

$$
\sum_{l=0}^{k-1}\beta_l^k = \sum_{l=0}^{k-1}\sum_{j=0}^{l}(k-j)^{r-1-p/2} = \sum_{j=0}^{k-1}(k-j)^{r-p/2} \leqslant \zeta(p/2-r).
$$

Choosing ε so that $2^p\bar{c}_p\varepsilon\zeta(p/2-r)<1$, (7) yields

$$
\mathbb{E}_{x_0}\left[V(X_N)^r\right] \leqslant \frac{\mathscr{C}_2(\varepsilon,x_0)\left(V^r(x_0)\vee 1\right)}{1-2^p\bar{c}_p\varepsilon\zeta(p/2-r)},
$$

with $\mathscr{C}_2(\varepsilon,x_0)$ given by (6). This concludes our proof. □

Remark 2.7. Note that the last display gives an explicit form of an upper bound for $\mathscr{B}_{V,r}(x_0)$, which, in particular, shows how it depends on the various quantities of Assumption 2.1. However, this bound may not be optimal (or near optimal) for specific models.

The next proposition helps to extend the above result from any $r<p/2-1$ to $\varsigma(s,p)$ as stipulated in Theorem 2.2. However it is also a stand-alone result that is applicable to certain models where Theorem 2.2 is not directly applicable. These

are cases where one directly does not have any good estimate of the conditional centered moment Ξ_n as required in Theorem 2.8, but have suitable upper bounds for its $\|\cdot\|_\theta$ norm. As a simple example, let X_n be a stochastic process taking values in $[-\mathfrak{c}_0, \infty)$, whose temporal evolution is given by

$$X_{n+1} = \mathfrak{c}_1 + X_n/2 + Y_n$$

where $\mathfrak{c}_0$ and $\mathfrak{c}_1$ are (real-valued) constants, and $\{Y_n\}$ is an $\mathcal{F}_n$-adapted martingale difference process (that is, $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = 0$) and $\sup_n \mathbb{E}(|Y_n|^p) < \infty$ for $p > 2$. Then Theorem 2.2 is not applicable, but the following proposition can be applied with $\theta = 1$ to $V(x) = x + \mathfrak{c}_0$.

Proposition 2.8. *Let $\Xi_n \equiv \mathbb{E}_{x_0}\left[|V(X_{n+1}) - \mathbb{E}(V(X_{n+1})|\mathcal{F}_n)|^p|\mathcal{F}_n\right]$ denote the centered conditional p-th moment of $V(X_{n+1})$ given $\mathcal{F}_n$. Assume that (2.1-a) and (2.1-c) of Assumption 2.1 hold, and for $p > 2$, some $\theta \in [1, \infty]$, and some constant $0 < \tilde{\mathscr{B}}_\theta(x_0) < \infty$,*

$$\|\Xi_n\|_\theta = \mathbb{E}_{x_0}\left[\Xi_n^\theta\right]^{1/\theta} \leqslant \tilde{\mathscr{B}}_\theta(x_0), \qquad \textit{for all } n \geqslant 0.$$

Then $\mathscr{B}_{V,r}(x_0) \doteq \sup_{n\in\mathbb{N}} \mathbb{E}_{x_0}\left[V(X_n)^r\right] < \infty$ for $0 \leqslant r < \varsigma(\theta, p)$, where

$$\varsigma(\theta,p) = \begin{cases} p\left(1 - \frac{1}{2\theta}\right) - 1 & \text{for } \theta \in \left[1, \frac{p}{2}\right] \cup \left(\frac{p}{p-2}, \infty\right] \text{ when } 2 < p < 4, \\ & \text{for any } \ \theta \geqslant 1 \text{ when } p > 4; \\ p - 2 & \text{for } \theta \in \left(\frac{p}{2}, \frac{p}{p-2}\right] \text{ when } 2 < p < 4. \end{cases}$$

Here $\theta = \infty$ corresponds to the case that $\Xi_n = \mathbb{E}_{x_0}\left[|V(X_{n+1}) - \mathbb{E}(V(X_{n+1})|\mathcal{F}_n)|^p|\mathcal{F}_n\right] \leqslant \bar{\mathscr{B}}$ a.s., for some constant $\bar{\mathscr{B}} > 0$.

Proof of Proposition 2.8. The constants appearing in various estimates below (besides the ones that appeared before) will be denoted by $\hat{C}_i$'s. They will not depend on n but may depend on the parameters of the system and the initial position x_0.

Define $\mathscr{M}_n$, η and ξ_k as in the proof of Proposition 2.6. Fix N, $0 \leqslant k \leqslant N$, and define $\tau \equiv \tau(N, k)$ by

$$\tau = \inf\{j \geqslant k : \mathscr{M}_j - \mathscr{M}_k + \xi_k \geqslant A(N - k - 1)/2\}.$$

Clearly, $\tau \leqslant N$. For $j > k$, notice that on $\{\tau = j\}$,

$$\mathscr{M}_{j-1} - \mathscr{M}_k + \xi_k \leqslant A(N - k - 1)/2,$$

and hence on $\{\eta = k\} \cap \{\tau = j\}$

$$\mathscr{M}_N - \mathscr{M}_{j-1} \geqslant A(N - k - 1)/2 + V(X_N).$$

It follows that for $j > k$

$$\begin{aligned} &\mathbb{E}_{x_0}[V(X_N)\mathbb{1}_{\{\eta=k\}}\mathbb{1}_{\{\tau=j\}}] \\ &\leqslant \mathbb{E}_{x_0}\left[|\mathscr{M}_N - \mathscr{M}_{j-1}|^r \mathbb{1}_{\{|\mathscr{M}_N - \mathscr{M}_{j-1}| > A(N-k-1)/2\}}\mathbb{1}_{\{|\mathscr{M}_{j-1} - \mathscr{M}_k| + \xi_k > A(j-k-2)\vee 0\}}\mathbb{1}_{\{\tau=j\}}\right]. \end{aligned}$$

Notice that $\mathcal{S}_j \equiv \mathbb{E}_{x_0}\left[|\mathscr{M}_N - \mathscr{M}_{j-1}|^r \mathbb{1}_{|\mathscr{M}_N - \mathscr{M}_{j-1}| > A(N-k-1)/2}|\mathcal{F}_j\right]$ can be estimated by Lemma 2.5 as

$$\begin{aligned} \mathcal{S}_j &\leqslant (2/A(N-k-1))^{p-r}\mathbb{E}_{x_0}\left[|\mathscr{M}_N - \mathscr{M}_{j-1}|^p|\mathcal{F}_j\right] \\ &\leqslant 2^{p-1}\left[\mathbb{E}_{x_0}\left[|\mathscr{M}_N - \mathscr{M}_j|^p|\mathcal{F}_j\right] + |\mathscr{M}_j - \mathscr{M}_{j-1}|^p\right]/(N-k-1)^{p-r} \end{aligned}$$

$$\leqslant 2^{p-1}\left[\bar{c}_p(N-j)^{\frac{p}{2}-1}\mathbb{E}_{x_0}\left[\sum_{l=j}^{N-1}\Xi_l|\mathcal{F}_j\right]+|\mathscr{M}_j-\mathscr{M}_{j-1}|^p\right]/(N-k-1)^{p-r}.$$

Also, for $\tau=k$ by Lemma 2.5,

$$\begin{aligned}\mathbb{E}_{x_0}[V(X_N)\mathbb{1}_{\{\eta=k\}}\mathbb{1}_{\{\tau=k\}}] &\leqslant \mathbb{E}_{x_0}\left[\mathbb{1}_{\{\tau=k\}}\mathbb{E}_{x_0}\left[(|\mathscr{M}_N-\mathscr{M}_k|+\xi_k)^r\mathbb{1}_{||\mathscr{M}_N-\mathscr{M}_k|+\xi_k>A(N-k-1)}|\mathcal{F}_k\right]\right]\\ &\leqslant \hat{C}_1\mathbb{E}_{x_0}\left[\mathbb{1}_{\{\tau=k\}}\left((N-k-1)^{r-p/2-1}\sum_{l=k}^{N-1}\mathbb{E}_{x_0}\left[\Xi_l|\mathcal{F}_k\right]+(N-k-1)^{r-p}|\xi_k|^p\right)\right].\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}_{x_0}[V(X_N)\mathbb{1}_{\{\eta=k\}}] &= \sum_{j=k}^{N}\mathbb{E}_{x_0}[V(X_N)\mathbb{1}_{\{\eta=k\}}\mathbb{1}_{\{\tau=j\}}]\\ &\leqslant \hat{C}_1(N-k-1)^{r-p/2-1}\mathbb{E}_{x_0}\left[\mathbb{1}_{\{\tau=k\}}\sum_{l=k}^{N-1}\mathbb{E}_{x_0}\left[\Xi_l|\mathcal{F}_k\right]\right]\\ &\quad+\hat{C}_1(N-k-1)^{r-p}\bar{\mathscr{B}}_{\mathcal{D}}+\sum_{j=k+1}^{N}\mathbb{E}_{x_0}\left[\mathbb{1}_{\{\tau=j\}}\mathbb{1}_{\{|\mathscr{M}_{j-1}-\mathscr{M}_k|+\xi_k>A(j-k-2)\}}\mathcal{S}_j\right]\\ &\leqslant \hat{C}_2\Big[(N-k-1)^{r-\frac{p}{2}-1}\sum_{j=k}^{N}\sum_{l=j}^{N-1}\mathbb{E}_{x_0}\left[\Xi_l\mathbb{1}_{\{\tau=j\}}\right]+(N-k-1)^{p-r}\\ &\quad+(N-k-1)^{r-p}\sum_{j=k+1}^{N}|\mathscr{M}_j-\mathscr{M}_{j-1}|^p\mathbb{1}_{\{|\mathscr{M}_{j-1}-\mathscr{M}_k|+\xi_k>A(j-k-2)\vee 0\}}\Big].\end{aligned}\tag{8}$$

