EXIT TIME ANALYSIS FOR KESTEN'S STOCHASTIC RECURRENCE EQUATIONS

CHANG-HAN RHEE, JEEHO RYU, INSUK SEO

Abstract. This paper studies the exit times of the Kesten's stochastic recurrence equations. Depending on the Lyapunov exponent, the exit time scales polynomially or logarithmically as the radius of the exit boundary increases. The polynomial scaling corresponds to the convergent behavior, and the logarithmic scaling corresponds to the divergent behavior, respectively, in the deterministic counterpart.

1. Introduction

Since the seminal work of Kesten [6], the stochastic recurrent equation

$$\mathbf{X}_{n+1} = \mathbf{A}_{n+1}\mathbf{X}_n + \mathbf{B}_{n+1} \tag{1.1}$$

has been a subject of extensive research throughout the past decades. The precise conditions we impose on $(\mathbf{A}_i, \mathbf{B}_i)$ throughout this paper will be described in Definition 2.1, but here we mention that they are i.i.d. with lighter tails than that of $\boldsymbol{X}_n$'s stationary distribution. Note that otherwise, the overall tail behavior of $\mathbf{X}_{n+1}$ is simply dominated by the tail behavior of either $\mathbf{A}_n$ or $\mathbf{B}_n$. In case the recursion is contractive and stabilizes over time, the stationary distribution of $\boldsymbol{X}_n$ has been studied extensively. In particular, Kesten has shown that the stationary distribution of $\{\mathbf{X}_i\}_{i=1,2,\dots}$ is heavy-tailed even when $\mathbf{A}_i$ and $\mathbf{B}_i$ are light-tailed.

A curious phenomena called *edge of stability* was observed in a wide variety of neural network architectures and drew significant interest from the machine learning community. The phenomena has been analyzed and observed most prominently in the context of the deterministic gradient descent algorithm, where the dichotomy between convergent and divergent behavior of the recursion that drives the algorithm can be determined clearly. However, to understand the implication of the edge of stability and leverage such phenomena in practical applications, it is imperative to understand the stochastic counterpart, because the computation of precise gradient is prohibitively expensive, and hence, the stochastic version, stochastic gradient descent (SGD), is almost always employed. Kesten's recursion (1.1) has been associated with the dynamics of SGD in the context of linear regression or near local optima in more general contexts; see, for example, [2, 4, 5]. Despite the simplicity of the recursion (1.1), the analysis of the aforementioned dichotomy is far more intricate than the deterministic context. In particular, the analysis of the divergent behavior has not been well-understood.

In this paper, we characterize asymptotic limits of the exit times of $\boldsymbol{X}_n$. We study both contractive and explosive cases. Specifically, we show that when the Lyapunov exponent is negative,

$$\lim_{R\to\infty} \frac{\log \mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} = \alpha \quad \text{for all } \mathbf{x}_0 \in \mathbb{R}^d \,,$$

where $\tau_R(\mathbf{x})$ is the exit time from the ball $\mathcal{B}_R = \{\mathbf{y} \in \mathbb{R}^d : |\mathbf{y}| \le R\}$ and $\alpha > 0$ is the index of the power law associated with the tail of the stationary distribution of $\boldsymbol{X}_n$. On the other hand, when the Lyapunov exponent is positive,

$$\frac{1}{\gamma_L} \le \liminf_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} \le \limsup_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} < \infty \,.$$

where γ_L denotes Lyapunov exponent. Therefore, the behavior of the process (2.1) changes qualitatively according to the sign of the Lyapunov exponent; we observe a polynomial to logarithmic phase transition in the mean exit time.

In the critical case, i.e., the case $\gamma_L = 0$, some simulations show that the mean exit time grows poly-logarithimically, but we do not know at this moment whether it is the only possibility or not. The study of the critical case seems to be very difficult and is a possible future research topic.

The remainder of the article is organized as follows. Section 3 presents our main results. In Section 4, we provide several important time series models for which we can apply our main results. In Sections 5 and 6, we analyze the contractive regime and prove Theorems 3.4 and 3.5. In Section 7, we investigate the explosive regime and prove Theorem 3.7. In Appendix A, we provide some sufficient conditions to check Assumption 4, and in Appedix B, we explain why the examples given in Section 4 fall into our framework.

2. Preliminaries

This section sets the notation and reviews known results regarding Kesten's stochastic recurrence equation. Throughout the paper, we write $\mathbb{Z}^+$ and $\mathbb{Z}_0^+$ to denote the set of positive integers and the set of non-negative integers, respectively.

Definition 2.1. Let $d \ge 1$ and let $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$ be an independent and identically distributed sequence of random matrices and vectors, where $\mathbf{A}_n$ is a $d \times d$ random matrix and $\mathbf{B}_n$ is a $d \times 1$ random vector. Then, Kesten's stochastic recurrence equation on $\mathbb{R}^d$ associated with $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$ refers to the process $(\mathbf{X}_n)_{n\in\mathbb{Z}_0^+}$ defined by

$$\mathbf{X}_{n+1} = \mathbf{A}_{n+1}\mathbf{X}_n + \mathbf{B}_{n+1}, \quad n \in \mathbb{Z}_0^+. \tag{2.1}$$

(1) For $\mathbf{x} \in \mathbb{R}^d$, we denote by $(\mathbf{X}_n(\mathbf{x}))_{n\in\mathbb{Z}_0^+}$ the process (2.1) starting at $\mathbf{x}$, i.e., $\mathbf{X}_0(\mathbf{x}) = \mathbf{x}$.
(2) For a Borel probability measure μ on $\mathbb{R}^d$, we denote by $(\mathbf{X}_n(\mu))_{n\in\mathbb{Z}_0^+}$ the process driven by (2.1) with initial distribution μ, i.e., $\mathbb{P}(\mathbf{X}_0(\mu) \in \cdot) = \mu(\cdot)$.
(3) We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space associated with the random elements $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$. We denote by $\mathbb{E}$ the expectation associated with $\mathbb{P}$.

(4) Denote by $\mathcal{F}_n$ the σ-algebra on Ω generated by $(\mathbf{A}_k, \mathbf{B}_k)_{k=1}^n$. Then, it is clear from (2.1) that $\mathbf{X}_n(\mathbf{x})$ is $\mathcal{F}_n$-measurable for each $n \in \mathbb{Z}_0^+$ and furthermore $(\mathbf{X}_n(\mathbf{x}))_{n\in\mathbb{Z}_0^+}$ is a Markov process adapted to the filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}_0^+}$ (where $\mathcal{F}_0$ is the null σ-algebra).

For $R > 0$, denote by $\tau_R(\mathbf{x})$ the exit time of the process $\mathbf{X}_n(\mathbf{x})$ from the ball

$$\mathcal{B}_R = \{\mathbf{y} \in \mathbb{R}^d : |\mathbf{y}| \le R\}, \tag{2.2}$$

i.e.,

$$\tau_R(\mathbf{x}) := \inf\{n \in \mathbb{Z}_0^+ : |\mathbf{X}_n(\mathbf{x})| > R\},$$

where we set $\inf \phi := \infty$ as usual.
In this article, our primarily concern is the asymptotic behavior of the mean exit time $\mathbb{E}[\tau_R(\mathbf{x})]$, in the regime $R \to \infty$.

For $d \times d$ matrix $\mathbf{M}$, we write

$$:= \sup_{|\mathbf{x}|=1} |\mathbf{M}\mathbf{x}|$$

the matrix norm of $\mathbf{M}$. Throughout this article, it will be frequently used that this norm is sub-multiplicative:

$$\|\mathbf{M}\mathbf{N}\| \le \|\mathbf{M}\|\|\mathbf{N}\| . \tag{2.3}$$

For $n \in \mathbb{Z}^+$, define

$$\Pi_n := \mathbf{A}_n\mathbf{A}_{n-1}\cdots\mathbf{A}_1. \tag{2.4}$$

Then, it is well-known (cf. [3, Theorem 1, page 457] and [3, Theorem 2, page 460]) that the limit

$$\gamma_L = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\log \|\Pi_n\|\right] \tag{2.5}$$

exists, and moreover it almost surely holds that

$$\gamma_L = \lim_{n\to\infty} \frac{1}{n} \log \|\Pi_n\| \tag{2.6}$$

by the multiplicative ergodic theorem. The constant γ_L is called a *Lyapunov exponent* associated with sequence $(\mathbf{A}_n)_{n\in\mathbb{Z}^+}$.

At a heuristic level, it is clear that the behavior of the process $(\mathbf{X}_n(\mathbf{x}))_{n\in\mathbb{Z}_0^+}$ depends crucially on the sign of the Lyapunov exponent γ_L. If $\gamma_L < 0$, as the elements of Π_n converges to 0 exponentially fast, the process $\mathbf{X}_n$ contracts to $\mathbf{0}$. On the other hand, if $\gamma_L > 0$, as the elements of Π_n exponentially diverges to $+\infty$, we can expect that the process $\mathbf{X}_n$ is explosive in the sense that it escapes any bounded domain very quickly. The investigation of the mean exit time $\mathbb{E}[\tau_R(\mathbf{x})]$ also depends on the sign of the Lyapunov exponent.

2.1. **Known results.** For the contractive regime, i.e., for the case when γ_L defined in (2.5) is negative, the asymptotic behavior of the process (2.1) has been investigated extensively over the past decades. We briefly summarize important results before introducing our main contributions.

Existence and uniqueness of stationary measure. The following set of assumptions is commonly assumed in the study of the contractive regime.

Notation 2.2. We denote by $(\mathbf{A},\, \mathbf{B})$ the pair of random matrix and vector having the same distribution with $(\mathbf{A}_n,\, \mathbf{B}_n)$ appeared in Definition 2.1.

Assumption 1 (Standard conditions for contractive regime). *(1) The Lyapunov exponent is negative, i.e., $\gamma_L < 0$.*

(2) There exists $s > 0$ such that $\mathbb{E}\,[\|\mathbf{A}\|^s] < \infty$.

(3) $\mathbb{E}\log^+ |\mathbf{B}| < \infty$ where $\log^+ x := \log(x \vee 1)$.

(4) There is no fixed point in the sense that $\mathbb{P}(\mathbf{A}\mathbf{x} + \mathbf{B} = \mathbf{x}) < 1$ holds for all $\mathbf{x} \in \mathbb{R}^d$.

We will always work under Assumption 1 in the context of a contractive regime. In the next theorem, we recall Kesten's theorem (cf. [6]), which proves that Assumption 1 ensures the existence of a unique stationary distribution.

Theorem 2.3 (Kesten's Theorem). *Under Assumption 1, there exists a unique stationary distribution ν_∞ of the process* (2.1).

Proof. We refer to [6, Theorem 6. page 247] for a proof. ☐

Tail of the stationary measure. An important object in the study of the contractive regime is the function $h_{\mathbf{A}} : [0,\, \infty) \to [0,\, \infty]$ defined by

$$h_{\mathbf{A}}(s) = \lim_{n\to\infty} [\mathbb{E}\|\Pi_n\|^s]^{\frac{1}{n}} \quad ;\; s \ge 0\,. \tag{2.7}$$

whose existence is guaranteed by Kingman's subadditive theorem; see [1, page 167]. We now summarize basic facts regarding this function.

(1) Although in general it is notoriously difficult to compute the exact value of $h_{\mathbf{A}}(s)$, we can readily get lower and upper bounds when $\mathbb{P}(\det \mathbf{A} = 0) = 0$. Inserting (cf. (2.3)) the bound
$$\frac{1}{\|\mathbf{A}_1^{-1}\| \cdots \|\mathbf{A}_n^{-1}\|} \le \|\Pi_n\| \le \|\mathbf{A}_1\| \cdots \|\mathbf{A}_n\|\,,$$
to the definition (2.7), we get
$$\mathbb{E}\frac{1}{\|\mathbf{A}^{-1}\|^s} \le h_{\mathbf{A}}(s) \le \mathbb{E}\|\mathbf{A}\|^s\,, \qquad s \ge 0\,. \tag{2.8}$$

(2) Write[1]
$$\alpha_\infty := \sup\{s : \mathbb{E}\|\mathbf{A}\|^s < \infty\}\,, \tag{2.9}$$
where $\alpha_\infty > 0$ by Assumption 1-(2), so that by (2.8), we have
$$h_{\mathbf{A}}(\beta) < \infty \text{ for all } \beta \in [0,\, \alpha_\infty)\,.$$

(3) By the Hölder inequality, we can readily check that the function $\log h_{\mathbf{A}}$ and hence $h_{\mathbf{A}}$ is convex on $[0,\, \infty)$. Furthermore, it is known that in the case $\gamma_L < 0$, the function $h_{\mathbf{A}}$ is

[1] We note that in all of our application explained in Section 4, the tail of $\mathbf{A}$ is light so that $\alpha_\infty = \infty$.

decreasing at 0 due to Theorem 2.4 below. From this observation, we notice that there are two possible cases as illustrated in Figure 2.1:

(a) Firstly, the function $h_{\mathbf{A}}$ may be a decreasing function as in Figure 2.1-(left), and in this case Goldie-Grübel theorem (e.g., [1, Theorem 2.4.1]) shows that, with several additional technical assumptions, the tail of ν_∞ is exponentially light. This case is of independent interest but will not be handled in the current article.

(b) Another possibility is the case where $h_{\mathbf{A}}(s)$ diverges to $+\infty$ as s increases to $+\infty$ as in Figure 2.1-(right), and this case is handled by Kesten in [6]. We will focus mainly on this case; see Assumption 2. In this case, as $h_{\mathbf{A}}(0) = 1$, by the intermediate value theorem, we have a unique solution to the equation $h_{\mathbf{A}}(s) = 1$ (cf. as in Figure 2.1-(right)).

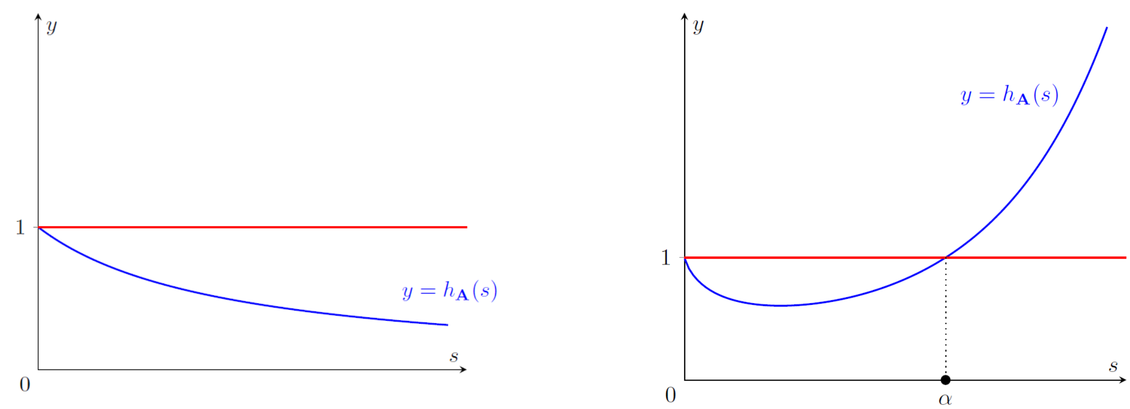

FIGURE 2.1. Tow possible shapes of $h_{\mathbf{A}}$.

Theorem 2.4. *Assume that the Lyapunov exponent γ_L is negative and $h_{\mathbf{A}}(s)$ is finite for some $s > 0$. Then there exists $s_0 > 0$ such that $h_{\mathbf{A}}(s_0) < 1$.*

Proof. We refer to [1, Lemma 4.4.2. page 168] or [6, page 231] for the proofs. □

Assumption 2 (*Condition on* $h_{\mathbf{A}}$). *There exists $s_1 > 0$ such that $h_{\mathbf{A}}(s_1) > 1$. In particular, under the Assumption 1, by the continuity of $h_{\mathbf{A}}$ which follows from its convexity, Theorem 2.4, and the intermediate value theorem, there exists $\alpha \in (0, \alpha_\infty)$ such that $h_{\mathbf{A}}(\alpha) = 1$.*

Remark 2.5 (Remarks on Assumption 2). (1) Throughout our discussion on contractive regime in the current article, the constant $\alpha > 0$ always refer to the one appeared in Assumption 2.