We next estimate the above two terms separately, and for that we need the following bound:

$$\begin{aligned}\mathbb{P}_{x_0}\left(\max_{k\leqslant j\leqslant l}|\mathscr{M}_j-\mathscr{M}_k|+\xi_k>\Upsilon\right) &\leqslant \mathbb{E}_{x_0}\left[\max_{k\leqslant j\leqslant l}|\mathscr{M}_j-\mathscr{M}_k|+\xi_k\right]^p/\Upsilon^p\\ &\leqslant 2^{p-1}\left(\left(\frac{p}{p-1}\right)^p\mathbb{E}_{x_0}\left[|\mathscr{M}_l-\mathscr{M}_k|^p\right]+\mathbb{E}_{x_0}|\xi_k|^p\right)/\Upsilon^p\\ &\leqslant 2^{p-1}\left(\left(\frac{p}{p-1}\right)^p\bar{c}_p(l-k)^{p/2-1}\sum_{j=k}^{l}\mathbb{E}_{x_0}[\Xi_j]+\bar{\mathscr{B}}^p_{\mathcal{D}}\right)/\Upsilon^p\\ &\leqslant \hat{C}_3\left((l-k)^{p/2}+1\right)/\Upsilon^p,\end{aligned}\tag{9}$$

where the second step used Doob's maximal inequality, the third employed the Burkholder-Davis-Gundy inequality and Hölder's inequalities (c.f. proof of Lemma 2.5), and the final step used the fact that $\mathbb{E}[\xi_l]\leqslant\|\xi_l\|_\theta\leqslant\tilde{B}_\theta(x_0)$. Now notice that

$$\begin{aligned}\sum_{j=k}^{N-1}\sum_{l=j}^{N-1}\mathbb{E}_{x_0}\left[\Xi_l\mathbb{1}_{\{\tau=j\}}\right] &= \sum_{l=k}^{N-1}\sum_{j=k}^{l}\mathbb{E}_{x_0}\left[\Xi_l\mathbb{1}_{\{\tau=j\}}\right]\\ &= \sum_{l=k}^{N-1}\mathbb{E}_{x_0}\left[\Xi_l\mathbb{1}_{\{\tau\leqslant l\}}\right]\\ &\leqslant \sum_{l=k}^{N-1}\|\Xi_l\|_\theta\mathbb{P}_{x_0}(\tau\leqslant l)^{1/\theta^*}\\ &\leqslant \tilde{\mathscr{B}}_\theta(x_0)\sum_{l=k}^{N-1}\mathbb{P}_{x_0}\left(\max_{k\leqslant j\leqslant l}(|\mathscr{M}_j-\mathscr{M}_k|+\xi_k)>A(N-k-1)/2\right)^{1/\theta^*}\\ &\leqslant \left(\frac{2}{A}\right)^{p/\theta^*}\tilde{\mathscr{B}}_\theta(x_0)\sum_{l=k}^{N-1}\left(\frac{\hat{C}_3\left((l-k)^{p/2}+1\right)}{(N-k-1)^p}\right)^{1/\theta^*}\\ &\leqslant \hat{C}_4(N-k-1)^{-\frac{p}{2\theta^*}+1}=\hat{C}_4(N-k-1)^{-\frac{p}{2}\left(1-\frac{1}{\theta}\right)+1},\end{aligned}\tag{10}$$

where the third inequality employed (9), and θ^* is the Hölder conjugate of θ (i.e., $\frac{1}{\theta}+\frac{1}{\theta^*}=1$).

Next notice that the term

$$\mathcal{A} \equiv \sum_{j=k+1}^{N} \mathbb{E}_{x_0}\left[|\mathscr{M}_j - \mathscr{M}_{j-1}|^p \mathbb{1}_{\{|\mathscr{M}_{j-1}-\mathscr{M}_k|+\xi_k > A(j-k-2)\}}\right]$$

can be estimated as

$$\begin{aligned}
\mathcal{A} &\leqslant \|\Xi_{k+1}\|_\theta + \|\Xi_{k+2}\|_\theta + \sum_{j=k+3}^{N} \mathbb{E}_{x_0}\left[\Xi_{j-1}\mathbb{1}_{\{|\mathscr{M}_{j-1}-\mathscr{M}_k|+\xi_k > A(j-k-2)\}}\right] \\
&\leqslant 2\tilde{\mathscr{B}}_\theta(x_0) + \sum_{j=k+3}^{N} \|\Xi_{j-1}\|_\theta \mathbb{P}_{x_0}\left(|\mathscr{M}_{j-1} - \mathscr{M}_k| + \xi_k > A(j-k-2)\right)^{1/\theta^*} \\
&\leqslant \tilde{\mathscr{B}}_\theta(x_0)\left[2 + A^{-p/\theta^*} \sum_{j=k+3}^{N} \left(\hat{C}_3\left((j-k-1)^{p/2}+1\right)/(j-k-2)^p\right)^{1/\theta^*}\right], \\
&\leqslant \hat{C}_5\left[1 + \sum_{j=k+3}^{N} 1/(j-k-2)^{p/2\theta^*}\right] \\
&\leqslant \begin{cases} \hat{C}_6 \equiv \hat{C}_5\left(1+\zeta(p/2\theta^*)\right) & \text{if } p/2\theta^* = p(1-1/\theta)/2 > 1, \\ \hat{C}_7(N-k-1) & \text{otherwise,} \end{cases}
\end{aligned} \tag{11}$$

where the third inequality is by (9) (recall that ζ denotes the Riemann zeta function). We now consider some cases.

Case 1: $\theta \leqslant p/2$: Suppose that $r < p\left(1 - \frac{1}{2\theta}\right) - 1$. Notice that this range of r implies that $p - r - 1 > \frac{p}{2\theta} \geqslant 1$. It follows from (8), (10), and the second case of (11) that

$$\begin{aligned}
\mathbb{E}_{x_0}[V^r(X_N)] &= \sum_{k=0}^{N} \mathbb{E}_{x_0}[V^r(X_N)\mathbb{1}_{\{\eta=k\}}] \\
&\leqslant \sum_{k=0}^{N-2} \mathbb{E}_{x_0}[V^r(X_N)\mathbb{1}_{\{\eta=k\}}] + \mathbb{E}_{x_0}[V^r(X_N)\mathbb{1}_{\{X_{N-1}\in\mathcal{D}\}}] + \sup_{x\in\mathcal{D}} V^r(x) \\
&\leqslant \hat{C}_8\Big[\sum_{k=0}^{N-2}(N-k-1)^{r-\frac{p}{2}-1-\frac{p}{2}\left(1-\frac{1}{\theta}\right)+1} + \sum_{k=0}^{N-2}(N-k-1)^{r-p} \\
&\qquad + \sum_{k=0}^{N-2}(N-k-1)^{r-p+1}\Big] + \mathscr{C}_3 + \sup_{x\in\mathcal{D}} V^r(x) \\
&= \hat{C}_8\left(\zeta\left(p\left(1-\frac{1}{2\theta}\right) - r\right) + \zeta(p-r) + \zeta(p-r-1)\right) + \mathscr{C}_3 + \sup_{x\in\mathcal{D}} V^r(x).
\end{aligned}$$

Case 2: $\theta > p/2$ and $p \geqslant 4$: Suppose that $r < p\left(1 - \frac{1}{2\theta}\right) - 1$. Notice that $\theta > p/2$ and $p \geqslant 4$ imply that $p/2\theta^* = p(1-1/\theta)/2 > 1$. Arguing along the lines of the previous case, it follows from (8), (10), and the first case of (11) that

$$\begin{aligned}
\mathbb{E}_{x_0}[V^r(X_N)] &\leqslant \sum_{k=0}^{N-2} \mathbb{E}_{x_0}[V^r(X_N)\mathbb{1}_{\{\eta=k\}}] + \mathbb{E}_{x_0}[V^r(X_N)\mathbb{1}_{\{X_{N-1}\in\mathcal{D}\}}] + \sup_{x\in\mathcal{D}} V^r(x) \\
&\leqslant \hat{C}_9\left(\sum_{k=0}^{N-2}(N-k)^{r-p\left(1-\frac{1}{2\theta}\right)} + \sum_{k=0}^{N-1}(N-k-1)^{r-p}\right) + \mathscr{C}_3 + \sup_{x\in\mathcal{D}} V^r(x)
\end{aligned}$$

$$\leqslant \hat{C}_{10}.$$

The other cases in the assertion follow similarly once we observe that $\theta > p/(p-2) \Leftrightarrow p/2\theta^* > 1$ and for $2 < p < 4$, $p/2 < p/(p-2)$. ☐

Proof of Theorem 2.2. From Proposition 2.6 and the growth assumption on φ, it follows that for any $1 \leqslant \theta < (p-2)/2s$,

$$\sup_n \|\Xi_n\|_\theta \leqslant \sup_n \left(\mathbb{E}_{x_0}(\varphi^\theta(X_n))\right)^{1/\theta} < \infty,$$

where $\|\cdot\|_\theta$ is the $\mathcal{L}^\theta(\Omega, \mathbb{P})$-norm (c.f. Proposition 2.8). The assertion now follows from Proposition 2.8 by letting $\theta \uparrow (p-2)/2s$. If $s = 0$, that is, $\Xi_n \leqslant \mathscr{C}$, for some constant $\mathscr{C}$ a.s., we take $\theta = \infty$ in Proposition 2.8. ☐

3. **Ergodicity of Markov processes.** In this section we present new results on Harris ergodicity of Markov processes which employ Theorem 2.2 at their core. It is important to point out that no minorization condition is assumed here; see [20] for a recent brief overview of ergodicity of Markov chains. Moreover, unlike some other results that stipulate locally compact state spaces, thus precluding them from being applied to (infinite-dimensional) Banach space-valued processes, the state space $\mathcal{E}$ here is simply assumed to be Polish. The results can, therefore, be potentially applied to discretized stochastic PDE models.

The first result assumes the existence of a transition density $q(\cdot,\cdot)$ of $\{X_n\}$ (with respect to some measure μ) with a generous growth condition on it. Note that if $\mathcal{E}$ is countable, then $q(x, y) = \mathbb{P}(X_{n+1} = y|X_n = x) \leqslant 1$ and thus automatically satisfies (3.2-d). The second result is a variant of the first which replaces the assumption of existence of density by a suitable continuity condition on the transition probability measures.

Definition 3.1. A function $V : \mathcal{E} \to [0, \infty)$ is *inf-compact* if the level sets, $\mathcal{K}_m = \{x : V(x) \leqslant m\}$ are compact for all $m \geqslant 0$.

Note that an inf-compact function V is lower-semicontinuous.

Theorem 3.2. *Let $\{X_n\}$ be a Markov process taking values in a Polish space $\mathcal{E}$ with transition kernel $\mathcal{P}$. Suppose that for an inf-compact function $V : \mathcal{E} \to [0, \infty)$, the following conditions hold:*

(3.2-a) *for all $n \in N$,*

$$\mathcal{P}V(x) - V(x) \leqslant -A, \quad \text{on} \quad \{x \notin \mathcal{D}\};$$

(3.2-b) *for some $p > 2$,*

$$\mathcal{P}|V(\cdot) - \mathcal{P}V(x)|^p(x) = \int |V(y) - \mathcal{P}V(x)|^p \mathcal{P}(x, dy) \leqslant \varphi(x),$$

where $\varphi : \mathcal{E} \to [0, \infty]$ satisfies $\varphi(x) \leqslant \mathscr{C}_\varphi(1 + V^s(x))$ for some $s < p/2 - 1$ and some constant $\mathscr{C}_\varphi > 0$. This is of course same as requiring $\mathbb{E}\left[\left|V(X_{n+1}) - \mathcal{P}V(X_n)\right|^p \middle| \mathcal{F}_n\right] \leqslant \varphi(X_n)$.

(3.2-c) $\sup_{x\in\mathcal{D}} V(x) < \infty$, *and* $\sup_{x\in\mathcal{D}} \mathcal{P}V(x) < \infty$,

Also, suppose that

(3.2-d) *$\mathcal{P}$ is weak Feller, ψ-irreducible, and admits a density q with respect to some Radon measure μ, that is, $\mathcal{P}(x,dy) = q(x,y)\mu(dy)$, and that for every compact set $\mathcal{K}$, there exist a constant $\mathfrak{c}_{\mathcal{K},0}$ and an exponent $0 < \tilde{r} \equiv \tilde{r}_{\mathcal{K}} < \varsigma(s,p)$ such that*

$$\sup_{y\in\mathcal{K}} q(x,y) \leqslant \mathfrak{c}_{\mathcal{K},0}\left(1 + V^{\tilde{r}}(x)\right).$$

Then

(i) *Under (3.2-a)–(3.2-c), $\sup_n \mathbb{E}_{x_0}(V^r(X_n)) \equiv \sup_n \mathcal{P}^n V^r(x_0) < \infty$ for any $0 \leqslant r < \varsigma(s,p)$, where $\varsigma(s,p)$ is as in Theorem 2.2.*

(ii) *Under the additional assumption of (3.2-d), $\{X_n\}$ is positive Harris recurrent (PHR) and aperiodic with a unique invariant distribution π, and for any x_0 and $\tilde{r} \in (0, \varsigma(s,p))$*

$$\int (V^r + 1)d|\mathcal{P}^n(x_0,\cdot) - \pi| \to 0 \quad \text{as } n \to \infty; \tag{12}$$

or equivalently,

$$\|\mathcal{P}^n(x_0,\cdot) - \pi\|_{V^r+1} \doteq \sup_{f:|f|\leqslant V^r+1} |\mathcal{P}^n f(x_0) - \pi(f)| \to 0 \quad \text{as } n \to \infty. \tag{13}$$

Proof. (i) follows from the Theorem 2.2. Since V is inf-compact, it follows from (i) that for every x_0, $\{\mathcal{P}^n(x_0,\cdot)\}$ is tight, and let π be one of its limit points. Since $\mathcal{P}$ is weak Feller, by the Krylov-Bogolyubov theorem [23, Theorem 7.1], π is invariant for $\mathcal{P}$, and uniqueness of π follows from the assumption of ψ-irreducibility [9, Proposition 4.2.2]. Hence, for every x_0, $\mathcal{P}^n(x_0,\cdot) \Rightarrow \pi$ (along the full sequence) as $n \to \infty$.

For (ii) we start by establishing the following claim.

Claim 1: Suppose that $f \leqslant V^r + 1$ for some $r \in (0, \varsigma(s,p))$. Then $\mathcal{P}^n f(x_0) \to \pi(f)$ as $n \to \infty$ for any $x_0 \in \mathcal{E}$.

Since V is lower semi-continuous we have by (generalized) Fatou's lemma,

$$\pi(V^r) \leqslant \liminf_{n\to\infty} \mathcal{P}^n V^r(x_0) \leqslant \mathscr{B}_{V,r}(x_0)$$

for any $r \in (0, \varsigma(s,p))$. Now let $f \leqslant V^r + 1$ for some $r \in (0, \varsigma(s,p))$ and fix $\varepsilon > 0$.

Since $\{\mathcal{P}^n(x_0,\cdot)\}$ is tight, for a given $\tilde{\varepsilon} > 0$, there exists a compact set $\mathcal{K}$ (which depends on x_0 and which we take of the form $\mathcal{K}_m = \{x : V(x) \leqslant m\}$ for sufficiently large m) such that

$$\sup_n \mathcal{P}^n(x_0, \mathcal{K}^c) \leqslant \tilde{\varepsilon} \quad \text{and} \quad \pi(\mathcal{K}^c) \leqslant \tilde{\varepsilon}.$$

Now by Hölder's inequality

$$\begin{aligned}\mathcal{P}^n f 1_{\mathcal{K}^c}(x_0) = \int f(y)1_{\mathcal{K}^c}(y)\mathcal{P}^n(x_0,dy) &\leqslant \int (V^r(y)+1)1_{\mathcal{K}^c}(y)\mathcal{P}^n(x_0,dy)\\ &\leqslant \left(\int V^{r'}(y)\mathcal{P}^n(x_0,dy)\right)^{r/r'} \left(\int \mathbb{1}_{\mathcal{K}^c}(y)\mathcal{P}^n(x_0,dy)\right)^{1-r/r'} + \mathcal{P}^n(x_0,\mathcal{K}^c),\\ &\leqslant \mathscr{B}^{r/r'}_{V,r'}(x_0)\tilde{\varepsilon}^{1-r/r'} + \tilde{\varepsilon} \end{aligned}\tag{14}$$

for some $r < r' < \varsigma(s,p)$. Similarly, $\pi(f\mathbb{1}_{\mathcal{K}^c}) \leqslant \mathscr{B}^{r/r'}_{V,r'}(x)\tilde{\varepsilon}^{1-r/r'} + \tilde{\varepsilon}$.

Since $f\mathbb{1}_{\mathcal{K}} \in L^1(\mu)$, there exist $\{h_m\} \subset C_b(\mathcal{E}, \mathbb{R})$ such that $h_m \to f1_{\mathcal{K}}$ in $L^1(\mu)$ as $m \to \infty$, and $\sup_x |h_m(x)| \leqslant \sup_{x\in\mathcal{K}} |f(x)|$ for $m \geqslant 1$. In fact, since $\operatorname{supp}(f\mathbb{1}_{\mathcal{K}}) \subset \mathcal{K}$, we can choose $\{h_m\}$ such that $\operatorname{supp}(h_m) \subset \mathcal{K}' \supset \mathcal{K}$ for some compact set $\mathcal{K}'$.

Observe that for $x \in \mathcal{E}$, $y \in \mathcal{K}'$ and $n \geqslant 1$,

$$\begin{aligned} q^n(x,y) = &\int q^{n-1}(x,z)q(z,y)d\mu(z) \leqslant \int q^{n-1}(x,z)\mathfrak{c}_{\mathcal{K}',0}\left(1+V^{\tilde{r}}(z)\right)d\mu(z) \\ &\leqslant \mathfrak{c}_{\mathcal{K}',0}\left(1+\mathbb{E}_x(V^{\tilde{r}}(X_{n-1}))\right) \leqslant \mathfrak{c}_{\mathcal{K}',0}\left(1+\mathscr{B}_{V,\tilde{r}}(x)\right) \equiv \mathscr{C}_{\mathcal{K}'}(x). \end{aligned}$$

Hence

$$\begin{aligned} \sup_n |\mathcal{P}^n f\mathbb{1}_{\mathcal{K}}(x_0) - \mathcal{P}^n h_m| &\leqslant \int_{\mathcal{K}'} |f(y)\mathbb{1}_{\mathcal{K}}(y) - h_m(y)| q^n(x_0,y)d\mu(y) \\ &\leqslant \mathscr{C}_{\mathcal{K}'}(x_0)\|f1_{\mathcal{K}} - h_m\|_1. \end{aligned} \tag{15}$$

Next, notice that π is absolutely continuous with μ. Indeed, if $\mu(\mathcal{A}) = 0$, then $\mathcal{P}(x,\mathcal{A}) = 0$, and hence $\pi(\mathcal{A}) = \int \pi(dx)\mathcal{P}(x,\mathcal{A}) = 0$. Let $g = d\pi/d\mu$. For any $M > 0$,

$$\begin{aligned} |\pi(h_m) - \pi(f1_{\mathcal{K}})| &\leqslant M\int |h_m - f\mathbb{1}_{\mathcal{K}}|\mathbb{1}_{\{g\leqslant M\}}d\mu + \int |h_m - f\mathbb{1}_{\mathcal{K}}|g\mathbb{1}_{\{g\geqslant M\}}d\mu \\ &\leqslant M\|h_m - f\mathbb{1}_{\mathcal{K}}\|_1 + 2\sup_{x\in\mathcal{K}}|f(x)|\int g\mathbb{1}_{\{g\geqslant M\}}d\mu. \end{aligned} \tag{16}$$

Write

$$\begin{aligned} \mathcal{P}^n f(x_0) - \pi(f) = &\left(\mathcal{P}^n f1_{\mathcal{K}}(x_0) - \mathcal{P}^n h_m(x_0)\right) + \left(\mathcal{P}^n h_m(x_0) - \pi(h_m)\right) \\ &+ \left(\pi(h_m) - \pi(f1_{\mathcal{K}}(x_0))\right) + \mathcal{P}^n f1_{\mathcal{K}^c}(x_0) - \pi(f1_{\mathcal{K}^c}(x_0)), \end{aligned} \tag{17}$$

and choose $\mathcal{K}$ such that (14) holds for $\tilde{\varepsilon}$ where $\tilde{\varepsilon}$ is chosen such that $\mathscr{B}_{V,r'}^{r/r'}(x_0)\tilde{\varepsilon}^{1-r/r'} + \tilde{\varepsilon} \leqslant \varepsilon/10$. Since $\int g d\mu = 1$, choose sufficiently large M such that $\int g\mathbb{1}_{\{g\geqslant M\}}d\mu \leqslant \varepsilon/(20\sup_{x\in\mathcal{K}}|f(x)|)$, then a sufficiently large m such that

$$\|f1_{\mathcal{K}} - h_m\|_1 \leqslant (\varepsilon/5\mathscr{C}_{\mathcal{K}'}(x_0)) \wedge (\varepsilon/10M).$$

Finally, since $\mathcal{P}^n(x_0,\cdot) \Rightarrow \pi$, and $h_m \in C_c(\mathcal{E},\mathbb{R})$, we have $(\mathcal{P}^n h_m(x_0) - \pi(h_m)) \to 0$ as $n \to \infty$. Hence, we can choose a sufficiently large n such that $|\mathcal{P}^n h_m(x_0) - \pi(h_m)| \leqslant \varepsilon/5$, and thus from (14), (15), (16) and (17),

$$|\mathcal{P}^n f(x_0) - \pi(f)| \leqslant \varepsilon.$$

This proves Claim 1, which, in particular, asserts that for any $x_0 \in \mathcal{E}$ and any Borel set $\mathcal{A}$, $\mathcal{P}^n(x,\mathcal{A}) \xrightarrow[n\to\infty]{} \pi(\mathcal{A})$.

Claim 2: For any $x_0 \in \mathcal{E}$, $\|\mathcal{P}^n(x_0,\cdot) - \pi\|_{TV} = 2\sup_{\mathcal{A}\in\mathcal{B}(\mathcal{E})}|\mathcal{P}^n(x_0,\mathcal{A}) - \pi(\mathcal{A})| \xrightarrow[n\to\infty]{} 0$.