(2) In order to check Assumption 2 for a specific model, by (2.8) and theorem 2.4 above, it suffices to find $s > 0$ such that $\mathbb{E}\frac{1}{\|\mathbf{A}^{-1}\|^s} > 1$. Clearly, the last inequality is guaranteed if and only if $\mathbb{P}(\|\mathbf{A}^{-1}\| < 1) > 0$. Thus, this assumption is easy to check whenever $\mathbf{A}$ is invertible.

(3) By the convexity of $h_{\mathbf{A}}$, we have $h_{\mathbf{A}} < 1$ on $(0, \alpha)$ and $h_{\mathbf{A}} > 1$ on (α, ∞).

Kesten [6] proved that under Assumptions 1 and 2, along with several technical assumptions, the stationary distribution ν_∞ has a heavy-tailed distribution with a power-law tail of exponent α given in Assumption 2. More precisely, it is proven that, there exists a positive function $g : \mathbb{S}^{d-1} \to (0, \infty)$ such that, for all $\mathbf{u} \in \mathbb{S}^{d-1}$,

$$\lim_{z\to\infty} z^\alpha \nu_\infty(\{\mathbf{x} : \mathbf{u}\cdot\mathbf{x} > z\}) = g(\mathbf{u}), \tag{2.10}$$

where

$$\mathbb{S}^{d-1} := \left\{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\right\}. \tag{2.11}$$

This result, known as Kesten's theorem for process (2.1), is interesting in that the stationary distribution is power-law tailed even in the case where $\mathbf{A}$ and $\mathbf{B}$ are light-tailed (e.g., normal-type distribution). An interested reader may refer to the original paper [6, Theorem 6] or textbook [1, page 169-171] for more details.

3. Main Results

3.1. **Main results for the contractive regime.**

This section presents our main result on the contractive regime, which provides asymptotics of the mean exit time $\mathbb{E}[\tau_R(\mathbf{x})]$ for $\mathbf{x} \in \mathbb{R}^d$ in the regime $R \to \infty$.

Technical assumptions. Our main result, stated in Theorem 3.4, requires additional assumptions. Kesten's theorem [6, Theorem 6] provides a set of sufficient conditions. However, working with them require further introduction of definitions and notations, and the verification of those assumptions are difficult for the example we consider. Instead, we provide a set of assumptions which are concise and easy to confirm for the examples given in Section 4. Our assumptions are weaker than those of [6, Theorem 6]; see Remarks 3.1, 3.2, and 3.3.

The first additional assumption is on the unboundedness of the support of the stationary distribution.

Assumption 3. *The support of ν_∞ is unbounded.*

This assumption is obviously necessary in our context. Without this, the process (2.1) is confined to the support of ν_∞, and hence, the mean of the exit time from a domain $\mathcal{B}_R$ with sufficiently large R would be $+\infty$.

Remark 3.1 (Remarks on Assumption 3). (1) In view of (2.10), the assumptions of [6, Theorem 6] imply Assumption 3. In particular, the assumption of [1, 6, Condition (A) in Section 4.4] regarding the denseness of the additive subgroup of $\mathbb{R}$ generated by the log of absolute value of largest eigenvalue of Π_n implies Assumption 3.

(2) Note that, when the support of a random vector $\mathbf{A}\mathbf{x} + \mathbf{B}$ is unbounded for all $\mathbf{x} \in \mathbb{R}^d$, then it is immediate that the Assumption 3 holds. This is indeed the case for the models we discuss in Section 4. More detailed scheme to verify Assumptipon 3 when the support

of $\mathbf{A}$ and $\mathbf{B}$ are bounded is explained in [1, Proposition 4.3.1, page 160] at which a characterization of the support ν_∞ is provided.

The following technical assumption prevents the effect of single outcome of random variable $\mathbf{A}, \mathbf{B}$ from dominating the long-term dynamics of (2.1). Again, it is weaker than assumptions of [6, Theorem 6].

Assumption 4. *There exist R_0, $z_0 > 1$, $C_0 > 0$, and $\alpha_+ \in (\alpha, \alpha_\infty]$ such that, we have*

$$\mathbb{P}\big(|\mathbf{A}\mathbf{x} + \mathbf{B}| > zR\big) \le \frac{C_0}{z^{\alpha_+}} \mathbb{P}\big(|\mathbf{A}\mathbf{x} + \mathbf{B}| > R\big) \tag{3.1}$$

for all $R \in [R_0, \infty)$, $z \in [z_0, \infty)$ and $\mathbf{x} \in \mathcal{B}_R$.

This assumption guarantees that $\mathbf{A}\mathbf{x} + \mathbf{B}$ has tails strictly lighter than power-law distribution with index α. If it has thicker tails than power-law distribution with index α, the behavior of the process (2.1) is simply dominated by large deviations of $\mathbf{A}$ or $\mathbf{B}$, rather than the structural properties of stochastic recurrence equations.

Remark 3.2 (Remarks on Assumption 4). (1) We say that the random matrix $\mathbf{A}$ or the random vector $\mathbf{B}$ has a tail lighter than the power-law distribution of index α if we can find R_1, $z_1 > 1$, $C_1 > 0$, and $\alpha_1 > \alpha$ such that

$$\mathbb{P}(|\mathbf{A}\mathbf{x}| > Rz) \le \frac{C_1}{z^{\alpha_1}} \mathbb{P}(|\mathbf{A}\mathbf{x}| > R) \quad \text{or} \quad \mathbb{P}(|\mathbf{B}| > Rz) \le \frac{C_1}{z^{\alpha_1}} \mathbb{P}(|\mathbf{B}| > R) \tag{3.2}$$

for all $R \in [R_1, \infty)$, $z \in [z_1, \infty)$ and $\mathbf{x} \in \mathbb{S}^{d-1}$. Then, one can conjecture that Assumption 4 is valid when both $\mathbf{A}$ and $\mathbf{B}$ have tails lighter than power-law distribution of index α.

(2) Considering the case $\mathbf{x} = \mathbf{0}$ in (3.1), we note that Assumption 4 implies that the random vector $\mathbf{B}$ has a tail lighter than power-law distribution of index α. In particular, by applying the layer-cake formula, we get

$$\mathbb{E}|\mathbf{B}|^\gamma < \infty \quad \text{for all } \gamma \in [0, \alpha_+) . \tag{3.3}$$

The final assumption concerns the non-degeneracy of the process (2.1).

Assumption 5. *There exists $n_0 \in \mathbb{Z}^+$ such that, for all $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ (cf. (2.4))*

$$\mathbb{P}\left[\mathbf{x} \cdot \Pi_{n_0} \mathbf{y} = 0\right] < 1 .$$

Remark 3.3 (Remarks on Assumption 5). (1) Of course, this assumption with $n_0 = 1$ holds if $\mathbf{A}$ is almost surely invertible, which is one of the assumptions of [6, Theorem 6]. We note that our models in Section 4 satisfy this assumption.

(2) On the other hand, our assumption is strictly weaker than this invertibility assumption. For instance, let $d = 2$ and $\mathbf{A} = \begin{pmatrix} c & cZ \\ cZ & cZ^2 \end{pmatrix}$ where Z is a standard normal random variables, and $c > 0$ is a constant small enough to guarantee that the Lyapunov exponent associated with $\mathbf{A}$ is negative. Then $\mathbf{A}$ is always non-invertible, while for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, we have $\mathbb{P}(\mathbf{x} \cdot \mathbf{A}\mathbf{y} = 0) = 0$ so that Assumption 5 is satisfied with $n_0 = 1$.

Main results. Now we are ready to state main results. We assume in the next theorem that Assumptions 1, 2, 3, 4 and 5 are in force. We emphasize that, taken together, all of these assumptions are weaker than those of Kesten's theorem.

Theorem 3.4. *For all $\mathbf{x}_0 \in \mathbb{R}^d$, it holds that*

$$\lim_{R\to\infty} \frac{\log \mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} = \alpha \quad \text{for all } \mathbf{x}_0 \in \mathbb{R}^d ,$$

where $\alpha > 0$ is the exponent appeared in Assumption 2.

The proof of Theorem 3.4, given in Section 6, is based on the refined construction of suitable sub- and super-martingales. More precisely, we prove in Propositions 6.1 and 6.2 that, for all $\gamma_1 \in (0, \alpha)$ and $\gamma_2 \in (\alpha, \alpha_\infty)$ where α_∞ is the one defined in (2.9), we find constants $C_{\gamma_1}, C'_{\gamma_1}, C_{\gamma_2} > 0$ such that for all $\mathbf{x}_0 \in \mathbb{R}^d$ and $R > 0$,

$$C_{\gamma_1} R^{\gamma_1} - C'_{\gamma_1} \left(|\mathbf{x}_0|^{\gamma_1} + 1\right) \le \mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \le C_{\gamma_2} \left(R^{\gamma_2} + 1\right) . \tag{3.4}$$

Then, Theorem 3.4 is a direct consequence of this bound.

Univariate case. Theorem 3.4 implies that $\mathbb{E}[\tau_R(\mathbf{x}_0)]$ grows on the order of R^α possibly multiplied by a sub-polynomial prefactor. One can also expect more precise estimate, e.g.,

$$\lim_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{R^\alpha} = C. \tag{3.5}$$

Obtaining such a sharp estimate in full generality under the current set of assumptions seems to be very difficult, as proving a bound of the form (3.4) is not feasible at this time for $\gamma_1 = \alpha$ or $\gamma_2 = \alpha$. However, for the univariate case, namely the case with $d = 1$, we can get a stronger result close to (3.5). The main benefit of the univariate case is that the the Lyapunov exponent γ_L and the function $h_{\mathbf{A}}(\cdot)$ simplify to

$$\gamma_L = \mathbb{E}[\log|\mathbf{A}|] \quad \text{and} \quad h_{\mathbf{A}}(s) = \mathbb{E}|\mathbf{A}|^s \tag{3.6}$$

as (2.3) now becomes an equality as the norm $\|\cdot\|$ is reduced to the absolute value of a real number. Thanks to this simplification, we get the follwing refined result of 3.4.

Theorem 3.5. *Suppose that $d = 1$ and that $\alpha \ge 2$. Then, for all $\mathbf{x}_0 \in \mathbb{R}$, we have that*

$$0 < \liminf_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{R^\alpha} \le \limsup_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{R^\alpha} < \infty . \tag{3.7}$$

Remark 3.6. We note that condition $\alpha \ge 2$ is equivalent to $\mathbb{E}\left[\mathbf{A}^2\right] \le 1$.

The proof of Theorem 3.4 is given in Section 6.6. We note that, the proof therein shows the lower bound

$$0 < \liminf_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{R^\alpha}$$

for all $\alpha > 0$, and the condition $\alpha \ge 2$ is required only in the proof of the last inequality of 3.7. We believe that for $\alpha \ge 2$, the estimate of the form 3.5 holds, but for $\alpha < 2$, we are not able to exclude the possibility of $\mathbb{E}[\tau_R(\mathbf{x}_0)] \simeq \kappa(R)R^\alpha$ where $\kappa(R)$ is a sub-polynomial prefactor satisfying $\kappa(R) \to +\infty$ as $R \to \infty$ slower than any polynomial.

3.2. **Main results for the explosive regime.** Now we consider the explosive regime, i.e., the case when the Lyapunov exponent of γ_L is positive, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \log \|\Pi_n\| = \gamma_L > 0 \tag{3.8}$$

almost surely. To estimate the mean of the exit time under this condition, we impose the following assumptions.

Assumption 6 (Assumptions for explosive regime). *(1)* $\mathbb{P}(\mathbf{A}$ *is singular*$) = 0$ *and moreover* $\mathbf{A}$ *is irreducible, i.e. there is no proper nontrivial subspace* V *of* $\mathbb{R}^d$ *such that* $\mathbf{A}V \subset V$ *almost surely.*

(2) We have that

$$\inf_{\mathbf{x}\in\mathbb{R}^d} \mathbb{E} \log |\mathbf{A}\mathbf{x} + \mathbf{B} - \mathbf{x}| > -\infty \,. \tag{3.9}$$

(3) There exist R_0, $z_0 > 1$, $C_0 > 0$, *and* $\beta_0 > 1$ *such that, we have*

$$\mathbb{P}\left[|\mathbf{A}\mathbf{x} + \mathbf{B}| > zR\right] \le \frac{C_0}{(\log z)^{\beta_0}} \mathbb{P}\left[|\mathbf{A}\mathbf{x} + \mathbf{B}| > R\right] \tag{3.10}$$

for all $R \in [R_0, \infty)$, $z \in [z_0, \infty)$ *and* $\mathbf{x} \in \mathcal{B}_R$. *[This is a much weaker version of Assumption 4 of the contractive regime]*

Part (1) of the previous assumption is standard in the study of Kesten's stochastic recurrence equation (2.1), as most models trivially satisfy it. We will explain the meaning of part (2) and (3) of Assumption 6 after stating the main theorem. In the explosive regime, the norm $\|\Pi_n\| = \|\mathbf{A}_n \cdots \mathbf{A}_1\|$ exponentially diverges to $+\infty$, we can guess that the exit time is logarithmically small, and the next result confirm that it is indeed the case. Of course, we assume (3.8) and Assumption 6 in the next theorem.

Theorem 3.7. *For all* $\mathbf{x}_0 \in \mathbb{R}^d$, *we have that*

$$\frac{1}{\gamma_L} \le \liminf_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} \le \limsup_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} < \infty \,.$$

The proof of this theorem is given in Section 7. In particular, we prove in Proposition 7.5 a strong lower bound of the form

$$\liminf_{R\to\infty} \frac{\tau_R(\mathbf{x}_0)}{\log R} \ge \frac{1}{\gamma_L}$$

which holds almost surely. Then, the lower bound given in Theorem 3.7 is a direct consequence of Fatou's lemma. The proof of upper bound is quite technical and again based on highly complicated construction of a sub-martingale which leads us an estimate of the form

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \le \kappa_1(1 + \log R) - \kappa_2 \log^+ |\mathbf{x}_0|$$

for some constants κ_1, $\kappa_2 > 0$. We further note that we can even compute these two constants explicitly.

Remark 3.8 (Remarks on Assumption 6 and Theorem 3.7). (1) If part (2) of Assumption 6 is violated, the escape from the domain $\mathcal{B}_R$ may take much longer time scale, esepcially when the dynamical system starts from $\mathbf{x}_0$ such that

$$\mathbb{E}\log|\mathbf{A}\mathbf{x}_0 + \mathbf{B} - \mathbf{x}_0| = -\infty\ . \tag{3.11}$$

In particular, (3.11) implies that $\mathbf{X}_1(\mathbf{x}_0)$ is very close to $\mathbf{x}_0$, and indicates that it might be very hard to escape from a neighborhood of $\mathbf{x}_0$.

(2) In view of Theorem 3.7, we confirmed that the escaping time from a domain $\mathcal{B}_R$ is of $O(\log R)$. However, if the random vector $\mathbf{A}\mathbf{x} + \mathbf{B}$ for some $\mathbf{x}$ is super-heavy tail of the form $\mathbb{P}\left[|\mathbf{A}\mathbf{x} + \mathbf{B}| \geq z\right] \sim \frac{1}{(\log z)^\gamma}$ for some $\gamma < 1$, then the escape from the domain is not governed by the structure of the process (2.1) but by a single large outcome and hence the analysis is trivialized significantly. For this reason we assume part (3) of Assumption 6 to investigate the structural property of the process (2.1).

(3) We conjecture that it holds that

$$\lim_{R\to\infty} \frac{\mathbb{E}[\tau_R(\mathbf{x}_0)]}{\log R} = \frac{1}{\gamma_L}$$

as the multiplicative factor $\mathbf{A}_n$ satisfies, in the sense of (3.8), $\|\mathbf{A}_n \cdots \mathbf{A}_1\| \sim e^{\gamma_L n}$ and hence we can guess that $|\mathbf{X}_n|$ grows as $e^{\gamma_L n}$. Our lower bound is in coincidence with this conjecture, while getting a tight upper bound leading this result seems to be very difficult at this moment and probably requires additional assumption on the behavior of $(\mathbf{A}, \mathbf{B})$ for the reason indicated in part (1) of the current remark.