To see this we first prove that for μ-a.a $z \in \mathcal{E}$, $q^n(x_0,z) \to g(z) \equiv (d\pi/d\mu)(z)$. Since g is the stationary density of $\{X_n\}$, for μ-a.a $z \in \mathcal{E}$, $g(z) = \int g(x)q(x,z)d\mu(x)$. For each $z \in \mathcal{E}$, define the function f_z by $f_z(x) = q(x,z)$ and notice that because of condition (3.2-d), the function f_z satisfies the hypothesis of *Claim 1*. Consequently, for μ-a.a z, as $n \to \infty$

$$q^n(x_0,z) = \int q(x,z)\mathcal{P}^{n-1}(x_0,dx) = \mathcal{P}^{n-1}f_z(x_0)$$

$$\xrightarrow[n\to\infty]{} \pi(f_z) = \int q(x,z)d\pi(x) = \int q(x,z)g(x)d\mu(x) = g(z).$$

By Scheffe's lemma, $\int |q^n(x_0,z) - g(z)|d\mu(z) \to 0$, which is equivalent to the assertion in the claim (since $\|\mathcal{P}^n(x_0,\cdot) - \pi\|_{TV} = \int |q^n(x_0,z) - g(z)|d\mu(z)$).

Finally, by Hölder's inequality for some $r' \in (r, \varsigma(s,p))$

$$\begin{aligned}
&\int (V^r(y)+1)d|\mathcal{P}^n(x,\cdot) - \pi|(y) \\
&\leqslant \left(\int V^{r'}(y)(\mathcal{P}^n(x,dy) + \pi(dy))\right)^{r/r'} \|\mathcal{P}^n(x,\cdot) - \pi\|_{TV}^{1-r/r'} + \|\mathcal{P}^n(x,\cdot) - \pi\|_{TV} \\
&\qquad \leqslant 2\mathscr{B}_{V,r'}(x)^{r/r'}\|\mathcal{P}^n(x,\cdot) - \pi\|_{TV}^{1-r/r'} + \|\mathcal{P}^n(x,\cdot) - \pi\|_{TV} \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

The equivalence of (12) and (13) follows from Lemma A.1. □

Remark 3.3. Claim 2 also follows from [9, Theorem 4.3.4] (also see [8]), which states that when $\{X_n\}$ has a unique invariant distribution, π, setwise convergence of $\mathcal{P}^n(x_0,\cdot)$ to π for every x_0 is equivalent of the convergence of the former to the latter in the total variation norm. This in turn is equivalent to $\{X_n\}$ being aperiodic and PHR. The equivalence of the setwise convergence of $\mathcal{P}^n(x,\cdot)$ and convergence in total-variation norm is a unique feature of PHR chains. In the context of Theorem 3.2, a direct proof of this assertion appears to be more illuminating than invoking [9, Theorem 4.3.4], whose proof uses different techniques. However, we employ [9, Theorem 4.3.4] in the following variant of Theorem 3.2 which does not require the existence of the transition density $q(x,y)$.

Theorem 3.4. *Let $\{X_n\}$ be a Markov chain as in Theorem 3.2 satisfying (3.2-a)–(3.2-c). Suppose further that $\{X_n\}$ is weak Feller, ψ-irreducible, and that for each $\mathcal{A} \in \mathcal{B}(\mathcal{E})$, there exists $n_0 \equiv n_0(\mathcal{A})$ such that the mapping $x \in \mathcal{E} \to \mathcal{P}^{n_0}(x,\mathcal{A})$ is continuous. Then the assertion of Theorem 3.2 holds.*

Proof. We observe from the proof of Theorem 3.2 that $\{X_n\}$ has a unique invariant distribution π, and for every x_0, $\mathcal{P}^n(x_0,\cdot) \Rightarrow \pi$ as $n \to \infty$ and $\pi(V^r) \vee \sup_n \mathcal{P}^n V^r(x_0) \leqslant \mathscr{B}_{V,r}(x_0) < \infty$ for $0 < r < \varsigma(s,p)$. We now claim that $\mathcal{P}^n(x_0,\mathcal{A}) \to \pi(\mathcal{A})$ for any $\mathcal{A} \in \mathcal{B}(\mathcal{E})$. This simply follows from the weak convergence of $\mathcal{P}^n(x_0,\cdot)$ to π and the fact that the mapping $y \longmapsto \mathcal{P}^{n_0}(y,\mathcal{A})$ is continuous (by our hypothesis) and bounded. Indeed for $n \geqslant n_0$,

$$\mathcal{P}^n(x_0,\mathcal{A}) = \int \mathcal{P}^{n_0}(y,\mathcal{A})\mathcal{P}^{n-n_0}(x_0,dy) \xrightarrow[n\to\infty]{} \int \mathcal{P}^{n_0}(y,\mathcal{A})\pi(dy) = \pi(\mathcal{A}).$$

where the last equality holds because π is the invariant distribution of $\{X_n\}$. Consequently, by [9, Theorem 4.3.4] (see Remark 3.3 above), $\|\mathcal{P}^n(x_0,\cdot) - \pi\|_{TV} \xrightarrow[n\to\infty]{} 0$. The rest of the proof follows the same arguments as that of Theorem 3.2. □

Remark 3.5. The continuity assumption in Theorem 3.4 is, of course, implied by the stronger condition that for each $\mathcal{A} \in \mathcal{B}(\mathcal{E})$, the one-step transition probability $\mathcal{P}(\cdot,\mathcal{A}) : \mathcal{E} \to [0,1]$ is continuous. The latter condition is equivalent to the strong Feller property of $\{X_n\}$.

4. **Applications.** This sections is devoted to establishing stability of a broad class of multiplicative systems through various applications of the results established in the prequel.

4.1. **Discrete time switching systems.** Let $\mathbb{H}$ be a Hilbert space and $\mathcal{E}$ a Polish space. Suppose there exists a sequence of measurable maps $P_n : \mathbb{H} \times \mathcal{E} \times \mathcal{E} \to [0, 1]$ such that for each $x \in \mathbb{H}$, the function $P_n(x, \cdot, \cdot)$ is a transition probability kernel. Consider a discrete-time $\mathcal{F}_n$-adapted process $\{Z_n\} \equiv \{(X_n, Y_n)\}$ taking values in $\mathbb{H} \times \mathcal{E}$, whose dynamics is defined by the following rule: given the state $(X_n, Y_n) = (x_n, y_n)$,

(SS-1) first, Y_{n+1} is selected randomly according to the (possibly) time-inhomogenous transition probability distribution $P_n(x_n, y_n, \cdot) \equiv P_{n,x_n}(y_n, \cdot)$,

(SS-2) next given $Y_{n+1} = y_{n+1}$,
$$X_{n+1} = H_n(x_n, y_{n+1}, \xi_{n+1}),$$
where $\{\xi_k : k = 1, \dots\}$ is a sequence of independent random variables taking values in a Banach space $\mathbb{B}$, ξ_{n+1} is independent of $\sigma\{\mathcal{F}_n, Y_{n+1}\}$ and $H_n : \mathbb{H} \times \mathcal{E} \times \mathbb{B} \to \mathbb{H}$.

In general $\{(X_n, Y_n)\}$ is a (possibly) time-inhomogeneous Markov process but clearly, neither $\{X_n\}$ nor $\{Y_n\}$ is Markovian on its own. The stochastic system $\{(X_n, Y_n)\}$ is known as a *discrete-time switching system* or a *stochastic hybrid system* (and sometimes also known as *iterated function system with place dependent probabilities* [1]). Stochastic hybrid systems are extensively used to model practical phenomena where system parameters are subject to sudden changes. These systems have found widespread applications in various disciplines including the synthesis of fractals and the modeling of biological networks [12], target tracking [19], communication networks [10], and control theory [3, 2, 4], to name a few. There is a considerable literature addressing classical weak stability questions concerning the existence and uniqueness of invariant measures of iterated function systems, see e.g., [21, 13, 26, 5, 11] and the references therein. Comprehensive sources studying various properties of these systems including results on stability in both continuous and discrete time can be found in [14, 29] (and also the references therein). In most of these works, $\{Y_n\}$ is often assumed to be a stand-alone finite or countable state-space Markov chain.

We consider a broad class of coupled switching or hybrid systems whose dynamics is described by (SS-1) and (SS-2) with H_n of the form
$$H_n(x, y, z) = L_n(x, y) + F_n(x, y) + G_n(x, y, z),$$
where $L_n, F_n : \mathbb{H} \times \mathcal{E} \to \mathbb{H}$ and $G_n : \mathbb{H} \times \mathcal{E} \times \mathbb{B} \to \mathbb{H}$. In other words, $\{X_n\}$ satisfies
$$X_{n+1} = L_n(X_n, Y_{n+1}) + F_n(X_n, Y_{n+1}) + G_n(X_n, Y_{n+1}, \xi_{n+1}) \tag{18}$$
where the ξ_n are $\mathbb{B}$-valued random variables. Systems of the form (18), of course, subsume multiplicative systems of the type
$$X_{n+1} = X_n + F_n(X_n, Y_{n+1}) + G^0_n(X_n, Y_{n+1})\xi_{n+1}.$$

We will make the following assumptions on the system (18):

Assumption 4.1.

(SS-3) *For* $\|x\| > B$ *and any* $y \in \mathcal{E}$,

$$P_{n,x}\langle F_n(x,\cdot), L_n(x,\cdot)\rangle(y) = \int \langle F_n(x,y'), L_n(x,y')\rangle P_{n,x}(y,dy') \leqslant -\mathfrak{m}_0\|x\|^{-(1+\gamma)},$$

for some constants $\mathfrak{m}_0$ *and exponent* $\gamma \geqslant 0$.

(SS-4) *The following growth conditions hold:*

- $\|L_n(x,y)\| \leqslant \mathfrak{m}_{L,1}(y)\|x\| + \mathfrak{m}_{L,2}(y)$ *and* $\|\bar{L}_n(x,y)\| \leqslant \mathfrak{m}_{\bar{L}}(y)(1+\|x\|)^{l_1}$, *where* $\bar{L}_n(x,y) = L_n(x,y) - P_{n,x}L_n(x,\cdot)(y)$.
- $\|F_n(x,y)\| \leqslant \mathfrak{m}_F(y)(1+\|x\|)^{f_0}$, $\bar{F}_n(x,y) \leqslant \mathfrak{m}_{\bar{F}}(y)(1+\|x\|)^{f_1}$, *and* $\|G_n(x,y,z)\| \leqslant \mathfrak{m}_G(y)(1+\|x\|)^{g_0}\Psi(z)$, *where* $\Psi : \mathbb{B} \to [0,\infty)$ *and* $\bar{F}_n(x,y) = F_n(x,y) - P_{n,x}F_n(x,\cdot)(y)$.
- *For any* $p > 0$, *the constants* $\bar{\mathfrak{m}}_{F,p}, \bar{\mathfrak{m}}_{\bar{F},p}, \bar{\mathfrak{m}}_{G,p}, \bar{\mathfrak{m}}_{L,1,p}, \bar{\mathfrak{m}}_{L,2,p}$ *and* $\bar{\mathfrak{m}}_{\bar{L},p}$ *are finite, and* $\bar{\mathfrak{m}}_{L,1,2} \leqslant 1$, *where the preceding constants are defined by*

$$\bar{\mathfrak{m}}_{\chi,p} \doteq \sup_{n,x,z} \int \mathfrak{m}^p_\chi(y)P_{n,x}(z,dy), \quad \chi = F, \bar{F}, G, \{L,1\}, \{L,2\}, \bar{L}. \tag{19}$$

(SS-5) *The exponents satisfy:*

- *(a)* $f_0 < (1+\gamma)/2$ *or (b)* $f_0 = (1+\gamma)/2$ *and* $\bar{\mathfrak{m}}_{F,2} \leqslant 2\mathfrak{m}_0$;
- $g_0 < \gamma \wedge 1/2$, *and* $l_1 \vee f_1 < 1/2$.

(SS-6) *The* ξ_n *are independent* $\mathbb{B}$*-valued random variables with distribution* ν_n*; for each* n, ξ_{n+1} *is independent of* $\sigma\{\mathcal{F}_n, Y_{n+1}\}$, *and for any* $p > 0$, $m^p_* = \sup_n \mathbb{E}(\Psi(\xi_n)^p) < \infty$

Proposition 4.2. *Under Assumption 4.1 we have* $\sup_n \mathbb{E}_{x_0}\|X_n\|^m < \infty$ *for any* $m > 0$ *and* $x_0 \in \mathbb{H}$. *If the functions* G_n *are centered with respect to the variable* z *in the sense that* $\hat{G}_n(x,y) \doteq \int_{\mathbb{B}} G_n(x,y,z)\nu_{n+1}(dz) = 0$ *for all* $n \geqslant 1$, $x \in \mathbb{H}$ *and* $y \in \mathcal{E}$, *then we only need* $g_0 < 1/2$ *instead of* $g_0 < \gamma \wedge 1/2$ *in (SS-5) for the above assertion to be true.*

Remark 4.3. Several comments are in order at this stage:

- Due to the growth assumption on G_n in (SS-4) and the condition (SS-6), for each n, x and y, the function $z \to G_n(x,y,z)$ is Bochner integrable, and hence $\hat{G}_n(x,y) \doteq \int_{\mathbb{B}} G_n(x,y,z)\nu_{n+1}(dz)$ is well defined (the integral is defined in Bochner sense).
- One scenario where the functions G_n are centered (with respect to the variable z) occurs when considering multiplicative stochastic system driven by zero-mean random variables. Specifically, in such models the G_n are of the form $G_n(x,y,z) = G^0_n(x,y)z$ and the ξ_n are mean zero-random variables. Also notice for these models, $\Psi(z) = \|z\|_{\mathbb{B}}$.
- Suppose that the G_n are not centered in the variable z. If $\gamma < 1/2$, (SS-5) requires that the growth exponent of G_n, $g_0 < \gamma$. However, this could be extended to the boundary case of $g_0 = \gamma$ (when $\gamma < 1/2$) provided the averaged growth constants $\bar{\mathfrak{m}}_{\chi,p}$ (c.f. (19)) meet certain conditions. If $g_0 = \gamma$ and $f_0 < (1+\gamma)/2$, then the assertion of Proposition 4.2 is true provided $\left(\bar{\mathfrak{m}}_{G,2}m^2_*\right)^{1/2} < \mathfrak{m}_0$. If $g_0 = \gamma$ and $f_0 = (1+\gamma)/2$, then the same assertion holds provided $\left(\bar{\mathfrak{m}}_{G,2}m^2_*\right)^{1/2} + \bar{\mathfrak{m}}_{F,2}/2 < \mathfrak{m}_0$.

- Condition (SS-3) is implied by the simpler condition:

$$\langle F_n(x,y), L_n(x,y)\rangle \leqslant -\mathfrak{m}_0\|x\|^{1+\gamma} \quad \text{for all } \|x\| > B \text{ and for all } y.$$

  Similarly, for many models a stronger (but easier to check) form of the condition (SS-4), where the 'constants' $\mathfrak{m}_\chi$ (for $\chi = F, \bar{F}, G, \{L,1\}, \{L,2\}, \bar{L}$) do not depend on y, suffices. In that case the corresponding averaged constants (given by (19)) are of course given by $\bar{\mathfrak{m}}_{\chi,p} = \mathfrak{m}_\chi^p$, and are therefore trivially finite.
- One common example of L_n is $L_n(x,y) \equiv L_n(x) = x$ or $U_n x$ for some unitary operator U_n. If $L_n(x,y) \equiv L_n(x)$, then centered L_n, that is, $\bar{L}_n \equiv 0$, and the condition on the corresponding growth exponent l_1 is trivially satisfied.
- Clearly, $f_1 \leqslant f_0$, where, recall that, f_1 and f_0 are the growth rates of $\bar{F}_n(x,y) = F_n(x,y) - P_x F_n(x,\cdot)(y)$ (centered F_n) and F_n, respectively. In some models, without any other information or suitable estimates on $\bar{F}$, f_1 may just have to be taken the same as f_0, in which case condition (SS-5) implies that the above result on uniform bounds on moments applies to systems for which $f_0 < 1/2$ (and not $(1+\gamma)/2$). However, in some other models the optimal growth rate f_1 of $\bar{F}_n$ can indeed be lower than that of F_n. For example, as we noted before for the function L_n, if $F_n(x,y) \equiv F_n(x)$, then $\bar{F}_n(x,y) \equiv 0$ (that is, in particular, $f_1 = 0$), and this along with Theorem 3.2 leads to Corollary 4.4 about Harris ergodicty of a large class of multiplicative Markovian systems.

Proof of Proposition 4.2. Besides the different parameters in Assumption 4.1, other constants appearing in various estimates below will be denoted by $\mathfrak{m}_i$'s. They will not depend on n but may depend on the parameters of the system.

For the proof we will only consider the case of (SS-5)-(a), where $f_0 < (1+\gamma)/2$; the proofs in the cases of (SS-5)-(b) and the second point in Remark 4.3 follow from (22) and some minor modification of the arguments. For each n, define the functions $\hat{G}_n : \mathbb{H}\times\mathcal{E} \to \mathbb{H}$ and $\tilde{G}$, $\bar{G}_n : \mathbb{H}\times\mathcal{E}\times\mathbb{B} \to \mathbb{H}$ by

$$\hat{G}_n(x,y) = \int_{\mathbb{B}} G_n(x,y,z)\nu_{n+1}(dz), \quad \tilde{G}_n(x,y,z) = G_n(x,y,z) - \hat{G}_n(x,y), \quad \text{and}$$
$$\begin{aligned}\bar{G}_n(x,y,z) &= G_n(x,y,z) - P_{n,x}\hat{G}_n(x,\cdot)(y)\\ &= G_n(x,y,z) - \mathbb{E}(G(X_n,Y_{n+1},\xi_{n+1})|(X_n,Y_n)=(x,y))\end{aligned}$$

(recall that ν_n is the distribution measure of ξ_n), and notice that by (SS-4) and (SS-6) for any $p > 0$,

$$\begin{aligned}\mathbb{E}\left[|\hat{G}_n(X_n,Y_{n+1})|^p|\mathcal{F}_n\right] &= \int_{\mathcal{E}}\left(\int_{\mathbb{B}} G_n(x,y,z)\nu_{n+1}(dz)\right)^p P_{n,X_n}(Y_n,dy)\\ &\leqslant \int_{\mathcal{E}}\int_{\mathbb{B}} \mathfrak{m}_G^p(y)(1+\|X_n\|)^{pg_0}\Psi(z)^p\nu_{n+1}(dz)P_{n,X_n}(Y_n,dy)\\ &\leqslant \bar{\mathfrak{m}}_{\hat{G},p}(1+\|X_n\|)^{pg_0}, \end{aligned}\tag{20}$$

where $\bar{\mathfrak{m}}_{\hat{G},p} = \bar{\mathfrak{m}}_{G,p}m_*^p$ (recall $m_*^p = \sup_k \mathbb{E}[\Psi(\xi_k)^p] < \infty$). It now easily follows that $\bar{G}_n$ and $\tilde{G}_n$ satisfy the following growth conditions:

$$\|\tilde{G}_n(x,y,z)\| \leqslant \mathfrak{m}_{\tilde{G}}(y)(1+\|x\|)^{g_0}\Psi(z), \quad \text{and} \quad \|\bar{G}_n(x,y,z)\| \leqslant \mathfrak{m}_{\bar{G}}(y)(1+\|x\|)^{g_0}\Psi(z)$$

for some functions $\mathfrak{m}_{\bar{G}}(y)$ and $\mathfrak{m}_{\tilde{G}}(y)$ (depending on y), where $\bar{\mathfrak{m}}_{\chi,p} < \infty$ for $\chi = \tilde{G}, \bar{G}$ (see (19) for definition of $\bar{\mathfrak{m}}_{\chi,p}$). Consequently, for any $p > 0$

$$\mathbb{E}\left[\|\tilde{G}_n(X_n, Y_{n+1}, \xi_{n+1})\|^p \big| \mathcal{F}_n\right] \leqslant \bar{\mathfrak{m}}_{\tilde{G},p} m_*^p (1 + \|X_n\|)^{pg_0},$$
$$\mathbb{E}\left[\|\bar{G}_n(X_n, Y_{n+1}, \xi_{n+1})\|^p \big| \mathcal{F}_n\right] \leqslant \bar{\mathfrak{m}}_{\bar{G},p} m_*^p (1 + \|X_n\|)^{pg_0}.$$

Also,

$$\begin{aligned}\mathbb{E}\left[\|L_n(X_n, Y_{n+1})\|^2 \big| \mathcal{F}_n\right] &\leqslant \|X_n\|^2 + 2\bar{\mathfrak{m}}_{L,2,2}^{1/2}\|X_n\| + \bar{\mathfrak{m}}_{L,2,2} = \left(\bar{\mathfrak{m}}_{L,2,2}^{1/2} + \|X_n\|\right)^2 \\ \mathbb{E}\left[\|F_n(X_n, Y_{n+1})\|^2 \big| \mathcal{F}_n\right] &\leqslant \bar{\mathfrak{m}}_{F,2}(1 + \|X_n\|)^{2f_0}.\end{aligned} \tag{21}$$