(4) If $\mathbb{P}(\mathbf{A}$ is singular$) = 0$ is satisfied, Assumption 5 implies (1) of Assumption 6. Indeed, suppose that there exists such $V \subset \mathbb{R}^d$. Taking any nonzero vectors $v \in V$ and $v^\perp \in V^\perp$, it holds that

$$\mathbb{P}(v^\perp \cdot \Pi_{n_0} v = 0) = 1\ ,$$

violating Assumption 5. Also, as mentioned in the assumption, (3) of Assumption 6 is a direct consequence of Assumption 4. Hence, if Assumptions 4 and 5 are guaranteed, only (2) of Assumption 6 is left to be shown to fulfill the whole Assumption for explosive regime.

4. Applications

Before delving into the main proofs, we will first introduce some significant examples that emerge in various fields such as machine learning or economics that fall into the framework explained in the previous section.

4.1. **Stochastic gradient descent on a quadratic loss function.**

Vanilla Mini-batch stochastic gradient descent. The first example appears in the field of machine learning, as observed in [4]. To be concrete, the stochastic gradient descent (SGD) method with mini-batch runs on a quadratic loss function has a form of (2.1). The quadratic loss function $F : \mathbb{R}^d \to \mathbb{R}$ for this model is defined by

$$F(\mathbf{x}) = \frac{1}{2}\mathbb{E}_{(\mathbf{a},\, b)}\left[(\mathbf{a}\cdot\mathbf{x} - b)^2\right]$$

where the data $(\mathbf{a},\, b) \in \mathbb{R}^d \times \mathbb{R}$ is a random element with unknown distribution μ. We assume at this moment that we can sample the random data $(\mathbf{a},\, b)$ according to μ although we do not know exactly what the distribution is. Then ,the mini-batch SGD finding approximate minimum of F with batch size m is conducted as follows.

(1) We set $\mathbf{X}_0 = \mathbf{x}_0 \in \mathbb{R}^d$ any initial point.
(2) At each nth step, $n \in \mathbb{Z}^+$, sample m independent data $(\mathbf{a}_1^{(n)},\, b_1^{(n)}),\ \dots,\ (\mathbf{a}_m^{(n)},\, b_m^{(n)})$ (according to the law μ) and then define

$$F_n(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}(\mathbf{a}_i^{(n)}\cdot\mathbf{x} - b_i^{(n)})^2 \ .$$

(3) Perform the gradient descent with learning rate $\eta > 0$ to the approximating function F_n, i.e.,

$$\mathbf{X}_{n+1} = \mathbf{X}_n - \eta\nabla F_{n+1}(\mathbf{X}_n) \ . \tag{4.1}$$

Then, a simple computation shows that the (4.1) can be re-written as

$$\mathbf{X}_{n+1} = \mathbf{A}_{n+1}\mathbf{X}_n + \mathbf{B}_{n+1} \quad ;\ n \in \mathbb{Z}_0^+$$

where

$$\mathbf{A}_n = \mathbf{I}_d - \frac{\eta}{m}\sum_{i=1}^{m}\mathbf{a}_i^{(n)}(\mathbf{a}_i^{(n)})^\dagger \quad \text{and} \quad \mathbf{B}_n = \frac{\eta}{m}\sum_{i=1}^{m} b_i^{(n)}\mathbf{a}_i^{(n)} \ . \tag{4.2}$$

Here, $\mathbf{I}_d$ denotes $d \times d$ identity matrix and $\mathbf{v}^\dagger$ denotes the transpose of $\mathbf{v}$ and therefore $\mathbf{a}_i^{(n)}(\mathbf{a}_i^{(n)})^\dagger$ denotes a $d \times d$ matrix.

By assuming that the data are normal as in [4] , i.e.,

$$\mathbf{a}\sim\mathcal{N}(\mathbf{0},\, \Sigma) \quad \text{and} \quad b \sim \mathcal{N}(0,\, \sigma_B^2) \tag{4.3}$$

for some $\sigma_B > 0$, we can verify that all the assumptions of the previous section holds. We refer to Appendix B for a brief explanation.

An interesting remark is that, by tuning learning rate η from a large number to small one as in the real application of this algorithm, the lyapunov exponent decrease from positive number to the negative number. The algorithm will be stabilized and find the minimum when we take η sufficiently small, (but not that small so that the algorithm is way too slow) so that the Lyapunov exponenent associated with the matrix $\mathbf{A}_n$ becomes negative.

Mini-batch stochastic gradient descent with momentum. We can sometimes accelerate the mini-batch SGD algorithim by exploiting the momentum (reference?). This algorithm is a variant of the vanilla SGD explained in (4.1). All the settings of mini-batch SGD with momentum are identical to the one with vanilla one, but we update not only $\mathbf{X}_n$ but also momentum $\mathbf{V}_n$ together. For a parameter $\gamma \in [0,\, 1)$ tuning the strength of the momentum, we update $(\mathbf{X}_{n+1},\, \mathbf{V}_{n+1})$ by

$$\begin{cases} \mathbf{X}_{n+1} = \mathbf{X}_n - \eta \mathbf{V}_{n+1} \ , \\ \mathbf{V}_{n+1} = \gamma \mathbf{V}_n + (1-\gamma) \nabla F_{n+1}(\mathbf{X}_n) \ . \end{cases}$$

Note that the case $\gamma = 0$ corresponds to the vanilla mini-batch SGD. Then, writing

$$\mathbf{Y}_n := \begin{pmatrix} \mathbf{X}_n \\ \mathbf{V}_n \end{pmatrix} \ , \quad \mathbf{C}_n = \begin{pmatrix} \mathbf{I}_d - \eta(1-\gamma)\mathbf{A}_n & -\eta\gamma \mathbf{I}_d \\ (1-\gamma)\mathbf{A}_n & \gamma \mathbf{I}_d \end{pmatrix}, \quad \mathbf{D}_n = \begin{pmatrix} -\eta(1-\gamma)\mathbf{B}_n \\ (1-\gamma)\mathbf{B}_n \end{pmatrix} \ ,$$

where $\mathbf{A}_n$ and $\mathbf{B}_n$ are the ones defined in (4.2), we get

$$\mathbf{Y}_{n+1} = \mathbf{C}_{n+1}\mathbf{Y}_n + \mathbf{D}_{n+1} \quad ; \ n \in \mathbb{Z}_0^+$$

and therefore $(\mathbf{Y}_n)_{n \in \mathbb{Z}_0^+}$ can be regarded as a $2d$-dimensional process (2.1). Again, by assuming (4.3), we can check that the process $(\mathbf{Y}_n)_{n \in \mathbb{Z}_0^+}$ also satisfies the assumptions of Section 3. We will explain this in a companion paper of the current article.

4.2. **Time series models.** The remaining examples refer to financial models. As mentioned in [1, Chapter 1], the ARCH(p) and GARCH($1, q$) are well known examples of stochastic recurrence equation in econometrics or financial engineering.

Autoregressive conditional heteroskedasticity (ARCH) model. ARCH(p) is a time series modeling log-returns of assets. More precisely, in this model, the speculative prices of certain asset at time t is denoted by $P_t > 0$ and define the log-return of P_t as

$$X_t = \log \frac{P_{t+1}}{P_t} \quad ; \ t \in \mathbb{Z}^+ \ .$$

Then, for each $p \in \mathbb{Z}^+$, ARCH(p) models the behavior of X_t as

$$X_t = \sigma_t Z_t \quad ; \ t \geq p \tag{4.4}$$

where $(Z_t)_{t \in \mathbb{Z}^+}$ is an i.i.d. sequence of random variables with mean zero and variance one, and where

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_t X_{t-i}^2 \tag{4.5}$$

for some non-negative constants α_0, α_1, $\dots$, $\alpha_p \geq 0$ such that $\alpha_0 \alpha_p > 0$. We can re-interpret this equation as

$$\begin{pmatrix} X_{t+1}^2 \\ X_t^2 \\ X_{t-1}^2 \\ \vdots \\ X_{t-p+2}^2 \end{pmatrix} = \begin{pmatrix} \alpha_1 Z_{t+1}^2 & \alpha_2 Z_{t+1}^2 & \cdots & \alpha_{p-1} Z_{t+1}^2 & \alpha_p Z_{t+1}^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} X_t^2 \\ X_{t-1}^2 \\ X_{t-2}^2 \\ \vdots \\ X_{t-p+1}^2 \end{pmatrix} + \begin{pmatrix} \alpha_0 Z_{t+1}^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and therefore, by letting

$$\mathbf{X}_t = \begin{pmatrix} X_t^2 \\ X_{t-1}^2 \\ X_{t-2}^2 \\ \vdots \\ X_{t-p+1}^2 \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 & \alpha_2 Z_t^2 & \cdots & \alpha_{p-1} Z_t^2 & \alpha_p Z_t^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 Z_t^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.6}$$

we can write the ARCH(p) model exactly the same as (2.1).

Note that one might be interested in the exit time of X_t itself, instead of $\mathbf{X}_t$. This is not an issue since, if we denote by τ_R and $\widehat{\tau}_R$ the exit time of the process $\mathbf{X}_t$ and X_t^2, respectively (disregarding the initial location at this moment), we have that, for all $R > 0$

$$\tau_R \leq \widehat{\tau}_R \leq \tau_{\sqrt{p}R} + p ,$$

and therefore we can deliver the estimate of τ_R directly to that of $\widehat{\tau}_R$.

If Z is assumed to be normally distributed and $\alpha_1 \neq 0$, we proved that this model satisfies all the assumptions for our result. The proof is presented in the appendix A.

Generalized autoregressive conditional heteroskedasticity (GARCH) model. For $q \in \mathbb{Z}^+$, GARCH(1, q) is a variation of ARCH and defined by a sequence $(X_t)_{t\in\mathbb{Z}^+}$ of random variables as in (4.4) for $t \geq q$ with the same sequence $(Z_t)_{t\in\mathbb{Z}^+}$ of random variables and with a new relation for σ_t:

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 ,$$

for some non-negative constants α_0, α_1, β_1, ..., $\beta_q \geq 0$ such that $\alpha_0\alpha_1\beta_q > 0$. This equation can be rewritten as:

$$\begin{pmatrix} \sigma_{t+1}^2 \\ \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-q+2}^2 \end{pmatrix} = \begin{pmatrix} \alpha_1 Z_{t+1}^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \sigma_{t-2}^2 \\ \vdots \\ \sigma_{t-q+1}^2 \end{pmatrix} + \begin{pmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and therefore, by letting

$$\mathbf{X}_t = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \sigma_{t-2}^2 \\ \vdots \\ \sigma_{t-q+1}^2 \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.7}$$

we again get the expression of (2.1) for the GARCH model.

Similar to the ARCH model, we proved that the assumptions holds true if Z is normally distributed, see appendix.

5. Preliminary Results for contractive Regime

We study in this and the next section the contractive regime. In particular, we shall always assume that $\gamma_L < 0$ and that Assumptions 1, 2, 3, 4, and 5 throughout Sections 5 and 6.

The proof of Theorems 3.4 and 3.5 are given in the next section. In this section, we provide several preliminary results required in the proof given in the next section.

5.1. **Coupling.** Recall that $(\Omega, \mathcal{E}, \mathbb{P})$ denotes the probability space containing $\{(\mathbf{A}_n, \mathbf{B}_n) : n \in \mathbb{Z}^+\}$. Since the process $(\mathbf{X}_t(\mathbf{x}))_{t\geq 0}$ is completely determined by outcomes of $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$ and its starting location $\mathbf{X}_0$, we can couple all the processes $(\mathbf{X}_n(\mathbf{y}))_{n\in\mathbb{Z}_0^+}$, $\mathbf{y} \in \mathbb{R}^d$, at the same probability space. Note that, under this coupling, all the processes $(\mathbf{X}_n(\mathbf{y}))_{n\in\mathbb{Z}_0^+}$, $\mathbf{y} \in \mathbb{R}^d$ share the outcomes $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$. From this moment on, when we consider several processes start at different starting point, we shall assume that they are always coupled throughout this manner.

Note that, we can deduce by direct computation from (2.1) that

$$\mathbf{X}_n(\mathbf{x}) = \Pi_n\mathbf{x} + \mathbf{A}_2 \cdots \mathbf{A}_n\mathbf{B}_1 + \cdots + \mathbf{A}_n\mathbf{B}_{n-1} + \mathbf{B}_n \ . \tag{5.1}$$

Thus, under the coupling explained above, for all $n \in \mathbb{Z}^+$ and $\mathbf{y}, \mathbf{z} \in \mathbb{R}^d$, we have

$$\mathbf{X}_n(\mathbf{y}) - \mathbf{X}_n(\mathbf{z}) = \Pi_n(\mathbf{y} - \mathbf{z}) \ . \tag{5.2}$$

Remark 5.1. We shall assume in this section Assumptions 1, 2, and 3 so that by Theorem 2.3 there exists $\alpha > 0$ satisfying $h_{\mathbf{A}}(\alpha) = 1$ and there exists a stationary measure ν_∞ of (2.1). Also, for the contractive case, it is known that the infinite series

$$\sum_{n=1}^{\infty} \mathbf{A}_1 \cdots \mathbf{A}_{n-1}\mathbf{B}_n$$

converges to a random variable $\mathbf{R}$ with distribution ν_∞ almost surely. see [6, page 235]. This representation will be used to prove Lemma 5.3.

5.2. **Finiteness of exit time.** In the contractive regime, since the process $\mathbf{X}_t$ tends to $\mathbf{0}$ due to contracting force, and therefore it is not even clear whether the exit time τ_R for large R is finite or not. In this section, we prove that for any $R > 0$, the exit time τ_R is finite almost surely even in the contractive regime.

Remark 5.2. Before starting, we will summarize handy facts regarding the evolution of the moment of $\|\Pi_n\|$ which will be used frequently in the argument given in Sections 5 and 6. Namely, by Remark 2.5-(3), we have that

$$\begin{cases} \lim_{n\to\infty} \mathbb{E}\|\Pi_n\|^\gamma = 0, & \gamma \in (0, \alpha) \ , \\ \lim_{n\to\infty} \mathbb{E}\|\Pi_n\|^\gamma = \infty. & \gamma \in (\alpha, \infty) \ . \end{cases} \tag{5.3}$$

Moreover, for all $\gamma \in (\alpha, \infty)$, we can find $n_\gamma \in \mathbb{Z}_0^+$ and $\delta_\gamma > 0$ such that

$$\mathbb{E}\|\Pi_n\|^\gamma \geq (1 + \delta_\gamma)^n \quad \text{for all } n > n_0 \ . \tag{5.4}$$

We first nvestigate the tail of the stationary measure ν_∞ (cf. Remark 5.1). Note that we did not assume the requirements of Kesten's theorem, and therefore the next lemma cannot be obtained from it.

Lemma 5.3. *For all $\beta \in (0, \alpha)$, we have that*

$$\int_{\mathbb{R}^d} |\mathbf{x}|^\beta \nu_\infty(d\mathbf{x}) < \infty . \tag{5.5}$$

Proof. Let us fix $\beta \in (0, \alpha)$. By Remark 5.2, we can take $n_1 \in \mathbb{Z}^+$ such that

$$\mathbb{E}\|\Pi_{n_1}\|^\beta < 1 . \tag{5.6}$$

For $n \in \mathbb{Z}^+$, define

$$\mathbf{R}_n = \sum_{k=0}^{n} \Pi_{kn_1} \mathbf{W}_k$$

where, for $k \in \mathbb{Z}_0^+$,

$$\mathbf{W}_k := \mathbf{B}_{kn_1+1} + \mathbf{A}_{kn_1+1}\mathbf{B}_{kn_1+2} + \cdots + \mathbf{A}_{kn_1+1} \cdots \mathbf{A}_{(k+1)n_1-1}\mathbf{B}_{(k+1)n_1}$$

so that by Remark 5.1, the series $(\mathbf{R}_n)_{n\in\mathbb{Z}^+}$ converges almost surely to a random variable $\mathbf{R}$ with distribution ν_∞.