Now writing $G(X_n, Y_{n+1}, \xi_{n+1}) = \hat{G}_n(X_n, Y_{n+1}) + \tilde{G}(X_n, Y_{n+1}, \xi_{n+1})$, we have

$$\begin{aligned}\|X_{n+1}\|^2 = \|L_n(X_n, Y_{n+1})\|^2 &+ \|F_n(X_n, Y_{n+1})\|^2 + \|\hat{G}_n(X_n, Y_{n+1})\|^2 \\ &+ \|\tilde{G}(X_n, Y_{n+1}, \xi_{n+1})\|^2 \\ &+ 2\langle L_n(X_n, Y_{n+1}), F_n(X_n, Y_{n+1})\rangle \\ &+ 2\langle (L_n + F_n + \hat{G}_n)(X_n, Y_{n+1}), \tilde{G}(X_n, Y_{n+1}, \xi_{n+1})\rangle \\ &+ 2\langle (L_n + F_n)(X_n, Y_{n+1}), \hat{G}_n(X_n, Y_{n+1})\rangle.\end{aligned}$$

Denoting the term $\langle (L_n + F_n + \hat{G}_n)(X_n, Y_{n+1}), \tilde{G}(X_n, Y_{n+1}, \xi_{n+1})\rangle$ by J_{n+1} , we have

$$\begin{aligned}\mathbb{E}\left[J_{n+1} | \mathcal{F}_n\right] &= \int_{\mathbb{B}}\int_{\mathcal{E}} \langle (L_n + F_n + \hat{G}_n)(X_n, y), \tilde{G}(X_n, y, z)\rangle P_{n,X_n}(Y_n, dy)\nu_{n+1}(dz) \\ &= \int_{\mathcal{E}} \left\langle (L_n + F_n + \hat{G}_n)(X_n, y), \int_{\mathbb{B}} \tilde{G}(X_n, y, z)\nu_{n+1}(dz)\right\rangle P_{n,X_n}(Y_n, dy) \\ &= 0.\end{aligned}$$

Also by Cauchy-Schwartz inequality, (20) and (21)

$$\begin{aligned}&\mathbb{E}\left[|\langle F_n(X_n, Y_{n+1}), \hat{G}_n(X_n, Y_{n+1})\rangle| \big| \mathcal{F}_n\right] \\ &\leqslant \left(\mathbb{E}\left[\|F_n(X_n, Y_{n+1})\|^2 \big| \mathcal{F}_n\right]\right)^{1/2} \left(\mathbb{E}\left[\|\hat{G}_n(X_n, Y_{n+1})\|^2 \big| \mathcal{F}_n\right]\right)^{1/2} \\ &\leqslant \bar{\mathfrak{m}}_{F,2}^{1/2}\bar{\mathfrak{m}}_{\hat{G},2}^{1/2}(1 + \|X_n\|)^{f_0+g_0},\end{aligned}$$

and similarly,

$$\mathbb{E}\left[|\langle L_n(X_n, Y_{n+1}), \hat{G}_n(X_n, Y_{n+1})\rangle| \big| \mathcal{F}_n\right] \leqslant \bar{\mathfrak{m}}_{\hat{G},2}^{1/2}\left(\bar{\mathfrak{m}}_{L,2,2}^{1/2} \vee 1 + \|X_n\|\right)^{1+g_0}.$$

Hence, on $\{\|X_n\| > B\}$,

$$\begin{aligned}\mathbb{E}\left[\|X_{n+1}\|^2 | \mathcal{F}_n\right] &\leqslant \|X_n\|^2 + 2\bar{\mathfrak{m}}_{L,2,2}^{1/2}\|X_n\| + \bar{\mathfrak{m}}_{L,2,2} + \bar{\mathfrak{m}}_{F,2}(1 + \|X_n\|)^{2f_0} \\ &\quad + (\bar{\mathfrak{m}}_{\hat{G},p} + \bar{\mathfrak{m}}_{\tilde{G},p} m_*^p)(1 + \|X_n\|)^{2g_0} - 2\mathfrak{m}_0\|X_n\|^{1+\gamma} \\ &\quad + 2\bar{\mathfrak{m}}_{F,2}^{1/2}\bar{\mathfrak{m}}_{\hat{G},2}^{1/2}(1 + \|X_n\|)^{f_0+g_0} + 2\bar{\mathfrak{m}}_{\hat{G},2}^{1/2}\left(\bar{\mathfrak{m}}_{L,2,2}^{1/2} \vee 1 + \|X_n\|\right)^{1+g_0}.\end{aligned} \tag{22}$$

Since $\delta_0 \doteq 2(f_0 \vee g_0) \vee (f_0 + g_0) \vee (1 + g_0) < 1 + \gamma$, by (SS-5) it follows from the above inequality that we can choose $C > B$ large enough so that for $\|x_n\| > C$,

$$\mathbb{E}\left[\|X_{n+1}\|^2 - \|X_n\|^2 \mid \mathcal{F}_n\right] \leqslant \mathfrak{m}_1\left(\|X_n\|^{\delta_0} - \|X_n\|^{1+\gamma}\right) < 0.$$

Also notice that choosing $C > B \vee 1$ we have for $\|X_n\| > C$

$$\sqrt{\mathbb{E}\left[\|X_{n+1}\|^2|\mathcal{F}_n\right]} + \|X_n\| \leqslant \mathfrak{m}_2(1+\|X_n\|)^{1\vee\delta_0/2} \leqslant 2^{1\vee\delta_0/2}\mathfrak{m}_2\|X_n\|^{1\vee\delta_0/2}.$$

Therefore for $\|X_n\| > C$,

$$\begin{aligned}\mathbb{E}\left[\|X_{n+1}\||\mathcal{F}_n\right] - \|X_n\| \leqslant \sqrt{\mathbb{E}\left[\|X_{n+1}\|^2|\mathcal{F}_n\right]} - \|X_n\| &= \frac{\mathbb{E}\left[\|X_{n+1}\|^2 \mid \mathcal{F}_n\right] - \|X_n\|^2}{\sqrt{\mathbb{E}\left[\|X_{n+1}\|^2|\mathcal{F}_n\right]} + \|X_n\|}\\ &\leqslant \mathfrak{m}_3\left(\|X_n\|^{\delta_0 - 1\vee\delta_0/2} - \|X_n\|^{1+\gamma-1\vee\delta_0/2}\right).\end{aligned}$$

Because of assumption (SS-5), notice that

$$\|x\|^{\delta_0-1\vee\delta_0/2} - \|x\|^{1+\gamma-1\vee\delta_0/2} \xrightarrow[\|x\|\to\infty]{} \begin{cases} -\infty & \gamma > 0,\\ -\mathfrak{m}_3 & \gamma = 0.\end{cases}$$

In either case, there exist a constant $A > 0$, and a sufficiently large C, such that

$$\mathbb{E}\left[\|X_{n+1}\||\mathcal{F}_n\right] - \|X_n\| \leqslant -A \quad \text{on } \|X_n\| > C. \tag{23}$$

Next, notice that

$$\begin{aligned}&\Big|\|X_{n+1}\| - \mathbb{E}\left[\|X_{n+1}\|\big|\mathcal{F}_n\right]\Big|\\ \leqslant\ &\Big|\|X_{n+1}\| - \|\mathbb{E}[X_{n+1}|\mathcal{F}_n]\|\Big| + \Big|\|\mathbb{E}[X_{n+1}|\mathcal{F}_n]\| - \mathbb{E}\left[\|X_{n+1}\|\big|\mathcal{F}_n\right]\Big|\\ \leqslant\ &\|X_{n+1} - \mathbb{E}\left[X_{n+1}\big|\mathcal{F}_n\right]\| + \Big|\mathbb{E}\left[\|X_{n+1}\| - \|\mathbb{E}[X_{n+1}|\mathcal{F}_n]\|\big|\mathcal{F}_n\right]\Big|\\ \leqslant\ &\|X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\| + \mathbb{E}\left[\|X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\|\big|\mathcal{F}_n\right].\end{aligned}$$

Hence,

$$\begin{aligned}\Xi_n &= \mathbb{E}\left[\Big|\|X_{n+1}\| - \mathbb{E}\left[\|X_{n+1}\|\big|\mathcal{F}_n\right]\Big|^p\Big|\mathcal{F}_n\right] \leqslant\ 2^p\mathbb{E}\left[\|X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\|^p\Big|\mathcal{F}_n\right]\\ &= 2^p\mathbb{E}\left[\|\bar{L}(X_n,Y_{n+1}) + \bar{F}(X_n,Y_{n+1}) + \bar{G}(X_n,Y_{n+1},\xi_{n+1})\|^p\Big|\mathcal{F}_n\right]\\ &\leqslant \mathfrak{m}_4(1+\|X_n\|)^{p(l_1\vee f_1\vee g_0)} \equiv \phi_p(X_n),\end{aligned} \tag{24}$$

where $\phi_p(x) \doteq \mathfrak{m}_4(1+\|x\|)^{p(l_1\vee f_1\vee g_0)}$. Since $l_1\vee f_1\vee g_0 < 1/2$, for large enough p, we have $p(l_1\vee f_1\vee g_0) < p/2-1$. It now follows from Theorem 2.2 (using $V(x) = \|x\|$) that for any $r \in (0, \varsigma(s = p(l_1\vee f_1\vee g_0), p))$, $\sup_n \mathbb{E}\|X_n\|^r < \infty$. Since $p > 0$ is arbitrarily large, the assertion follows.