By the elementary inequality

$$(x_1 + \cdots + x_n)^r \le n^r (x_1^r + \cdots + x_n^r) \tag{5.7}$$

which holds for all $x_1, \ldots, x_n, r \ge 0$, the submultiplicativeness of the matrix norm, and independence of sequence $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$, we have

$$\mathbb{E}|\mathbf{W}_k|^\beta \le n_1^\beta \sum_{i=0}^{n_1-1} \left(\mathbb{E}\|\mathbf{A}\|^\beta\right)^i \mathbb{E}|\mathbf{B}|^\beta < \infty , \tag{5.8}$$

where $\mathbb{E}\|\mathbf{A}\|^\beta < \infty$ follows from Assumption 2. For d-dimensional random vector $\mathbf{X}$, we define the norm $\|\cdot\|_\beta$ by

$$\|\mathbf{X}\|_\beta := \left(\mathbb{E}|\mathbf{X}|^\beta\right)^{\max\{1,\, 1/\beta\}} .$$

Then, by the triangle inquality and the independence of random variables, we obtain

$$\|\mathbf{R}_n\|_\beta \le \sum_{k=0}^{n} \|\Pi_{kn_1}\mathbf{W}_k\|_\beta \le \sum_{k=0}^{n} \|\Pi_{n_1}\|_\beta^k \|\mathbf{W}_1\|_\beta \le \frac{\|\mathbf{W}_1\|_\beta}{1 - \|\Pi_{n_1}\|_\beta} .$$

where the last inequality follows from $\|\Pi_{n_1}\|_\beta < 1$ which comes from (5.6). Hence, by Fatou's inequality, we get

$$\|\mathbf{R}\|_\beta \le \liminf_{n\to\infty} \|\mathbf{R}_n\|_\beta \le \frac{\|\mathbf{W}_1\|_\beta}{1 - \|\Pi_{n_1}\|_\beta} < \infty$$

which leads to (5.5). $\square$

The next lemma regarding the ergodic behavior of the process (2.1) is a consequence of the previous one.

Lemma 5.4. *The followings hold.*

(1) For all $\psi \in C_c^\infty(\mathbb{R}^d)$ and a Borel probability measure π, we have

$$\lim_{n\to\infty} \mathbb{E}\psi(\mathbf{X}_n(\pi)) = \int_{\mathbb{R}^d} \psi d\nu_\infty \ .$$

(2) For all $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathbb{Z}^+$, we have $\mathbb{P}(\mathbf{X}_n(\mathbf{x}) = \mathbf{x}) < 1$.

Proof. (1) Let us fix $\psi \in C_c^\infty(\mathbb{R}^d)$. By writing $\alpha_0 = \min\left\{\frac{\alpha}{2}, 1\right\}$ where α as defined in 2.3, we deduce from (5.3) that

$$\lim_{n\to\infty} \mathbb{E}\|\Pi_n\|^{\alpha_0} = 0 \ . \tag{5.9}$$

Since ψ is compactly supported, it is straightforward that ψ is Hölder continuous with exponent α_0, i.e., there exists $M_\psi > 0$ such that

$$|\psi(\mathbf{x}) - \psi(\mathbf{y})| \le M_\psi |\mathbf{x} - \mathbf{y}|^{\alpha_0} \quad \text{for all } \mathbf{x},\, \mathbf{y} \in \mathbb{R}^d \ ,$$

and therefore, for any $\mathbf{y}$, $\mathbf{z} \in \mathbb{R}^d$, by (5.2), we have

$$|\mathbb{E}\left[\psi(\mathbf{X}_n(\mathbf{y})) - \psi(\mathbf{X}_n(\mathbf{z}))\right]| \le M_\psi \mathbb{E}\left[|\mathbf{X}_n(\mathbf{y}) - \mathbf{X}_n(\mathbf{z})|^{\alpha_0}\right] \le M_\psi |\mathbf{y} - \mathbf{z}|^{\alpha_0} \cdot \mathbb{E}\|\Pi_n\|^{\alpha_0} \ .$$

Hence, for each $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned} \left|\mathbb{E}\psi(\mathbf{X}_n(\mathbf{x})) - \int_{\mathbb{R}^d} \psi \, d\nu_\infty\right| &= \left|\int_{\mathbb{R}^d} \mathbb{E}\left[\psi(\mathbf{X}_n(\mathbf{x})) - \psi(\mathbf{X}_n(\mathbf{y}))\right] \nu_\infty(d\mathbf{y})\right| \\ &\le M_\psi \, \mathbb{E}\|\Pi_n\|^{\alpha_0} \int_{\mathbb{R}^d} |\mathbf{x} - \mathbf{y}|^{\alpha_0} \nu_\infty(d\mathbf{y}) \end{aligned}$$

where the first identity uses the stationarity of ν_∞. Since $\alpha_0 < \alpha$, by Lemma 5.3 and (5.9), we get

$$\lim_{n\to\infty} \mathbb{E}\psi(\mathbf{X}_n(\mathbf{x})) = \int_{\mathbb{R}^d} \psi d\nu_\infty \quad \text{for all } \mathbf{x} \in \mathbb{R}^d \ .$$

Thus, the conclusion of the lemma follows from the dominated convergence theorem as ψ is bounded.

(2) Suppose that there exists $n_0 \in \mathbb{Z}^+$ and $\mathbf{x}_0 \in \mathbb{R}^d$ such that $\mathbb{P}(\mathbf{X}_{n_0}(\mathbf{x}_0) = \mathbf{x}_0) = 1$ so that we have

$$\mathbb{P}(\mathbf{X}_{kn_0}(\mathbf{x}_0) = \mathbf{x}_0) = 1 \quad \text{for all } k \in \mathbb{Z}^+ \ . \tag{5.10}$$

Hence, for $\psi \in C_c^\infty(\mathbb{R}^d)$, we have

$$\lim_{k\to\infty} \mathbb{E}\left[\psi(\mathbf{X}_{kn_0}(\mathbf{x}_0))\right] = \int_{\mathbb{R}^d} \psi \, d\nu_\infty \ .$$

Thus, by (5.10), we have $\int_{\mathbb{R}^d} \psi \, d\nu_\infty = \psi(\mathbf{x}_0)$. Namely, ν_∞ is the Dirac measure at $\mathbf{x}_0$. This contradicts to Assumption 1-(3). □

Now we are ready to prove the finiteness of exit time.

Proposition 5.5. *For all $\mathbf{x} \in \mathbb{R}^d$ and $R > 0$, we have $\tau_R(\mathbf{x}) < \infty$ almost surely.*

Proof. We fix $\mathbf{x} \in \mathbb{R}^d$ and $R > 0$, and suppose on the contrary that

$$\mathbb{P}\left[\tau_R(\mathbf{x}) = \infty\right] = p_0 > 0 \ . \tag{5.11}$$

It is immediate from (5.11) that $\mathbf{x} \in \mathcal{B}_R$.

[Step 1] We first recursively define some probabilities $(q_n)_{n\in\mathbb{Z}^+}$ and distributions $(\mu_n)_{n\in\mathbb{Z}^+}$. First, we set

$$q_1 = \mathbb{P}\left[\mathbf{A}\mathbf{x} + \mathbf{B} \in \mathcal{B}_R\right]$$
$$\mu_1(\cdot) = \mathbb{P}\left[\mathbf{A}\mathbf{x} + \mathbf{B} \in \cdot \mid \mathbf{A}\mathbf{x} + \mathbf{B} \in \mathcal{B}_R\right] \ .$$

Note that μ_1 is supported on $\mathcal{B}_R$. Then, we recursively define, for $n \in \mathbb{Z}^+$,

$$q_{n+1} = \mathbb{P}\left[\mathbf{A}\mathbf{Z}_n + \mathbf{B} \in \mathcal{B}_R\right] \ ,$$
$$\mu_{n+1}(\cdot) = \mathbb{P}\left[\mathbf{A}\mathbf{Z}_n + \mathbf{B} \in \cdot \mid \mathbf{A}\mathbf{Z}_n + \mathbf{B} \in \mathcal{B}_R\right] \ ,$$

where $\mathbf{Z}_n \in \mathbb{R}^d$ is a random vector distributed according to μ_n and independent of $\mathbf{A}$ and $\mathbf{B}$. Then, by the Markov property, we can write

$$\mathbb{P}\left[\mathbf{X}_1(\mathbf{x}),\, \dots,\, \mathbf{X}_n(\mathbf{x}) \in \mathcal{B}_R\right] = q_1 q_2 \cdots q_n \ \text{ and} \tag{5.12}$$
$$\mathbb{P}\left[\mathbf{X}_1(\mu_k),\, \dots,\, \mathbf{X}_n(\mu_k) \in \mathcal{B}_R\right] = q_{k+1} q_{k+2} \cdots q_{k+n} \tag{5.13}$$

for all $n,\, k \in \mathbb{Z}^+$. In particular, by (5.12), we have, for all $n \in \mathbb{Z}^+$,

$$p_0 := \mathbb{P}\left[\tau_R(\mathbf{x}) = \infty\right] \le \mathbb{P}\left[\mathbf{X}_1(\mathbf{x}),\, \dots,\, \mathbf{X}_n(\mathbf{x}) \in \mathcal{B}_R\right] = q_1 q_2 \cdots q_n \ .$$

Since $p_0 > 0$, we must have

$$\lim_{n\to\infty} q_n = 1 \ . \tag{5.14}$$

[Step 2] Compactness argument along with Skorokhod representation theorem
Since each measure μ_n is supported on the compact set $\mathcal{B}_R$, by the Prokhorov theorem we can find a subsequence $\{\mu_{n_k}\}_{k\in\mathbb{Z}^+}$of $\{\mu_n\}_{n\in\mathbb{Z}^+}$ that converges weakly to a probability measure μ_∞ supported on $\mathcal{B}_R$.

Let us fix $n \in \mathbb{Z}^+$. Note from (5.13) that

$$\mathbb{P}\left[\mathbf{X}_n(\mu_{n_k}) \in \mathcal{B}_R\right] \ge q_{n_k+1} q_{n_k+2} \cdots q_{n_k+n}$$

and therefore by (5.14), we have

$$\lim_{k\to\infty} \mathbb{P}\left[\mathbf{X}_n(\mu_{n_k}) \in \mathcal{B}_R\right] = 1 \ . \tag{5.15}$$

We next claim that

$$\mathbb{P}\left[\mathbf{X}_n(\mu_\infty) \in \mathcal{B}_R\right] = 1 \ . \tag{5.16}$$

To that end, it suffices to prove that

$$\mathbb{P}\left[\mathbf{X}_n(\mu_\infty) \in \mathcal{B}_{R+\epsilon}\right] = 1 \text{ for all } \epsilon > 0 \ . \tag{5.17}$$

Fix $\epsilon > 0$. By the Skorokhod representation thoerem, we can find a sequence of random variables $\{\mathbf{y}_k\}_{k\in\mathbb{Z}^+}$ and $\mathbf{y}_\infty$ such that $\mathbf{y}_k \sim \mu_{n_k}$ for $k \in \mathbb{Z}^+$, $\mathbf{y}_\infty \sim \mu_\infty$ and $\mathbf{y}_k \to \mathbf{y}_\infty$ almost surely. Then, by (5.2), we have

$$\begin{aligned}
\mathbb{P}\left[\mathbf{X}_n(\mu_{n_k}) \in \mathcal{B}_R\right] - \mathbb{P}\left[\mathbf{X}_n(\mu_\infty) \in \mathcal{B}_{R+\epsilon}\right] &= \mathbb{P}\left[\mathbf{X}_n(\mathbf{y}_k) \in \mathcal{B}_R\right] - \mathbb{P}\left[\mathbf{X}_n(\mathbf{y}_\infty) \in \mathcal{B}_{R+\epsilon}\right] \\
&\le \mathbb{P}\left[|\mathbf{X}_n(\mathbf{y}_k) - \mathbf{X}_n(\mathbf{y}_\infty)| > \epsilon\right] \\
&= \mathbb{P}\left[|\Pi_n(\mathbf{y}_k - \mathbf{y}_\infty)| > \epsilon\right] \\
&\le \mathbb{P}\left[|\mathbf{y}_k - \mathbf{y}_\infty| \ge \frac{\epsilon}{M}\right] + \mathbb{P}\left[\|\Pi_n\| \ge M\right] \ .
\end{aligned}$$

Since $\mathbf{y}_k$ converges to $\mathbf{y}_\infty$ in probability, by sending $k \to \infty$ and then $M \to \infty$, and recalling (5.15), we get

$$1 - \mathbb{P}_{\mu_\infty}\left[\mathbf{X}_n \in \mathcal{B}_{R+\epsilon}\right] \le 0 \ .$$

This implies (5.17), and hence (5.16).

[Step 3] Contradiction to the non-boundedness of support of ν_∞

Let us take an auxiliary function $\psi : \mathbb{R}^d \to [0,\,1]$ which is a smooth function on $\mathbb{R}^d$ satisfying

$$\psi|_{\mathcal{B}_R} \equiv 0 \ , \quad \psi|_{\mathcal{B}^c_{R+1}} \equiv 1 \ , \ \text{ and } \ |D\psi|_\infty \le 2 \ .$$

Then, by Lemma 5.4, we get

$$\lim_{n\to\infty} \mathbb{E}\psi(\mathbf{X}_n(\mu_\infty)) = \int_{\mathbb{R}^d} \psi \, d\nu_\infty \ .$$

However, by (5.16) and the fact that $\psi|_{\mathcal{B}_R} \equiv 0$, we can notice that the left-hand side of the previous display is 0. This implies that $\int_{\mathbb{R}^d} \psi \, d\nu_\infty = 0$, and in particular $\nu_\infty(\mathcal{B}^c_{R+1}) = 0$. This contradicts to the non-boundedness of the support of ν_∞ from the Assumption 3, and the proof is completed. ☐

5.3. **Scattering property of Π_n.** We denote by

$$\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$$

the d-dimensional unit sphere and denote by σ_d the uniform surface measure on $\mathbb{S}^{d-1}$ normalized in a way that $\sigma_d(\mathbb{S}^{d-1}) = 1$. We denote by

$$\mathcal{D}_r(\mathbf{z}) := \left\{\mathbf{y} \in \mathbb{S}^{d-1} : |\mathbf{y} - \mathbf{z}| < r\right\}$$

the disk of radius $r > 0$ (in fact, the intersection of ball of radius r and unit sphere) on the surface $\mathbb{S}^{d-1}$ centered at $\mathbf{z}$.

The main result of this subsection is the following proposition which is a consequence of Assumption 5 and asserts that, the matrix Π_n is uniformly acts on a fixed vector $\mathbf{x} \in \mathbb{R}^d$ regardless of its direction in terms of β-th moment.

Proposition 5.6. *Suppose that Assumption 5 holds. Then, for all $\beta \in (0, \alpha_\infty)$, there exist $\delta_0 > 0$ and $N_0 > 0$ such that, for all $n \geq N_0$ and $\mathbf{x} \in \mathbb{R}^d$, we have*

$$\mathbb{E}|\Pi_n \mathbf{x}|^\beta \geq \delta_0 \mathbb{E}\|\Pi_n\|^\beta \cdot |\mathbf{x}|^\beta .$$

The proof of this proposition requires several preliminary technical lemmata below. We start with an elemetary observation about finite Borel measures on $\mathbb{S}^{d-1}$.

Lemma 5.7. *Let μ be a finite Borel measure on $\mathbb{S}^{d-1}$ and let $A \subset \mathbb{S}^{d-1}$ be a Borel set. Suppose that*

$$\liminf_{\delta\to 0} \frac{\mu(\mathcal{D}_\delta(\mathbf{z}))}{\sigma_d(\mathcal{D}_\delta(\mathbf{z}))} = 0 \quad \textit{for all } \mathbf{z} \in A .$$

Then, we have $\mu(A) = 0$.

Proof. By the Lebesgue decomposition theorem, we can decompose $\mu = \mu_a + \mu_s$ such that $\mu_a \ll \sigma_d$ and $\mu_s \perp \sigma_d$ where both μ_a and μ_s are positive measures. Then, by [7, Theorem 7.14 at page 143], for almost all (with respect to σ_d or equivalently μ_a) $\mathbf{z} \in A$, we have

$$\frac{d\mu_a}{d\sigma_d}(\mathbf{z}) = \lim_{\delta\to 0} \frac{\mu_a(\mathcal{D}_\delta(\mathbf{z}))}{\sigma_d(\mathcal{D}_\delta(\mathbf{z}))} \leq \liminf_{\delta\to 0} \frac{\mu(\mathcal{D}_\delta(\mathbf{z}))}{\sigma_d(\mathcal{D}_\delta(\mathbf{z}))} = 0$$

and thus we get $\mu_a(A) = 0$.