If $G_n(x,y,z)$ are centered, that is, if $\hat{G}_n \equiv 0$, then of course $\bar{\mathfrak{m}}_{\hat{G},p}$ can be taken to be 0 for all $p > 0$, and from (22) , $\delta_0 = 2(f_0 \vee g_0)$. Consequently, we do not need $g_0 < \gamma$ to have $\delta_0 < 1+\gamma$. □

Corollary 4.4. *Consider the class of $\{\mathcal{F}_n\}$-adapted Markov processes taking values in $\mathbb{R}^d$, whose dynamics is defined by*

$$X_{n+1} = L(X_n) + F(X_n) + G(X_n)\xi_{n+1}, \tag{25}$$

where $F, L : \mathbb{R}^d \to \mathbb{R}^d$, $G : \mathbb{R}^d \to \mathbb{M}^{d\times d'}$ are continuous functions, and $d \leqslant d'$. Assume that

(M-1) *F, G and L satisfy the growth conditions (a) $\|L(x)\| \leqslant \|x\|$ for $\|x\| > B$, (b) $\|F(x)\| \leqslant \mathfrak{m}_F(1+\|x\|)^{\gamma_0}$, and (c) $\|G(x)\| \leqslant \mathfrak{m}_G(1+\|x\|)^{g_0}$;*

(M-2) *for some constants* $\mathfrak{m}_0, B$ *and exponent* $\gamma \geqslant 0$,

$$\langle F(x), L(x)\rangle \leqslant -\mathfrak{m}_0\|x\|^{1+\gamma} \quad \text{for } \|x\| > B;$$

(M-3) *the exponents satisfy: (a)* $\gamma_0 < (1+\gamma)/2$, *or* $\gamma_0 = (1+\gamma)/2$ *and* $\mathfrak{m}_F \leqslant \mathfrak{m}_0/2$; *(b)* $g_0 < 1/2$;

(M-4) *the* ξ_n *are i.i.d* $\mathbb{R}^{d'}$*-valued random variables with density* ρ *with respect to Lebesgue measure,* λ_{leb}; $\rho(z) > 0$ *for all* $z \in \mathbb{R}^{d'}$, $\sup_{z\in\mathbb{R}^{d'}} \rho(z) < \infty$, *and for each* $p > 0$, $m_*^p = \mathbb{E}(\|\xi_1\|^p) < \infty$;

(M-5) *for some* $\theta \geqslant 0$ *and* $\varepsilon_0 > 0$,

$$u^T G(x)G(x)^T u \geqslant \varepsilon_0 u^T u/(1+\|x\|)^\theta \quad \text{for all } u, x \in \mathbb{R}^d.$$

If, in addition, $\mathbb{E}(\xi_1) = 0$, *then (a)* $\{X_n\}$ *is PHR and aperiodic with a unique invariant distribution* π, *(b)* $\sup_n \mathbb{E}_{x_0}(\|X_n\|^r) \vee \mathbb{E}_\pi\|X_n\|^r < \infty$, *and (c) (12) or equivalently, (13) holds with* $V(u) = \|u\|$ *for any* x_0 *and* $r > 0$. *If* $\mathbb{E}(\xi_1) \neq 0$, *then the same assertion is true provided* $g_0 < \gamma \wedge 1/2$

Proof. Since L, F and G are continuous, it follows by the dominated convergence theorem that $\{X_n\}$ is weak-Feller. From assumption (M-5) it follows that GG^T is positive definite (in particular, non singular) and that $\det(G(x)G^T(x)) \geqslant \varepsilon_0^d/(1+\|x\|)^{\theta d}$. Note that $\mathcal{P}(x,\cdot)$ admits a density $q(x,\cdot)$. Specifically,

$$q(x,y) = \frac{1}{\sqrt{\det(G(x)G^T(x))}} \cdot \rho\left(G(x)_R^-(y - L(x) - H(x)\right) \leqslant \sup_z \rho(z)(1+\|x\|)^{\theta d/2}/\varepsilon_0^{d/2},$$

where $G(x)_R^- = G^T(x)\left(G(x)G(x)^T\right)^{-1}$ is the Moore-Penrose pseudoinverse (in particular, right inverse) of $G(x)$. Moreover, since $\rho(z) > 0$ a.s, for each x, $q(x,y) > 0$ a.s in y (with respect to λ_{leb}), and consequently, $\{X_n\}$ is λ_{leb}-irreducible. This shows that Condition (3.2-d) of Theorem 3.2 holds. The various assertions now follow from Theorem 3.2 and Proposition 4.2 ☐

Remark 4.5. The condition (M-5) is much weaker than the uniform ellipticity condition that is sometimes imposed on GG^T for these kinds of models — the latter requiring, for some $\varepsilon_0 > 0$, $u^T G(x)G(x)^T u \geqslant \varepsilon_0 u^T u$ for all $u, x \in \mathbb{R}^d$. Corollary 4.4 also holds, with some possible minor modifications, for systems of the form (25) taking values in other locally compact spaces with ξ_n admitting a density ρ with respect to the Haar measure. In particular, for such systems taking values in a countable state space like $\mathbb{Z}^d$ or $\mathbb{Q}^d$, notice that the transition probability mass function (density with respect to the counting measure) $q(x,y)$ naturally exists and $q(x,y) \leqslant 1$, that is, the bound on q in condition (3.2-d) of Theorem 3.2 is trivially satisfied. Hence condition (M-5) in Corollary 4.4 is not needed in this case. However, depending on the specific model, one may still require G to have full row rank for establishing irreducibility of the chain.

As an important application, the above corollary can be used to establish ergodicity of numerical schemes of stochastic differential equations (SDEs).

Example 4.6. Euler-Maruyama scheme for ergodic SDEs: Consider the SDE

$$X(t) = X(0) + \int_0^t F(X(s))ds + \int_0^t G(X(s))dW(s),$$

and suppose that X is ergodic with invariant / equilibrium distribution π - which is typically unknown. Approximating this equilibrium distribution is an important computational problem in various areas including statistical physics, machine learning, mathematical finance etc. Since numerically solving the corresponding (stationary) Kolomogorov PDE for π is computationally expensive even when the dimension is as low as 3, one commonly resorts to discretization schemes like the Euler-Maruyama method:

$$X^{\Delta}(t_{n+1}) = X^{\Delta}(t_n) + F(X^{\Delta}(t_n))\Delta + \Delta^{1/2}G(X^{\Delta}(t_n))\xi_{n+1}.$$

Here the ξ_n are iid $N(0, I)$-random variables, and $\{t_n\}$ is a partition of $[0, \infty)$ with $t_{n+1} - t_n = \Delta$ - the step size of discretization. However, the use of such discretization techniques in approximating π is justified provided one can establish (a) ergodicity of the discretized chain $\{X^{\Delta}(t_n)\}$ with a unique invariant distribution π^{Δ}, and (b) convergence of π^{Δ} to π as $\Delta \to 0$. This is a hard problem involving infinite time horizon, and usual error analysis of Euler-Maruyama schemes, which has of course been well studied in the literature, is not useful here, as they are over finite time intervals. In comparison, much less is available on theoretical error analyses of these types of infinite-time horizon approximation problems, and some important results in this direction have been obtained by Talay [28, 27, 7]. A recent paper [6] (also see the references therein for more background on the problem) conducts a thorough large deviation error analysis of the problem in an appropriate scaling regime.

This short example does not attempt to address both the points (a) and (b) of this problem as that requires a separate paper-long treatment. Here, we are only interested in the point (a) above - which is ergodicity of the discretized chain $\{X^{\Delta}(t_n)\}$. It is well known that ergodicity of X does not guarantee the ergodicity of the discretized chain X^{Δ}. Discretization can destroy the underlying Lyapunov structure of an ergodic SDE!

In [28, 27] among several other important results, Talay et al. in particular showed that the chain $\{X^{\Delta}(t_n)\}$ is ergodic with unique invariant measure π^{Δ} and $\mathbb{E}(f(X^{\Delta}(t_n)) \to \pi^{\Delta}(f)$ as $n \to \infty$ for any $f \in C^{\infty}(\mathbb{R}^d, \mathbb{R})$ such that f and all its derivatives have polynomial growth under the assumption (i) $\langle F(x), x\rangle \leqslant -\mathfrak{m}_0\|x\|^2$, for $\|x\| > B$, (ii) F and G are C^{∞} with bounded derivatives of all order and (iii) GG^T is uniformly elliptic and bounded. An application of Corollary 4.4 shows that this result can be significantly improved with stronger convergence results under weaker hypothesis (c.f (M-1) -(M-5)). In particular, uniform ellipticity and boundedness conditions on GG^T, which are quite restrictive for many models, can be removed.

4.2. **Moment stability of linear stochastic control systems.** Consider the system

$$X_{n+1} = AX_n + Bu_n + \xi_{n+1} \tag{26}$$

We are interested in the problem of finding conditions under which a linear stochastic system with possibly unbounded additive stochastic noise is globally stabilizable with bounded control inputs $\{u_n\}$. Stabilization of stochastic linear systems with bounded control is a topic of significant interest in control engineering because of its importance in diverse fields; suboptimal control strategies such as receding-horizon control, and rollout algorithms, among others, can be easily constructed incorporating such constraints, and have become popular in applications. Here we simply refer to [25] and references therein for a detailed background on this topic.

Of course, boundedness of some moments of the noise component is necessary for attaining (moment) stability of the system. Specifically, we consider the following problem:

Problem: Suppose $\mathbb{U} \doteq \{z \in \mathbb{R}^m : \|z\| \leqslant U_{\max}\}$. We consider admissible possible k-history dependent control policies of the type $\pi = \{\pi_n\}$ so that $\pi_n : \mathbb{R}^{d\times k} \to \mathbb{U}$, and for every $y_1, y_2 \ldots, y_k \in \mathbb{R}^d$, $\pi_n(y_1, \ldots, y_k) \in \mathbb{U}$. Given $r \geqslant 1$ and $U_{\max} > 0$, find an admissible policy $\pi = \{\pi_n\}_{n\in\mathbb{N}}$ with control authority $U_{\max}$, such that the system (26) with $u_n = \pi_n(X_{n-k+1}, \ldots, X_{n-1}, X_n)$ is r-th moment stable, that is, for every initial condition $X_0 = x_0$, $\sup_n \mathbb{E}_{x_0}\|X_n\|^r < \infty$.

It is known that mean square boundedness holds for systems with bounded controls where A is Schur stable, that is, all eigenvalues of A are contained in the open unit disk (the proof uses Foster-Lyapunov techniques from [18]). In the more general framework, under the assumption that the pair (A, B) is only stabilizable (which in particular allows the eigenvalues of A to lie on the closed unit disk), [25] using [22] showed that there exist a k-history dependent control policy that ensures moment stability of (26), provided the control authority $U_{\max}$ is chosen sufficiently large. It was conjectured in [25], that the lower bound on $U_{\max}$ can possibly be lifted with newer techniques, and here we demonstrate that is indeed the case. The following result is an easy corollary of Proposition 4.2. For simplicity, we assume that A is orthogonal and (A, B) is reachable in k-steps. The steps from there to the more general case are similar to that in [25]. In case B has full row rank, it will follow that k can be taken to be 1, that is, the resulting policy is stationary feedback.