On the other hand, by [7, Theorem 7.15 at page 143], we know that

$$\lim_{\delta\to 0} \frac{\mu_s(\mathcal{D}_\delta(\mathbf{z}))}{\sigma_d(\mathcal{D}_\delta(\mathbf{z}))} = \infty, \quad \text{a.e. } \mu_s .$$

Since the limit at the left-hand side is 0 for all $\mathbf{z} \in A$, we can conclude that $\mu_s(A) = 0$ as well. This completes the proof. $\square$

The next technical lemma is a key in the proof of Proposition 5.6.

Lemma 5.8. *Suppose that Assumption 5 holds. Let $\phi : \mathbb{R}^d \to [0, \infty)$ be a non-negative measurable function satisfies $\phi > 0$ on $\mathbb{R}^d \setminus \{\mathbf{0}\}$. Fix $n \in \mathbb{Z}^+$, $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ and suppose that $\mathbb{E}[\phi(\Pi_n \mathbf{x})] < \infty$. Define a subset $S = S(\mathbf{x})$ of $\mathbb{S}^{d-1}$ by the collection of all $\mathbf{z} \in \mathbb{S}^{d-1}$ such that*

$$\liminf_{\delta\to 0} \frac{1}{\sigma_d(\mathcal{D}_\delta(\mathbf{z}))} \mathbb{E}\left[\phi(\Pi_n \mathbf{x}) \cdot \mathbf{1}\left\{\Pi_n \mathbf{x} \neq \mathbf{0},\, \frac{\Pi_n \mathbf{x}}{|\Pi_n \mathbf{x}|} \in \mathcal{D}_\delta(\mathbf{z})\right\}\right] > 0 .$$

Then, the set S contains a normal basis of $\mathbb{R}^d$.

Proof. Define a positive Borel measure μ on $\mathbb{S}^{d-1}$ by

$$\mu(D) := \mathbb{E}\left[\phi(\Pi_n \mathbf{x}) \cdot \mathbf{1}\left\{\Pi_n \mathbf{x} \neq \mathbf{0},\, \frac{\Pi_n \mathbf{x}}{|\Pi_n \mathbf{x}|} \in D\right\}\right],$$

for any Borel measurable set D of $\mathbb{S}^{d-1}$. Then, by Lemma 5.7, we have $\mu(\mathbb{S}^{d-1} \setminus S) = 0$ and thus by condition on ϕ, we have

$$\mathbf{1}\left\{\Pi_n \mathbf{x} \neq \mathbf{0},\, \frac{\Pi_n \mathbf{x}}{|\Pi_n \mathbf{x}|} \in \mathbb{S}^{d-1} \setminus S\right\} = 0$$

almost surely with respect to $\mathbb{P}$.

Now suppose on the contrary that S does not contain a basis of $\mathbb{R}^d$ so that there is a non-zero vector $\mathbf{w} \in \mathbb{R}^d$ perpendicular to all the vectors in S. Then, we have

$$\mathbb{P}(\mathbf{w} \cdot \Pi_n \mathbf{x} \neq 0) \leq \mathbb{P}\left(\Pi_n \mathbf{x} \neq \mathbf{0},\, \frac{\Pi_n \mathbf{x}}{|\Pi_n \mathbf{x}|} \in \mathbb{S}^{d-1} \setminus S\right) = 0$$

and thus we get a contradiction to Assumption 5. □

The next lemma asserts that, for each matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ and basis $\{\mathbf{z}_1,\, \dots,\, \mathbf{z}_d\} \in \mathbb{S}^{d-1}$, the norm of vectors $\mathbf{M}\mathbf{z}_i$ can be controlled from below.

Lemma 5.9. *Let $\mathbf{M} \in \mathbb{R}^{d\times d}$, let $\{\mathbf{z}_1,\, \dots,\, \mathbf{z}_d\} \in \mathbb{S}^{d-1}$ be a basis of $\mathbb{R}^d$, and let $\beta > 0$. Denote by $\mathbf{Z} \in \mathbb{R}^{d\times d}$ the matrix whose kth column is $\mathbf{z}_k$ for $k = 1,\, \dots,\, d$. Then, we have*

$$\sum_{i=1}^{d} |\mathbf{M}\mathbf{z}_i|^\beta \geq \frac{\|\mathbf{M}\|^\beta}{d^\beta \|\mathbf{Z}^{-1}\|^\beta}\ . \tag{5.18}$$

Proof. Let $\mathbf{x}_0 \in \mathbb{S}^{d-1}$ be the unit vector satisfy $|\mathbf{M}\mathbf{x}_0| = \|\mathbf{M}\|$. Write $\mathbf{x}_0 = a_1\mathbf{z}_1 + \cdots + a_d\mathbf{z}_d = \mathbf{Z}\mathbf{a}$ where $\mathbf{a} = (a_1,\, \dots,\, a_d)^\dagger$ so that

$$|\mathbf{a}| = |\mathbf{Z}^{-1}\mathbf{x}_0| \leq \|\mathbf{Z}^{-1}\|$$

since $|\mathbf{x}_0| = 1$. Thus, by (5.7), we have

$$\|\mathbf{M}\|^\beta = |\mathbf{M}\mathbf{x}_0|^\beta \leq d^\beta \sum_{i=1}^{d} |a_i|^\beta |\mathbf{M}\mathbf{z}_i|^\beta \leq d^\beta \|\mathbf{Z}^{-1}\|^\beta \sum_{i=1}^{d} |\mathbf{M}\mathbf{z}_i|^\beta\ .$$

□

Now we are ready to prove Proposition 5.6.

Proof of Proposition 5.6. Through scaling, it suffices to show that

$$\mathbb{E}|\Pi_n \mathbf{x}|^\beta \geq \delta_0 \mathbb{E}\|\Pi_n\|^\beta$$

holds for all $\mathbf{x} \in \mathbb{S}^{d-1}$. Suppose that this statement of proposition does not hold so that we can find a sequence $N_1 < N_2 < \cdots$ of positive integers and vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots \in \mathbb{S}^{d-1}$ such that

$$\mathbb{E}|\Pi_{N_k}\mathbf{x}_k|^\beta < \frac{1}{k}\mathbb{E}\|\Pi_{N_k}\|^\beta. \tag{5.19}$$

We can assume here that $N_1 > n_0$. Moreover, since $\mathbb{S}^{d-1}$ is compact, by extracting a suitable subseuqence, we can assume in addition that there exists $\mathbf{x}_\infty \in \mathbb{S}^{d-1}$ such that

$$\lim_{k\to\infty} |\mathbf{x}_\infty - \mathbf{x}_k| = 0\ . \tag{5.20}$$

Then, by (5.7) and (5.19),

$$\begin{aligned}\mathbb{E}|\Pi_{N_k}\mathbf{x}_\infty|^\beta &\leq 2^\beta(\mathbb{E}|\Pi_{N_k}\mathbf{x}_k|^\beta + \mathbb{E}|\Pi_{N_k}(\mathbf{x}_\infty - \mathbf{x}_k)|^\beta) \\ &\leq 2^\beta\, \mathbb{E}\|\Pi_{N_k}\|^\beta \left(\frac{1}{k} + |\mathbf{x}_\infty - \mathbf{x}_k|^\beta\right)\ .\end{aligned}$$

Thus, write $\rho_k = 2^\beta(\frac{1}{k} + |\mathbf{x}_\infty - \mathbf{x}_k|^\beta)$ so that by (5.20), we have

$$\lim_{k\to\infty} \rho_k = 0 \quad \text{and} \quad \mathbb{E}|\Pi_{N_k}\mathbf{x}_\infty|^\beta \le \rho_k \mathbb{E}\|\Pi_{N_k}\|^\beta \ . \tag{5.21}$$

Now applying Lemma 5.8 with $\phi(\mathbf{x}) = |\mathbf{x}|^\beta$ and $\mathbf{x} = \mathbf{x}_\infty$, which is possible since $\beta \in (0,\, \alpha_\infty)$, (cf. Assumption 2)

$$\mathbb{E}[\phi(\Pi_{n_0}\mathbf{x}_\infty)] = \mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta\right] \le \mathbb{E}\|\Pi_{n_0}\|^\beta < \infty \ ,$$

we can find a normal basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_d\} \in S(\mathbf{x}_\infty)$ [see the definition in Lemma 5.8 for $S(\mathbf{x}_\infty)$] of $\mathbb{R}^d$. We next claim that, for all $\ell \in \{1,\, \ldots,\, d\}$,

$$\lim_{k\to\infty} \max \frac{\mathbb{E}|\Pi_{N_k - n_0}\mathbf{z}_\ell|^\beta}{\mathbb{E}\|\Pi_{N_k-n_0}\|^\beta} = 0 \ . \tag{5.22}$$

To that end, for $\ell \in \{1,\, \ldots,\, d\}$ and $\delta > 0$, define

$$\mathcal{T}_{\ell,\,\delta} := \left\{ \Pi_{n_0}\mathbf{x}_\infty \ne \mathbf{0},\ \frac{\Pi_{n_0}\mathbf{x}_\infty}{|\Pi_{n_0}\mathbf{x}_\infty|} \in \mathcal{D}_\delta(\mathbf{z}_\ell) \right\} \ .$$

Since $\mathbf{z}_\ell \in S(\mathbf{x}_\infty)$, we can find $\epsilon_0,\, \delta_0 > 0$ such that, for all $\delta \in (0,\, \delta_0)$,

$$\mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right] \ge \epsilon_0 \cdot \sigma_d(\mathcal{D}_\delta(\mathbf{z}_\ell)) \ . \tag{5.23}$$

We note from (5.7) that, for all $\mathbf{y} \in \mathcal{D}_\delta(\mathbf{z}_\ell)$ and $n \in \mathbb{Z}^+$,

$$|\Pi_n \mathbf{z}_\ell|^\beta \le 2^\beta(|\Pi_n\mathbf{y}|^\beta + |\Pi_n(\mathbf{z}_\ell - \mathbf{y})|^\beta) \le 2^\beta(|\Pi_n\mathbf{y}|^\beta + \delta^\beta\|\Pi_n\|^\beta) \ .$$

Since $\frac{\Pi_{n_0}\mathbf{x}_\infty}{|\Pi_{n_0}\mathbf{x}_\infty|} \in \mathcal{D}_\delta(\mathbf{z}_\ell)$ under $\mathcal{T}_{\ell,\,\delta}$, applying the previous bound with $\mathbf{y} = \frac{\Pi_{n_0}\mathbf{x}_\infty}{|\Pi_{n_0}\mathbf{x}_\infty|}$, we get, for any $k \in \mathbb{Z}^+$, (recall that $N_k \ge N_1 > n_0$)

$$\begin{aligned}
&\mathbb{E}\left[|\Pi_{N_k-n_0}\mathbf{z}_\ell|^\beta\right] \cdot \mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right] \\
&= \mathbb{E}\left[|\mathbf{A}_{N_k}\cdots\mathbf{A}_{n_0+1}\mathbf{z}_\ell|^\beta \cdot |\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right] \\
&\le 2^\beta\mathbb{E}\left[\left(\left|\mathbf{A}_{N_k}\cdots\mathbf{A}_{n_0+1}\frac{\Pi_{n_0}\mathbf{x}_\infty}{|\Pi_{n_0}\mathbf{x}_\infty|}\right|^\beta + \delta^\beta\|\mathbf{A}_{N_k}\cdots\mathbf{A}_{n_0+1}\|^\beta\right) \cdot |\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right] \\
&\le 2^\beta\, \mathbb{E}|\Pi_{N_k}\mathbf{x}_\infty|^\beta + 2^\beta\delta^\beta\, \mathbb{E}\|\Pi_{N_k-n_0}\|^\beta \cdot \mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right] \\
&\le 2^\beta\mathbb{E}\|\Pi_{N_k-n_0}\|^\beta \cdot \left(\rho_k\mathbb{E}\|\Pi_{n_0}\|^\beta + \delta^\beta\mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right]\right)
\end{aligned}$$

where the last line follows from (5.21) and submultiplicativity of matrix operator norm. Dividing both sides by $\mathbb{E}\|\Pi_{N_k-n_0}\|^\beta\mathbb{E}\left[|\Pi_{n_0}\mathbf{x}_\infty|^\beta \cdot \mathbf{1}_{\mathcal{T}_{\ell,\,\delta}}\right]$ and applying (5.23), we get, for all $\delta \in (0,\, \delta_0)$,

$$\frac{\mathbb{E}|\Pi_{N_k-n_0}\mathbf{z}_\ell|^\beta}{\mathbb{E}|\Pi_{N_k-n_0}\mathbf{z}_\ell|^\beta} \le 2^\beta\left(\frac{\rho_k}{\epsilon_0 \cdot \sigma_d(\mathcal{D}_\delta(\mathbf{z}_\ell))}\mathbb{E}\|\Pi_{n_0}\|^\beta + \delta^\beta\right) \le C_\beta\left(\frac{\rho_k}{\delta^{d-1}} + \delta^\beta\right) \tag{5.24}$$

for some constant $C_\beta > 0$ since $\sigma_d(\mathcal{D}_\delta(\mathbf{z}_\ell)) = O(\delta^{d-1})$. Hence, by (5.21), we get

$$\limsup_{k\to\infty} \frac{\mathbb{E}|\Pi_{N_k-n_0}\mathbf{z}_\ell|^\beta}{\mathbb{E}\|\Pi_{N_k-n_0}\|^\beta} \le C\delta^\beta$$

and hency by letting $\delta \to 0$ we get (5.22).

On the other hand, by Lemma 5.9, we get

$$\sum_{\ell=1}^{d} \frac{\mathbb{E}|\Pi_{N_k-n_0}\mathbf{z}_\ell|^\beta}{\mathbb{E}\|\Pi_{N_k-n_0}\|^\beta} \ge \frac{1}{d^\beta \|\mathbf{Z}^{-1}\|^\beta}$$

where $\mathbf{Z} \in \mathbb{R}^{d\times d}$ is the matrix whose kth column is $\mathbf{z}_k$ for $k = 1, \ldots, d$. Letting $k \to \infty$ and recalling (5.22) yield a contradiction. □

6. Escape Time Analysis for contractive Regime

The purpose of the current section is to prove Theorem 3.4. Hence, we assume that the Lyapunov exponent γ_L is negative, and moreover assume Assumptions 1, 2, 3, 4 and 5 throughout section.

The proof of Theorem 3.4 is based on the following propositions on the estimate of the mean exit time in the super- and sub-critical regime. We recall the constnat $\alpha_+ \in (\alpha, \infty)$ from Assumption 4. We note that, we shall use $C_\gamma, D_\gamma, \ldots$ to denote constants depending only on γ (except for the distribution of $(\mathbf{A}, \mathbf{B})$), and different appearances of C_γ, for instance, may denote different constants.

Proposition 6.1. *For each $\gamma \in (\alpha, \alpha_+)$, there exists a constant $C_\gamma > 0$ such that, for all $\mathbf{x}_0 \in \mathbb{R}^d$ and $R > 0$,*

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \le C_\gamma \left(R^\gamma + 1\right) . \tag{6.1}$$

Proposition 6.2. *For each $\gamma \in (0, \alpha)$, there exist constants $C_\gamma, D_\gamma > 0$ such that, for all $\mathbf{x}_0 \in \mathbb{R}^d$ and $R > 0$,*

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \ge C_\gamma R^\gamma - D_\gamma \left(|\mathbf{x}_0|^\gamma + 1\right) . \tag{6.2}$$

The proof of these propositions are given in Sections 6.3 and 6.4, respectively. We assume these propositions now and then complete the proof of Theorem 3.4.

Proof of Theorem 3.4. Fix $\mathbf{x}_0 \in \mathbb{R}^d$. Let $\gamma \in (\alpha, \alpha_+)$. Then, by Proposition 6.1, we have

$$\frac{\log \mathbb{E}\left[\tau_R(\mathbf{x}_0)\right]}{\log R} \le \frac{\log C_\gamma}{\log R} + \frac{\log(R^\gamma + 1)}{\log R}$$

and therefore by letting $R \to \infty$ and then $\gamma \searrow \alpha$, we get

$$\limsup_{R\to\infty} \frac{\log \mathbb{E}\left[\tau_R(\mathbf{x}_0)\right]}{\log R} \le \alpha . \tag{6.3}$$

On the other hand, for $\gamma \in (0, \alpha)$, we take R sufficiently large so that, by Proposition 6.1, $\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \ge C_\gamma R^\gamma$. Then, taking logarithm both sides, dividing both sides by $\log R$, and sending $R \to \infty$ and then $\gamma \nearrow \alpha$, we get

$$\liminf_{R\to\infty} \frac{\log \mathbb{E}\left[\tau_R(\mathbf{x}_0)\right]}{\log R} \geq \alpha \ . \tag{6.4}$$

Combining (6.3) and (6.4) completes the proof. □

6.1. **Control of escaping location.** We control, based on Assumption 4, that we can not exit the ball $\mathcal{B}_R$ far away from the ball. This is consequence of the assumption that the tail of $\mathbf{Ax} + \mathbf{B}$ is not that heavy.

Notation. From now on, we simply write $\mathbf{X}_{\tau_R}(\mathbf{x})$ instead of $\mathbf{X}_{\tau_{R(\mathbf{x})}}(\mathbf{x})$ as there is no risk of confusion.

Lemma 6.3. *For all $\gamma \in [0, \alpha_+)$, there exists a constant $C_\gamma > 0$ such that*

$$\mathbb{E}|\mathbf{X}_{\tau_R}(\mathbf{x})|^\gamma \leq C_\gamma(R^\gamma + 1) \ .$$

Proof. Let $\gamma \in [0, \alpha_+)$, $R > R_0$ and $\mathbf{x} \in \mathcal{B}_R$. By the layer-cake formula and the fact that

$$\lim_{z\to\infty} z^\gamma \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > z) = 0$$

which follows from (3.1), we can write

$$\begin{aligned}
&\mathbb{E}\left[|\mathbf{Ax} + \mathbf{B}|^\gamma \cdot \mathbf{1}\left\{|\mathbf{Ax} + \mathbf{B}| > R\right\}\right] \\
&\quad = R^\gamma \left[\mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > R) + \int_1^\infty \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > Rz) \cdot \gamma z^{\gamma-1} dz\right] \ .
\end{aligned}$$

Since $\gamma \in [0, \alpha_+)$, again by (3.1), we have

$$\int_{z_0}^\infty \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > Rz) \cdot \gamma z^{\gamma-1} dz \leq \frac{\gamma C_0}{\alpha_+ - \gamma} \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > R)$$

plus

$$\int_1^{z_0} \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > Rz) \cdot \gamma z^{\gamma-1} dz \leq \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > R) \cdot \int_1^{z_0} \gamma z^{\gamma-1} dz = z_0^\gamma \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > R) \ ,$$

therefore we can conclude that

$$\mathbb{E}\left[|\mathbf{Ax} + \mathbf{B}|^\gamma \cdot \mathbf{1}\left\{|\mathbf{Ax} + \mathbf{B}| > R\right\}\right] \leq \left(\frac{\gamma C_0}{\alpha_+ - \gamma} + z_0^\gamma\right) R^\gamma \cdot \mathbb{P}(|\mathbf{Ax} + \mathbf{B}| > R) \ . \tag{6.5}$$

Consequently, by the strong Markov property,

$$\mathbb{E}|\mathbf{X}_{\tau_R}(\mathbf{x})|^\gamma \leq \sup_{\mathbf{y}\in\mathcal{B}_R} \mathbb{E}\left[|\mathbf{Ay} + \mathbf{B}|^\gamma \,\middle|\, |\mathbf{Ay} + \mathbf{B}| > R\right] \leq \left(\frac{\gamma C_0}{\alpha_+ - \gamma} + z_0^\gamma\right) R^\gamma \ .$$

□

6.2. **Moment estimates.**

Lemma 6.4. *Let $\gamma \in [0, \alpha_\infty)$ and $n \in \mathbb{Z}^+$. Then, there exists a constant $C_{n,\gamma} > 0$ such that*

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma < C_{n,\gamma}(|\mathbf{x}|^\gamma + 1) \quad \textit{for all } \mathbf{x} \in \mathbb{R}^d.$$

In particular, $\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma < \infty$*.*

Proof. Since $\gamma \in [0,\, \alpha_\infty)$, we have that $\mathbb{E}\|\mathbf{A}\|^\gamma < \infty$ and $\mathbb{E}|\mathbf{B}|^\gamma < \infty$ by Assumption 2 and (3.3), respectively. Thus, by (5.1) and (5.7), we have

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma \le (n+1)^\gamma \left((\mathbb{E}\|\mathbf{A}\|^\gamma)^n |\mathbf{x}|^\gamma + \sum_{i=0}^{n-1} (\mathbb{E}\|\mathbf{A}\|^\gamma)^{n-1} \mathbb{E}|\mathbf{B}|^\gamma \right)$$

and hence we get the desired result. □

Lemma 6.5. *Let* $\gamma \in [0,\, \alpha_\infty)$ *and* $n \in \mathbb{Z}^+$*. It holds that*

$$\inf_{\mathbf{x}:|\mathbf{x}|=1} \mathbb{E}|\Pi_n \mathbf{x} - \mathbf{x}|^\gamma > 0 \,. \tag{6.6}$$

Proof. We define $F_{n,\gamma} : \mathbb{R}^d \to \mathbb{R}$ as

$$F_{n,\gamma}(\mathbf{x}) = \mathbb{E}|\Pi_n \mathbf{x} - \mathbf{x}|^\gamma \,.$$

We will prove that $F_{n,\gamma} > 0$ on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ and that $F_{n,\gamma}$ is continuous on $\mathbb{R}^d$. Then, (6.6) follows immediately.

Let us first show the continuity of $F_{n,\gamma}$. If $\gamma < 1$, by the elementary inequality

$$|a^\gamma - b^\gamma| \le |a-b|^\gamma \quad \text{and} \quad (a+b)^\gamma \le a^\gamma + b^\gamma \quad \text{for } a,\, b \ge 0 \,,$$

we have

$$|F_{n,\gamma}(\mathbf{x}) - F_{n,\gamma}(\mathbf{y})| \le \mathbb{E}|\Pi_n(\mathbf{x}-\mathbf{y}) - (\mathbf{x}-\mathbf{y})|^\gamma \le (\mathbb{E}|\Pi_n|^\gamma + 1)|\mathbf{x}-\mathbf{y}|^\gamma \tag{6.7}$$

and the continuity follows immediately. On the other hand, if $\gamma \ge 1$, by the elementary inequality

$$|a^\gamma - b^\gamma| \le \gamma |a-b|(a+b)^{\gamma-1} \quad ;\ a,\, b \ge 0 \,,$$

which holds because of the mean-value theorem, then by Hölder inequality we have

$$\begin{aligned}
&|F_{n,\gamma}(\mathbf{x}) - F_{n,\gamma}(\mathbf{y})| \\
&\le \gamma \mathbb{E}\left[|\Pi_n(\mathbf{x}-\mathbf{y}) - (\mathbf{x}-\mathbf{y})| \cdot (||\Pi_n \mathbf{x} - \mathbf{x}| + |\Pi_n \mathbf{y} - \mathbf{y}|)^{\gamma-1} \right] \\
&\le \gamma \mathbb{E}\left[\Pi_n(\mathbf{x}-\mathbf{y}) - (\mathbf{x}-\mathbf{y})|^\gamma \right]^{\frac{1}{\gamma}} \mathbb{E}\left[(|\Pi_n \mathbf{x} - \mathbf{x}| + |\Pi_n \mathbf{y} - \mathbf{y}|)^\gamma \right]^{\frac{\gamma-1}{\gamma}}
\end{aligned}$$

Hence, by the bound (cf. (5.7)),

$$\mathbb{E}|\Pi_n \mathbf{x} - \mathbf{x}|^\gamma \le 2^\gamma \mathbb{E}(\|\Pi_n\|^\gamma + 1)|\mathbf{x}|^\gamma \tag{6.8}$$

we can conclude that

$$|F_{n,\gamma}(\mathbf{x}) - F_{n,\gamma}(\mathbf{y})| \le C_{n,\gamma} |\mathbf{x}-\mathbf{y}| \left(|\mathbf{x}|^{\gamma-1} + |\mathbf{y}|^{\gamma-1} + 1 \right)$$

and hence the continuity follows.

Next we shall prove that $F_{n,\gamma} > 0$. Suppose that $F_{n,\gamma}(\mathbf{x}_0) = 0$, i.e., $\Pi_n \mathbf{x}_0 = \mathbf{x}_0$ almost surely, for some $\mathbf{x}_0 \in \mathbb{R}^d \setminus \{\mathbf{0}\}$. Then, inductively, we have that $\Pi_{kn} \mathbf{x}_0 = \mathbf{x}_0$ almost surely for all $k \ge 1$. Let us take $\beta \in (\alpha,\, \alpha_+)$ so that $h_{\mathbf{A}}(\beta) > 1$ by Theorem 2.3-(4). Then, by

Proposition 5.6, we have

$$|\mathbf{x}_0|^\beta = \mathbb{E}|\Pi_{kn}\mathbf{x}_0|^\beta \ge \delta_0 \mathbb{E}\|\Pi_{kn}\|^\beta |\mathbf{x}_0| \ ,$$

and therefore, as $\mathbf{x}_0 \neq 0$, we get $\delta_0 \mathbb{E}\|\Pi_{kn}\|^\beta \le 1$ for all $k \in \mathbb{Z}^+$. This implies that

$$h_{\mathbf{A}}(\beta) = \lim_{k\to\infty} \left[\mathbb{E}\|\Pi_{kn}\|^\beta\right]^{\frac{1}{kn}} \le \lim_{k\to\infty} \delta_0^{-\frac{1}{kn}} = 1 \ .$$

This contradicts to $h_{\mathbf{A}}(\beta) > 1$ (as $\beta \in (\alpha,\, \alpha_\infty)$) and we are done. $\square$

Lemma 6.6. *Let $\gamma \in [0,\, \alpha_\infty)$ and $n \in \mathbb{Z}^+$. It holds that*

$$\inf_{\mathbf{x}\in\mathbb{R}^d} \mathbb{E}|\mathbf{X}_n(\mathbf{x}) - \mathbf{x}|^\gamma > 0 \ .$$

Proof. Define $F = F_{n,\gamma} : \mathbb{R}^d \to \mathbb{R}$ as

$$F_{n,\gamma}(\mathbf{x}) = \mathbb{E}|\mathbf{X}_n(\mathbf{x}) - \mathbf{x}|^\gamma \ .$$

Then, as in the proof of Lemma 6.5, we can verify that $F_{n,\gamma}$ is continuous on $\mathbb{R}^d$. The only difference in the proof is that, in the case $\gamma > 1$, we need to substitute the bound of $\mathbb{E}|\Pi_n\mathbf{x} - \mathbf{x}|^\gamma$ explained in (6.8) with a bound of $\mathbb{E}|\mathbf{X}_n(\mathbf{x}) - \mathbf{x}|^\gamma$. This can be done with Lemma 6.4 which provides

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x}) - \mathbf{x}|^\gamma < C_{n,\gamma}(|\mathbf{x}|^\gamma + 1) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

Moreover, by Lemma 5.4-(2), we have $F_{n,\gamma} > 0$. Finally, by (5.2) and (5.7), we have

$$F_{n,\gamma}(\mathbf{x}) \ge \frac{1}{2^\gamma}\mathbb{E}|\Pi_n\mathbf{x} - \mathbf{x}|^\gamma - \mathbb{E}|\mathbf{X}_n(\mathbf{0})|^\gamma \ge c_{n,\gamma}|\mathbf{x}|^\gamma - c'_{n,\gamma} \ , \tag{6.9}$$

where $c'_{n,\gamma} = \mathbb{E}|\mathbf{X}_n(\mathbf{0})|^\gamma$ and

$$c_{n,\gamma} = \frac{1}{2^\gamma} \inf_{\mathbf{y}:|\mathbf{y}|=1} \mathbb{E}|\Pi_n\mathbf{y} - \mathbf{y}|^\gamma > 0$$

where the last strict inequality is the content of Lemma 6.5. Summing up, $F_{n,\gamma}$ is a positive continuous function, which diverges to $+\infty$ as $|\mathbf{x}| \to \infty$ thanks to (6.9), and hence the assertion of the lemma follows. $\square$

Lemma 6.7. *Let $\gamma \in (\alpha,\, \alpha_\infty)$. Then, there exists $N_\gamma > 0$ such that, for all $n \ge N_\gamma$ and $\mathbf{x} \in \mathbb{R}^d$,*

$$\mathbb{E}\left[|\mathbf{X}_n(\mathbf{x})|^\gamma \,\middle|\, \mathbf{X}_{n-N_\gamma}(\mathbf{x})\right] \ge 2|\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma - \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma \ ,$$

and in particular, we have

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma \ge 2\mathbb{E}|\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma - \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma \ .$$

Proof. Fix $\gamma \in (\alpha,\, \alpha_\infty)$. Since $h_{\mathbf{A}}(\gamma) > 1$ by Theorem 2.3-(1), there exist constants $N_1 = N_1(\gamma) > 0$ and $\rho_0 = \rho_0(\gamma) > 0$ such that

$$\mathbb{E}\|\Pi_n\|^\gamma \ge (1+\rho_0)^n \quad \text{for all } n > N_1 \ . \tag{6.10}$$

Recall the constants N_0 and δ_0 obtained in Proposition 5.6. Then, by Proposition 5.6, for all $\mathbf{x} \in \mathbb{R}^d$ and $n > \max(N_0, N_1)$, we have

$$\mathbb{E}|\Pi_n \mathbf{x}|^\gamma \geq \delta_0 \cdot (1+\rho_0)^n |\mathbf{x}|^\gamma . \tag{6.11}$$

Let us take $N_\gamma > \max(N_0, N_1)$ large enough so that

$$\frac{\delta_0}{2^\gamma} \cdot (1+\rho_0)^{N_\gamma} \geq 2 . \tag{6.12}$$

By the same computation with (5.1), we can write

$$\mathbf{X}_n(\mathbf{x}) = \mathbf{A}_n \mathbf{A}_{n-1} \cdots \mathbf{A}_{n-N_\gamma+1} \mathbf{X}_{n-N_\gamma}(\mathbf{x}) + \mathbf{W} \tag{6.13}$$

where

$$\mathbf{W} = \mathbf{A}_n \cdots \mathbf{A}_{n-N_\gamma+2} \mathbf{B}_{n-N_\gamma+1} + \cdots + \mathbf{A}_n \mathbf{B}_{n-1} + \mathbf{B}_n . \tag{6.14}$$

Note that $\mathbf{W}$ has the same distribution with $\mathbf{X}_{N_\gamma}(\mathbf{0})$ by (5.1). By (5.7), Proposition 5.6, (6.11), and (6.12), for all $n \geq N_\gamma$,

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{X}_n(\mathbf{x})|^\gamma \,\middle|\, \mathbf{X}_{n-N_\gamma}(\mathbf{x})\right] &\geq \frac{1}{2^\gamma} \mathbb{E}\left[|\mathbf{A}_n \mathbf{A}_{n-1} \cdots \mathbf{A}_{n-N_\gamma+1} \mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma \,\middle|\, \mathbf{X}_{n-N_\gamma}(\mathbf{x})\right] - \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma \\
&\geq \frac{\delta_0}{2^\gamma} \cdot (1+\rho_0)^{N_\gamma} \mathbb{E}\|\mathbf{A}_n \mathbf{A}_{n-1} \cdots \mathbf{A}_{n-N_\gamma+1}\|^\gamma \cdot |\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma - \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma \\
&\geq 2|\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma - \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma .
\end{aligned}$$

This completes the proof of the first assertion. The second assertion follows directly from the first one by the tower property. □

From this moment on, for $\gamma \in (\alpha, \alpha_\infty)$, the constant N_γ always refers to the one obtained in the previous lemma.