Proposition 4.7. *Consider the system defined by* (26)*. Suppose that A is orthogonal and the pair (A, B) is reachable in k steps (that is,* $\operatorname{rank}(\mathcal{R}_k) = d$*, where $\mathcal{R}_k = [B\ AB\ A^2B\ \ldots\ A^{k-1}B]$). Then for any $U_{\max} > 0$, there exists a k-history dependent policy $\pi = \{\pi_n\}$ such that given $(X_{n-k+1}, \ldots, X_{n-1}, X_n) = (x_{n-k+1}, \ldots, x_{n-1}, x_n)$, $\pi_n(x_{n-k+1}, \ldots, x_{n-1}, x_n) \doteq f_{n \bmod k}(x_{\lfloor n/k\rfloor k})$ for some functions $f_0, f_1, \ldots, f_{k-1} : \mathbb{R}^d \to \mathbb{R}^m$ where $\|f_i(x)\| \leqslant U_{\max}$ for $i = 0, 1, 2, \ldots, k-1$, and for which* $\sup_n \mathbb{E}_{x_0}\|X_n\|^r < \infty$ *for any $x_0 \in \mathbb{R}^d$.*

Proof. Define $\hat{X}^{(k)}_n = X_{nk}$, and notice that by iterating (26) we get

$$\hat{X}^{(k)}_{n+1} = A^k \hat{X}^{(k)}_n + \mathcal{R}_k \begin{pmatrix} u_{(n+1)k-1} \\ \vdots \\ u_{nk+1} \\ u_{nk} \end{pmatrix} + \sum_{j=1}^{k} A^{k-1-j}\xi_{nk+i} \equiv A^k \hat{X}^{(k)}_n + \mathcal{R}_k \hat{u}^{(k)}_n + \hat{\xi}^{(k)}_n$$

Notice that $\mathbb{E}(\xi^{(k)}_n) = 0$ and $\sup_n \mathbb{E}\|\xi^{(k)}_n\|^p \leqslant \hat{\mathscr{C}_k}$ for some constant $\hat{\mathscr{C}_k} > 0$. Since $\mathcal{R}_k$ has full row rank, it has a right inverse $\mathcal{R}^-_k$. We define

$$\operatorname{sat}(y) = \begin{cases} y & \text{if } y \in B(0, \hat{U}_{\max}), \\ \hat{U}_{\max}\ y/\|y\| & \text{otherwise,} \end{cases}$$

and choose $\hat{u}_n^{(k)} = -\mathcal{R}_k^- A^k \operatorname{sat}(\hat{X}_n^{(k)})$, where $\hat{U}_{\max}$ is such that $\|\mathcal{R}_k^- A^k\|\hat{U}_{\max} \leqslant U_{\max}$. This yields the system

$$\hat{X}_{n+1}^{(k)} = A^k \hat{X}_n^{(k)} - A^k \operatorname{sat}(\hat{X}_n^{(k)}) + \hat{\xi}_n^{(k)}.$$

Since for $\|z\| > U_{\max}$, $\langle A^k z, -A^k \operatorname{sat}(z)\rangle = -\|z\|$ (recall that A is orthogonal), we have from Proposition 4.2 that there exists a constant $\mathfrak{c}_0^{(k,r)}$ such that

$$\sup_n \mathbb{E}\|\hat{X}_n^{(k)}\|^r = \sup_n \mathbb{E}\|X_{nk}\|^r < \mathfrak{c}_0^{(k)}.$$

It is now immediate from a sequential argument that for any $\ell = 0, 1, \ldots, k-1$, we have $\mathbb{E}\|X_{nk+\ell}\|^r \leqslant \mathfrak{c}_\ell^{(k,r)}$, where $\mathfrak{c}_\ell^{(k,r)} = 3^{r-1}\left(\|A\|^r \mathfrak{c}_{\ell-1}^{(k,r)} + \|\mathcal{R}_k^- A^k\|^r U_{\max}^r + \mathfrak{m}_*^r\right)$.

Notice that the original controls u_n are defined

$$u_n = -E_{k-(n \bmod k)}^T \mathcal{R}_k^- A^k \operatorname{sat}(X_{\lfloor n/k\rfloor k}),$$

where the matrices $E_j \in \mathbb{M}_{m\times km}$, $j = 1, 2, \ldots, k$, are defined by

$$E_j = \begin{bmatrix} \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} & \underbrace{\boldsymbol{I}_{m\times m}}_{j\text{-th block}} & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \end{bmatrix}$$

In particular, from the state at time nk, the present and the next $k-1$ controls $u_j, j = nk, nk+1, \ldots, nk+k-1$ can be computed. □

Appendix A. Appendix.

Lemma A.1. *Let ν be a signed measure on a complete separable metric space $\mathcal{E}$. Suppose that $g : \mathcal{E} \to [0, \infty)$ is a measurable function such that $|\nu|(g) = \int g\, d|\nu| < \infty$. Then*

$$\frac{1}{2}|\nu|(g) \leqslant \|\nu\|_g \leqslant |\nu|(g),$$

where we recall that $\|\nu\|_g = \sup_{\{f:|f|\leqslant g\}} |\nu(f)|$.

Proof. The second inequality is trivial since for any measurable f with $|f| \leqslant g$, we have $|\nu(f)| \leqslant |\nu|(|f|) \leqslant |\nu|(g)$. For the first inequality, let $\mathcal{E} = \mathcal{Y} \cup \mathcal{N}$ be the Hahn decomposition for ν (in particular, $\mathcal{Y} \cap \mathcal{N} = \varnothing$), with the corresponding Jordan decomposition $\nu = \nu^+ - \nu^-$ (i.e., $\operatorname{supp}(\nu^+) \subset \mathcal{Y}$ and $\operatorname{supp}(\nu^-) \subset \mathcal{N}$). Choose $f = g\mathbb{1}_{\mathcal{Y}}$. Then

$$\|\nu\|_g \geqslant |\nu(g\mathbb{1}_{\mathcal{Y}})| = |\nu^+(g\mathbb{1}_{\mathcal{Y}}) - \nu^-(g\mathbb{1}_{\mathcal{Y}})| = \nu^+(g\mathbb{1}_{\mathcal{Y}}) = \nu^+(g),$$

where the last equality is because $\operatorname{supp}(\nu^+) \subset \mathcal{Y}$. Similarly, choosing $f = g\mathbb{1}_{\mathcal{N}}$, we have $\|\nu\|_g \geqslant \nu^-(g)$, whence it follows that $2\|\nu\|_g \geqslant |\nu|(g)$. □

REFERENCES

[1] M. F. Barnsley, S. G. Demko, J. H. Elton and J. S. Geronimo, Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities, *Annales de l'Institut Henri Poincaré. Probabilités et Statistique*, **24** (1988), 367-394, *Erratum in ibid.*, **24** (1989), 589-590.

[2] D. Chatterjee, E. Cinquemani and J. Lygeros, Maximizing the probability of attaining a target prior to extinction, *Nonlinear Analysis: Hybrid Systems*, **5** (2011), 367-381, Special Issue related to IFAC Conference on Analysis and Design of Hybrid Systems (ADHS'09) - IFAC ADHS'09.

[3] D. Chatterjee and S. Pal, An excursion-theoretic approach to stability of discrete-time stochastic hybrid systems, *Applied Mathematics & Optimization*, **63** (2011), 217-237.

[4] O. L. V. Costa, M. D. Fragoso and R. P. Marques, *Discrete-Time Markov Jump Linear Systems*, Probability and its Applications (New York), Springer-Verlag London, Ltd., London, 2005.

[5] P. Diaconis and D. Freedman, Iterated random functions, *SIAM Review*, **41** (1999), 45-76 (electronic).

[6] A. Ganguly and P. Sundar, Inhomogeneous functionals and approximations of invariant distributions of ergodic diffusions: central limit theorem and moderate deviation asymptotics, *Stochastic Processes and their Applications*, **133** (2021), 74-110.

[7] C. Graham and D. Talay, *Stochastic Simulation and Monte Carlo Methods*, vol. 68 of Stochastic Modelling and Applied Probability, Springer, Heidelberg, 2013, Mathematical foundations of stochastic simulation.

[8] O. Hernández-Lerma and J.-B. Lasserre, Further criteria for positive Harris recurrence of Markov chains, *Proceedings of the American Mathematical Society*, **129** (2001), 1521-1524.

[9] O. Hernández-Lerma and J.-B. Lasserre, *Markov Chains and Invariant Probabilities*, vol. 211 of Progress in Mathematics, Birkhäuser Verlag, Basel, 2003.

[10] J. P. Hespanha, A model for stochastic hybrid systems with application to communication networks, *Nonlinear Analysis. Theory, Methods & Applications*, **62** (2005), 1353-1383.

[11] S. F. Jarner and R. L. Tweedie, Locally contracting iterated functions and stability of Markov chains, *Journal of Applied Probability*, **38** (2001), 494-507.

[12] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise*, vol. 97 of Applied Mathematical Sciences, 2nd edition, Springer-Verlag, New York, 1994.

[13] A. Lasota and J. A. Yorke, Lower bound technique for Markov operators and iterated function systems, *Random & Computational Dynamics*, **2** (1994), 41-77.

[14] X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London, 2006.

[15] S. P. Meyn and R. L. Tweedie, Stability of Markovian processes. I. Criteria for discrete-time chains, *Advances in Applied Probability*, **24** (1992), 542-574.

[16] S. P. Meyn and R. L. Tweedie, Stability of Markovian processes. II. Continuous-time processes and sampled chains, *Advances in Applied Probability*, **25** (1993), 487-517.

[17] S. P. Meyn and R. L. Tweedie, Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes, *Advances in Applied Probability*, **25** (1993), 518-548.

[18] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd edition, Cambridge University Press, London, 2009.

[19] M. Michel, *Jump Linear Systems in Automatic Control*, Marcel Dekker, New York, 1990.

[20] S. V. Nagaev, An alternative method of the proof of the ergodic theorem for general Markov chains, *Theory of Probability and its Applications*, **66** (2021), 364-375.

[21] M. Peigné, *Iterated Function Systems and Spectral Decomposition of the Associated Markov Operator*, in *Fascicule de Probabilités*, vol. 1993 of Publ. Inst. Rech. Math. Rennes, Univ. Rennes I, Rennes, 1993, 28pp.

[22] R. Pemantle and J. S. Rosenthal, Moment conditions for a sequence with negative drift to be uniformly bounded in L^r, *Stochastic Processes and their Applications*, **82** (1999), 143-155.

[23] G. D. Prato, *An Introduction to Infinite-Dimensional Analysis*, Universitext, Springer-Verlag, Berlin, 2006, Revised and extended from the 2001 original by Da Prato.

[24] P. E. Protter, *Stochastic Integration and Differential Equations*, vol. 21 of Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing.

[25] F. Ramponi, D. Chatterjee, A. Milias-Argeitis, P. Hokayem and J. Lygeros, Attaining mean square boundedness of a marginally stable stochastic linear system with a bounded control input, *IEEE Transactions on Automatic Control*, **55** (2010), 2414-2418.

[26] T. Szarek, Invariant measures for nonexpensive Markov operators on Polish spaces, *Dissertationes Mathematicae (Rozprawy Matematyczne)*, Dissertation, Polish Academy of Science, Warsaw, **415** (2003), 62pp.

[27] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, *Stochastics and Stochastic Reports*, **29** (1990), 13-36.

[28] D. Talay and L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, *Stochastic Analysis and Applications*, **8** (1990), 483-509 (1991).

[29] G. G. Yin and C. Zhu, *Hybrid Switching Diffusions*, vol. 63 of Stochastic Modelling and Applied Probability, Springer, New York, 2010, Properties and applications.

Received May 2022; revised November 2022; early access March 2023.