Lemma 6.8. *Let $\gamma \in (\alpha, \alpha_\infty)$. There exists $M_\gamma > 0$ such that, for all $n \geq M_\gamma$ and $\mathbf{x} \in \mathbb{R}^d$,*

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma > \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma + 1 .$$

Proof. Denote by $\mathbf{X}'_{N_\gamma}(\mathbf{x})$ an independent copy of $\mathbf{X}_{N_\gamma}(\mathbf{x})$ so that $\mathbf{X}_{n+N_\gamma}(\mathbf{x})$ has the same distribution with $\mathbf{X}_n(\mathbf{X}'_{N_\gamma}(\mathbf{x}))$. Then, by (5.7), (5.2),and Proposition 5.6, there exists $\delta_0(\gamma) > 0$ and $N_0(\gamma) \in \mathbb{Z}^+$ such that, for all $n \geq N_0(\gamma)$ and $\mathbf{x} \in \mathbb{R}^d$,

$$\mathbb{E}|\Pi_n \mathbf{x}|^\gamma \geq \delta_0 \mathbb{E}\|\Pi_n\|^\gamma \cdot |\mathbf{x}|^\gamma$$

holds. For all $n \in \mathbb{Z}^+$, by (5.7), (5.2) and Proposition 5.6, we have

$$\begin{aligned}
\mathbb{E}|\mathbf{X}_{n+N_\gamma}(\mathbf{x})|^\gamma + \mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma &= \mathbb{E}\left[|\mathbf{X}_n(\mathbf{X}'_{N_\gamma}(\mathbf{x}))|^\gamma + |\mathbf{X}_n(\mathbf{x})|^\gamma\right] \\
&\geq \frac{1}{2^\gamma} \mathbb{E}|\mathbf{X}_n(\mathbf{X}'_{N_\gamma}(\mathbf{x})) - \mathbf{X}_n(\mathbf{x})|^\gamma \\
&\geq \frac{\delta_0}{2^\gamma} \mathbb{E}\|\Pi_n\|^\gamma \cdot \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{x}) - \mathbf{x}|^\gamma \\
&\geq \frac{\delta_0}{2^\gamma} \mathbb{E}\|\Pi_n\|^\gamma \cdot \left(\inf_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{x}) - \mathbf{x}|^\gamma\right) .
\end{aligned}$$

By Lemma 6.4-(2) and (6.10), we can find $N_1(\gamma) > 0$ so that the right-hand side of the previous bound is greater than $2\mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma + 2$ for all $n \geq N_1(\gamma)$. Let us take

$$M_\gamma = \max\{N_0(\gamma),\, N_1(\gamma)\} + N_\gamma$$

so that for $n \geq M_\gamma$, the previous bound implies that, for all $\mathbf{x} \in \mathbb{R}^d$,

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma + \mathbb{E}|\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma > 2\mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma + 2\ .$$

This implies $\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma > \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma + 1$ since we have

$$\mathbb{E}|\mathbf{X}_{n-N_\gamma}(\mathbf{x})|^\gamma \leq \frac{1}{2}\left[\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma + \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma\right]$$

by Lemma 6.7. □

6.3. **Upper bound of mean exit time.** We start from a lemma which is useful in the proof of both upper and lower bounds. We denote by $\mathcal{F}_n$ the σ-algebra generated by $(\mathbf{A}_k,\, \mathbf{B}_k)_{k\in[\![0,\, n]\!]}$ so that the process $(\mathbf{X}_n)_{n\in\mathbb{Z}^+}$ is adapted to the filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$.

Lemma 6.9. *Let $\gamma \in (0,\, \alpha_\infty)$ and $N \in \mathbb{Z}^+$. Then, there exist a constant $C_{\gamma,\, N} > 0$ such that, for all $\mathbf{x} \in \mathbb{R}^d$ and $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$-stopping time τ,*

$$\max\{\mathbb{E}|\mathbf{X}_\tau(\mathbf{x})|^\gamma,\, \mathbb{E}|\mathbf{X}_{\tau+1}(\mathbf{x})|^\gamma,\, \dots,\, \mathbb{E}|\mathbf{X}_{\tau+N-1}(\mathbf{x})|^\gamma\} \leq C_{\gamma,\, N}\left(\mathbb{E}|\mathbf{X}_\tau(\mathbf{x})|^\gamma + 1\right)\ .$$

Proof. Fix $\gamma \in (0,\, \alpha_\infty)$. In view of (5.1) and (5.7), for all $n \in \mathbb{Z}^+$ and $\mathbf{x} \in \mathbb{R}^d$,

$$\mathbb{E}|\mathbf{X}_n(\mathbf{x})|^\gamma \leq 2^\gamma \mathbb{E}\|\Pi_n\|^\gamma \cdot |\mathbf{x}|^\gamma + 2^\gamma \mathbb{E}\left[|\mathbf{A}_2\cdots\mathbf{A}_n\mathbf{B}_1 + \cdots + \mathbf{A}_n\mathbf{B}_{n-1} + \mathbf{B}_n|^\gamma\right]\ .$$

Hence, for all $\mathbf{x} \in \mathbb{R}^d$,we have

$$\max\{\mathbb{E}|\mathbf{X}_0(\mathbf{x})|^\gamma,\, \mathbb{E}|\mathbf{X}_1(\mathbf{x})|^\gamma,\, \dots,\, \mathbb{E}|\mathbf{X}_{N-1}(\mathbf{x})|^\gamma\} \leq C_{\gamma,\, N}\left(|\mathbf{x}|^\gamma + 1\right)\ , \tag{6.15}$$

where

$$C_{\gamma,\, N} = \max_{n\in[\![0,\, N-1]\!]}\{2^\gamma\mathbb{E}\|\Pi_n\|^\gamma,\, 2^\gamma\mathbb{E}\left[|\mathbf{A}_2\cdots\mathbf{A}_n\mathbf{B}_1 + \cdots + \mathbf{A}_n\mathbf{B}_{n-1} + \mathbf{B}_n|^\gamma\right]\}\ .$$

We note that $C_{\gamma,\, N} < \infty$ because of $\mathbb{E}\|\mathbf{A}\|^\gamma < \infty$ and $\mathbb{E}|\mathbf{B}|^\gamma < \infty$ from Theorem 2.3-(2) and (3.3). Now the conclusion of the lemma follows from (6.15) and the strong Markov property since $\mathbf{X}_{\tau+n}(\mathbf{x}) = \mathbf{X}_n(\mathbf{X}_\tau(\mathbf{x}))$ for $n \in \mathbb{Z}^+$ and stopping time τ. □

For $\gamma \in (\alpha,\, \alpha_+)$, recall the constant $N_\gamma > 0$ from Lemma 6.7. We can infer from Lemma 6.7 that the norm $|\mathbf{X}_n(\mathbf{x})|^\gamma$ is diverging when we count the process by N_γ skipping. Hence, we write in this subsection

$$\widehat{\mathbf{X}}_n(\mathbf{x}) = \mathbf{X}_{nN_\gamma}(\mathbf{x}) \quad ;\ n \in \mathbb{Z}^+\ . \tag{6.16}$$

To analyze the behavior of this accelerated process, we define

$$\widehat{\tau}_R(\mathbf{x}) = \inf\left\{n \in \mathbb{Z}^+ : \tau_R(\mathbf{x}) \leq nN_\gamma\right\}\ . \tag{6.17}$$

We note that $\widehat{\mathbf{X}}_n(\mathbf{x})$ and $\widehat{\tau}_R(\mathbf{x})$ depend on γ, although we did not emphasize this dependency.

Lemma 6.10. *For each $\gamma \in (\alpha,\, \alpha_+)$, there exists a constant $C_\gamma > 0$ such that, for all $R > 0$ and $\mathbf{x} \in \mathbb{R}^d$,*

$$\mathbb{E}|\widehat{\mathbf{X}}_{\widehat{\tau}_R(\mathbf{x})}(\mathbf{x})|^\gamma < C_\gamma\,(R^\gamma + 1)\ .$$

Proof. Fix $\gamma \in (\alpha,\, \alpha_+)$, $R > 0$ and $\mathbf{x} \in \mathbb{R}^d$. By definition (6.17) of $\widehat{\tau}_R(\mathbf{x})$, we have

$$\widehat{\tau}_R(\mathbf{x})N_\gamma \in \{\tau_R(\mathbf{x}),\, \tau_R(\mathbf{x})+1,\, \dots,\, \tau_R(\mathbf{x})+N_\gamma-1\}\ .$$

Since $\tau_R(\mathbf{x})$ is $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$-stopping time, by Lemma 6.9, we have

$$\mathbb{E}|\widehat{\mathbf{X}}_{\widehat{\tau}_R(\mathbf{x})}(\mathbf{x})|^\gamma = \mathbb{E}|\mathbf{X}_{\widehat{\tau}_R(\mathbf{x})N_\gamma}(\mathbf{x})|^\gamma \le C_\gamma\left(\mathbb{E}|\mathbf{X}_{\tau_R(\mathbf{x})}(\mathbf{x})|^\gamma + 1\right)\ .$$

This completes the proof since $\mathbb{E}|\mathbf{X}_{\tau_R(\mathbf{x})}(\mathbf{x})|^\gamma \le C_\gamma R^\gamma$ by Lemma 6.3. □

Now we are ready to prove Proposition 6.1.

Proof of Proposition 6.1. Fix $\gamma \in (\alpha,\, \alpha_+)$ and $R > 0$. We also fix $\mathbf{x} \in \mathbb{R}^d$ and simply write $\widehat{\mathbf{X}}_n = \widehat{\mathbf{X}}_n(\mathbf{x})$.

Write

$$G_\gamma := \mathbb{E}|\mathbf{X}_{N_\gamma}(\mathbf{0})|^\gamma > 0$$

where the strict inequality follows from Lemma (5.4)-(2) and let us define $F = F_\gamma : \mathbb{R}^d \to \mathbb{R}$ by

$$F(\mathbf{x}) := |\mathbf{x}|^\gamma - G_\gamma \quad ;\ \mathbf{x} \in \mathbb{R}^d\ .$$

For $n \in \mathbb{Z}^+$, denote by $\widehat{\mathcal{F}}_n$ the σ-algebra generated by $(\mathbf{A}_k,\, \mathbf{B}_k)_{k\in[\![0,\, nN_\gamma]\!]}$, and define a $\widehat{\mathcal{F}}_n$-measurable random variable M_n by

$$M_n := |\widehat{\mathbf{X}}_n|^\gamma - \sum_{k=0}^{n-1} F(\widehat{\mathbf{X}}_k)\ .$$

Then, by Lemma 6.7, we have

$$\mathbb{E}\left[M_{n+1}\,|\,\widehat{\mathcal{F}}_n\right] \ge M_n$$

and therefore $(M_n)_{n\in\mathbb{Z}^+}$ is indeed an $(\widehat{\mathcal{F}}_n)_{n\in\mathbb{Z}^+}$-submartingale. By Lemma 6.8, we can find $K_\gamma > 0$ independent of $\mathbf{x}$ such that

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}F(\widehat{\mathbf{X}}_k) \ge \frac{1}{2} \quad \text{for all } n \ge K_\gamma\ . \tag{6.18}$$

Since $\widehat{\tau}_R(\mathbf{x})$ defined in (6.17) is an $(\widehat{\mathcal{F}}_n)_{n\in\mathbb{Z}^+}$-stopping time, by the optional stopping theorem, for all $n \in \mathbb{Z}^+$,

$$\mathbb{E}|\widehat{\mathbf{X}}_{n\wedge\widehat{\tau}_R(\mathbf{x})}|^\gamma - \mathbb{E}\sum_{k=0}^{n\wedge\widehat{\tau}_R(\mathbf{x})-1} F(\widehat{\mathbf{X}}_k) = \mathbb{E}\left[M_{n\wedge\widehat{\tau}_R(\mathbf{x})}\right] \ge M_0 = |\mathbf{x}|^\gamma\ .$$

Hence, for $n \geq K_\gamma$, by (6.18), we have

$$\begin{aligned}\mathbb{E}|\widehat{\mathbf{X}}_{n\wedge\widehat{\tau}_R(\mathbf{x})}|^\gamma &\geq \mathbb{E}\left[\sum_{k=0}^{n\wedge\widehat{\tau}_R(\mathbf{x})-1}\mathbb{E}\left[F(\widehat{\mathbf{X}}_k)\right](\mathbf{1}\{\widehat{\tau}_R(\mathbf{x})\geq K_\gamma\}+\mathbf{1}\{\widehat{\tau}_R(\mathbf{x})<K_\gamma\})\right]\\ &\geq \mathbb{E}\left[\frac{1}{2}(n\wedge\widehat{\tau}_R(\mathbf{x}))\mathbf{1}\{\widehat{\tau}_R(\mathbf{x})\geq K_\gamma\}+\sum_{k=0}^{n\wedge\widehat{\tau}_R-1}(-G_\gamma)\mathbf{1}\{\widehat{\tau}_R(\mathbf{x})<K_\gamma\}\right]\\ &\geq \frac{1}{2}\mathbb{E}\left[n\wedge\widehat{\tau}_R(\mathbf{x})\right]-K_\gamma(1+G_\gamma)\ ,\end{aligned} \tag{6.19}$$

where the second inequality follows from $F \geq -G_\gamma$.

By Proposition 5.5 and (6.17), we have $\widehat{\tau}_R(\mathbf{x}) < \infty$ almost surely, and therefore by the monotone convergence theorem,

$$\mathbb{E}\left[n\wedge\widehat{\tau}_R(\mathbf{x})\right] \nearrow \mathbb{E}\widehat{\tau}_R(\mathbf{x}) \quad \text{as } n\to\infty\ .$$

On the other hand, since

$$|\widehat{\mathbf{X}}_{n\wedge\widehat{\tau}_R(\mathbf{x})}|^\gamma \leq \max\left\{|\widehat{\mathbf{X}}_{\widehat{\tau}_R(\mathbf{x})}|^\gamma,\, R^\gamma\right\}\ ,$$

and since the random variable at the right-hand side has finite expectation by Lemma 6.10, by the dominated convergence theorem (along with the fact that $\widehat{\tau}_R(\mathbf{x}) < \infty$ almost surely), we get

$$\lim_{n\to\infty}\mathbb{E}|\widehat{\mathbf{X}}_{n\wedge\widehat{\tau}_R(\mathbf{x})}|^\gamma = \mathbb{E}|\widehat{\mathbf{X}}_{\widehat{\tau}_R(\mathbf{x})}|^\gamma \leq C_\gamma\left(R^\gamma+1\right)\ ,$$

where the last inequality follows from Lemma 6.10. Therefore, letting $n\to\infty$ in (6.19) and applying $\tau_R(\mathbf{x}) \leq \widehat{\tau}_R(\mathbf{x})N_\gamma$, we get (6.1). □

6.4. **Lower bound of mean exit time.** For the lower bound, we need the following construction of the Lyapunov function associated with process (2.1).

Lemma 6.11. *Let $r>0$. Suppose that $\mathbf{U}$ is a $d\times d$ random matrix and $\mathbf{V}$ is a d-dimensional random vector satisfying*

$$\mathbb{E}\left[\|\mathbf{U}\|^r\right]\leq 1, \quad \text{and} \quad \mathbb{E}\left[|\mathbf{V}|^r\right]\in(0,\,\infty)\ , \tag{6.20}$$

and $\mathbb{E}\left[\|\mathbf{U}\|^s\right]<1$ for all $s\in(0,\,r)$. Then, there exist constants $c_1,\dots,c_{\lfloor r\rfloor}>0$ such that the function

$$g(\mathbf{x}) = g_{\mathbf{U},\mathbf{V}}(\mathbf{x}) := \begin{cases}|\mathbf{x}|^r + c_1|\mathbf{x}|^{r-1}+\cdots+c_{\lfloor r\rfloor}|\mathbf{x}|^{r-\lfloor r\rfloor} & \text{if } r>1\\ |\mathbf{x}|^r & \text{if } r\leq 1\end{cases} \tag{6.21}$$

satisfies, for all $\mathbf{x}\in\mathbb{R}^d$,

$$\mathbb{E}\left[g(\mathbf{U}\mathbf{x}+\mathbf{V})\right]-g(\mathbf{x}) \leq \mathbb{E}g(\mathbf{V})\ .$$

We postpone the proof of this lemma to the next subsection and prove Proposition 6.2 by assuming it. For $\gamma\in(0,\,\alpha)$, by Theorem 2.3-(4), we have

$$\lim_{n\to\infty}\left(\mathbb{E}\|\mathbf{A}_1\cdots\mathbf{A}_n\|^\gamma\right)^{1/n} = h_{\mathbf{A}}(\gamma) < 1\ .$$

Thus, we can find $L_\gamma \ge 1$ such that

$$\mathbb{E}\|\mathbf{A}_1 \cdots \mathbf{A}_{L_\gamma}\|^\gamma < 1 \ . \tag{6.22}$$

For $n \in \mathbb{Z}^+$ and $i \in [\![0,\, L_\gamma - 1]\!]$, denote by $\widehat{\mathcal{F}}_n^{(i)}$ the σ-algebra generated by $(\mathbf{A}_k,\, \mathbf{B}_k)_{k\in[\![1,\, nL_\gamma+i]\!]}$ and write

$$\widehat{\mathbf{X}}_n^{(i)}(\mathbf{x}) := \mathbf{X}_{nL_\gamma+i}(\mathbf{x}) \quad ;\ \mathbf{x} \in \mathbb{R}^d$$

which is $\widehat{\mathcal{F}}_n^{(i)}$-measurable random variable. We again note that $\widehat{\mathcal{F}}_n^{(i)}$ and $\widehat{\mathbf{X}}_n^{(i)}(\mathbf{x})$ depend on γ although we did not highlight the dependency. The following lemma is a direct consequence of Lemma 6.11.

Lemma 6.12. *For $\gamma \in (0,\, \alpha)$, there exist constants $c_1,\, \dots,\, c_{\lfloor \gamma \rfloor} > 0$ (depending on γ) such that the function $g = g_\gamma : \mathbb{R}^d \to \mathbb{R}$ defined by*

$$g(\mathbf{x}) = |\mathbf{x}|^\gamma + c_1|\mathbf{x}|^{\gamma-1} + \cdots + c_{\lfloor\gamma\rfloor}|\mathbf{x}|^{\gamma-\lfloor\gamma\rfloor}$$

satisfies, for all $n \in \mathbb{Z}^+$, $i \in [\![0,\, L_\gamma - 1]\!]$, and $\mathbf{x} \in \mathbb{R}^d$,

$$\mathbb{E}\left[g(\widehat{\mathbf{X}}_{n+1}^{(i)}(\mathbf{x}))\,|\,\widehat{\mathcal{F}}_n^{(i)}\right] \le g(\widehat{\mathbf{X}}_n^{(i)}(\mathbf{x})) + \mathbb{E}\left[g(\mathbf{X}_{L_\gamma}(\mathbf{0}))\right] \ .$$

Proof. Fix $\gamma \in (0,\, \alpha)$. Fix $\mathbf{x} \in \mathbb{R}^d$ and simply write $\widehat{\mathbf{X}}_n^{(i)} = \widehat{\mathbf{X}}_n^{(i)}(\mathbf{x})$. Then, for all $n \in \mathbb{Z}^+$ and $i \in [\![0,\, L_\gamma - 1]\!]$, we can write

$$\widehat{\mathbf{X}}_{n+1}^{(i)} = \mathbf{U}_n^{(i)}\widehat{\mathbf{X}}_n^{(i)} + \mathbf{V}_n^{(i)} \tag{6.23}$$

where

$$\mathbf{U}_n^{(i)} = \mathbf{A}_{(n+1)L_\gamma+i}\mathbf{A}_{(n+1)L_\gamma+(i-1)} \cdots \mathbf{A}_{nL_\gamma+i+1} \ ,$$
$$\mathbf{V}_n^{(i)} = \mathbf{A}_{(n+1)L_\gamma+i} \cdots \mathbf{A}_{nL_\gamma+(i+2)}\mathbf{B}_{nL_\gamma+(i+1)} + \cdots + \mathbf{A}_{(n+1)L_\gamma+i}\mathbf{B}_{(n+1)L_\gamma+(i-1)} + \mathbf{B}_{(n+1)L_\gamma+i} \ .$$

In particular, by (6.22), we have $\mathbb{E}\|\mathbf{U}_n^{(i)}\|^\gamma < 1$ and therefore, by Proposition 6.11, there exists a function $g = g_\gamma : \mathbb{R}^d \to \mathbb{R}$ (independent of i) of the form

$$g(\mathbf{x}) = |\mathbf{x}|^\gamma + c_1|\mathbf{x}|^{\gamma-1} + \cdots + c_{\lfloor\gamma\rfloor}|\mathbf{x}|^{\gamma-\lfloor\gamma\rfloor}$$

with $c_1,\, \dots,\, c_{\lfloor\gamma\rfloor} > 0$ such that (cf. (6.23))

$$\mathbb{E}\left[g(\widehat{\mathbf{X}}_{n+1}^{(i)})\,|\,\widehat{\mathcal{F}}_n^{(i)}\right] \le g(\widehat{\mathbf{X}}_n^{(i)}) + \mathbb{E}\left[g(\mathbf{V}_n^{(i)})\right] \ . \tag{6.24}$$

The proof is completed since $\mathbf{V}_n^{(i)}$ and $\mathbf{X}_{L_\gamma}(\mathbf{0})$ obey the same distribution by (5.1). ☐

Now we are ready to prove Proposition 6.2.

Proof of Proposition 6.2. Fix $\gamma \in (0,\, \alpha)$ and $R > 0$. Fix $\mathbf{x} \in \mathbb{R}^d$ and simply write $\widehat{\mathbf{X}}_n^{(i)} = \widehat{\mathbf{X}}_n^{(i)}(\mathbf{x})$. Writing

$$G_\gamma := \mathbb{E}\left[g(\mathbf{X}_{L_\gamma}(\mathbf{0}))\right] \tag{6.25}$$

so that $G_\gamma > 0$ because of Lemma (5.4)-(2). Then, define a sequence of random variables $(M_n^{(i)})_{n\in\mathbb{Z}_0^+}$ by

$$M_n^{(i)} = g(\widehat{\mathbf{X}}_n^{(i)}) - nG_\gamma \quad ;\ n \in \mathbb{Z}_0^+$$

so that $(M_n^{(i)})_{n\in\mathbb{Z}_0^+}$ is not only adapted to the filtration $(\widehat{\mathcal{F}}_n^{(i)})_{n\in\mathbb{Z}_0^+}$, but also a $(\widehat{\mathcal{F}}_n^{(i)})_{n\in\mathbb{Z}_0^+}$-supermartingale by Lemma 6.12.

We next define, for $i \in [\![0,\, L_\gamma - 1]\!]$,

$$\widehat{\tau}_R^{(i)}(\mathbf{x}) = \inf\left\{n \in \mathbb{Z}^+ : \tau_R(\mathbf{x}) \le nL_\gamma + i\right\} \tag{6.26}$$

so that $\widehat{\tau}_R^{(i)}(\mathbf{x})$ is a $(\mathcal{F}_n^{(i)})_{n\in\mathbb{Z}_0^+}$-stopping time. Thus, by the optional stopping theorem, for all $n \in \mathbb{Z}^+$, we have

$$\mathbb{E}M^{(i)}_{n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})} \le \mathbb{E}M_0^{(i)} = \mathbb{E}g(\mathbf{X}_i(\mathbf{x}))\ . \tag{6.27}$$

By Lemma 6.9, we have

$$\max_{i\in[\![0,\, L_\gamma-1]\!]} \mathbb{E}g(\mathbf{X}_i(\mathbf{x})) \le C_\gamma(g(\mathbf{x}) + 1)\ .$$

Inserting this and the fact that

$$\mathbb{E}M^{(i)}_{n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})} = \mathbb{E}g(\widehat{\mathbf{X}}^{(i)}_{n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})}) - \mathbb{E}[n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})]\cdot G_\gamma$$

to (6.27) yields that, for some constants $C_\gamma,\, D_\gamma > 0$,

$$\mathbb{E}[n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})] \ge C_\gamma\mathbb{E}g(\widehat{\mathbf{X}}^{(i)}_{n\wedge\widehat{\tau}_R^{(i)}(\mathbf{x})}) - D_\gamma(|\mathbf{x}|^\gamma + 1)\ .$$

Since $\mathbb{P}[\widehat{\tau}_R^{(i)}(\mathbf{x}) < \infty] = 1$ by Proposition 5.5, we can apply the monotone convergence theorem and Fatou's lemma at the left- and right-hand side, respectively, along $n \to \infty$, we get

$$\mathbb{E}[\widehat{\tau}_R^{(i)}(\mathbf{x})] \ge C_\gamma\mathbb{E}g(\widehat{\mathbf{X}}^{(i)}_{\widehat{\tau}_R^{(i)}(\mathbf{x})}) - D_\gamma(|\mathbf{x}|^\gamma + 1)\ . \tag{6.28}$$

Noting from (6.26) that

$$(\widehat{\tau}_R^{(i)}(\mathbf{x}) - 1)L_\gamma + i < \tau_R(\mathbf{x})$$

and thus we get from (6.28) (along with trivial bound $g(\mathbf{x}) \ge |\mathbf{x}|^\gamma$) that

$$\mathbb{E}[\tau_R(\mathbf{x})] \ge C_\gamma \max_{i\in[\![0,\, L_\gamma-1]\!]} \mathbb{E}\left|\widehat{\mathbf{X}}^{(i)}_{\widehat{\tau}_R^{(i)}(\mathbf{x})}\right|^\gamma - D_\gamma(|\mathbf{x}|^\gamma + 1)\ . \tag{6.29}$$

On the other hand, in view of the definition of the stopping time $\widehat{\tau}_R^{(i)}(\mathbf{x})$, one of $\widehat{\mathbf{X}}^{(0)}_{\widehat{\tau}_R^{(0)}(\mathbf{x})},\, \widehat{\mathbf{X}}^{(1)}_{\widehat{\tau}_R^{(1)}(\mathbf{x})},\, \dots,\, \widehat{\mathbf{X}}^{(L_\gamma-1)}_{\widehat{\tau}_R^{(L_\gamma-1)}(\mathbf{x})}$ coincides with $\mathbf{X}_{\tau_R(\mathbf{x})}(\mathbf{x})$ and therefore,

$$\max_{i\in[\![0,\, L_\gamma-1]\!]} \mathbb{E}\left|\mathbf{Y}^{(i)}_{\widehat{\tau}_R^{(i)}(\mathbf{x})}\right|^\gamma \ge R^\gamma\ .$$

Inserting this to (6.29) completes the proof. □

6.5. **Proof of Lemma 6.11.** To prove Lemma 6.11, we need two elementary facts. For $a \in \mathbb{R}$ and $k \in \mathbb{Z}_0^+$, write

$$\binom{a}{k} = \frac{a(a-1)\cdots(a-k+1)}{k!}\ .$$

Lemma 6.13. *For all $x,\, y \ge 0$ and $r \ge 0$, we have*

$$(x+y)^r \le \left(x^r + \binom{r}{1} x^{r-1}y + \cdots + \binom{r}{\lfloor r \rfloor} x^{r-\lfloor r \rfloor} y^{\lfloor r \rfloor} \right) + y^r \ ,$$

where the right-hand side is $x^r + y^r$ when $r < 1$.

Proof. Write $\lfloor r \rfloor = n$ so that $n \le r < n+1$. Fix $y > 0$ and $n \in \mathbb{Z}^+$ (since the inequality is straighforward if $y = 0$ or $n = 0$) and define $f : [0, \infty) \to \mathbb{R}$ as

$$f(x) = \left(x^r + \binom{r}{1} x^{r-1}y + \cdots + \binom{r}{n} x^{r-n} y^n \right) + y^r - (x+y)^r.$$

Then, a straight forward computation yields

$$f^{(n+1)}(x) = r(r-1)\cdots(r-n)(x^{r-n-1} - (x+y)^{r-n-1}) \ ,$$

where $f^{(k)}$ denotes the kth derivative of f. Since $r - n - 1 < 0$, we have $f^{(n+1)} > 0$. Since we can readily check that $f(0) = f'(0) = \cdots = f^{(n)}(0) = 0$, we can inductively show that $f^{(n)} > 0, \ \ldots\ , f^{(1)} > 0$ and finally, $f > 0$. □

For a vector $\mathbf{v} = (v_1,\, \ldots,\, v_n) \in \mathbb{R}^n$, we write $\mathbf{v} > 0$ if $v_1,\, \ldots,\, v_n > 0$, and $\mathbf{v} \ge 0$ if $v_1,\, \ldots,\, v_n \ge 0$.

Lemma 6.14. *Let $\mathbb{H}$ be an $n \times n$ upper triangular matrix such that $\mathbb{H}_{1,\,1} \ge 0$ and $\mathbb{H}_{i,\,i} > 0$ for all $i \ge 2$. Then, there exists $\mathbf{v} = (v_1,\, \ldots,\, v_n) \in \mathbb{R}^n$ such that $\mathbf{v} > 0$ and $\mathbf{v}\mathbb{H} \ge 0$.*

Proof. We use induction on n. There is nothing to prove when $n = 1$. Suppose next that the statement holds for $n = k$. If $\mathbb{H}$ is a $(k+1) \times (k+1)$ upper triangular matrix such that all of its diagonal entries are positive. Let $\widetilde{\mathbb{H}}$ be a $k \times k$ upper triangular matrix defined by $\widetilde{\mathbb{H}}_{i,\,j} = \mathbb{H}_{i,\,j}$ for all $1 \le i,\, j \le k$. Then, since $\widetilde{\mathbb{H}}$ is an upper triangular matrix satisfying the condition for induction hypothesis, we can find $\widetilde{\mathbf{v}} = (v_1,\, \ldots,\, v_k) > 0$ such that $\widetilde{\mathbf{v}}\widetilde{\mathbb{H}} \ge 0$. Then, let $\mathbf{v} = (v_1,\, \ldots,\, v_k,\, t)$, so that

$$\mathbf{v}\mathbb{H} = \left(\widetilde{\mathbf{v}}\widetilde{\mathbb{H}},\ \sum_{i=1}^{k} v_i \mathbb{H}_{i,\,k+1} + t\mathbb{H}_{k+1,\,k+1} \right) \ .$$

Since $\mathbb{H}_{k+1,\,k+1} > 0$, we can take $t > 0$ large enough so that $\mathbf{v}\mathbb{H} \ge 0$. □

Let us now prove Lemma 6.11.

Proof of Lemma 6.11. Write $\lfloor r \rfloor = n$ so that $n \le r < n+1$. We define an upper triangular $(n+1) \times (n+1)$ matrix $\mathbb{H} = (h_{i,\,j})_{i,\,j \in [\![0,\,n]\!]}$ by

$$\begin{cases} h_{i,\,i} = \mathbb{E}\|\mathbf{U}\|^{r-i} - 1 & \text{for } i \in [\![0,\,n]\!] \\ h_{i,\,k} = \dbinom{r-i}{k-i} (\mathbb{E}[\|\mathbf{U}\|^{r-k}|\mathbf{V}|^{k-i}]) & \text{for } i \in [\![0,\,n]\!] \text{ and } k > i \\ 0 & \text{otherwise .} \end{cases}$$

Then, for all $i \in [\![0, n]\!]$, by Lemma 6.13,

$$\mathbb{E}\left[|\mathbf{U}\mathbf{x}+\mathbf{V}|^{r-i}\right] - |\mathbf{x}|^{r-i} \le \mathbb{E}(\|\mathbf{U}\| \cdot |\mathbf{x}| + |\mathbf{V}|)^{r-i} - |\mathbf{x}|^{r-i} \le \sum_{k=i}^{n} h_{i,\,k} |\mathbf{x}|^{r-k} + \mathbb{E}|\mathbf{V}|^{r-i} \ . \tag{6.30}$$

From (6.20), we have $h_{0,\,0} = 0$ and $h_{i,\,i} < 0$ for all $i \in [\![1, n]\!]$. Namely, the diagonal entry of the upper triangular matrix $\mathbb{H}$ is negative except for $(1, 1)$-component which is exactly zero. Hence, by Lemma 6.14, we can find a vector $\mathbf{c} = (1, c_1, \ldots, c_n) > 0$ such that all the entries of $\mathbf{c}\mathbb{H}$ are non-positive. By taking $g(\mathbf{x})$ as in (6.21) with this $c_1, \ldots, c_n$, we get from (6.30) that

$$\begin{aligned}\mathbb{E}\left[g(\mathbf{U}\mathbf{x}+\mathbf{V})\right] - g(\mathbf{x}) &\le \mathbb{E}g(\mathbf{V}) + \sum_{i=0}^{n}\sum_{k=0}^{n} c_i h_{i,\,k} |\mathbf{x}|^{r-k} \ . \\ &= \mathbb{E}g(\mathbf{V}) + \sum_{k=0}^{n} (\mathbf{c}\mathbb{H})_k |\mathbf{x}|^{r-k} \le \mathbb{E}g(\mathbf{V}) \ .\end{aligned}$$

□

6.6. **Univariate case.** We now analyze one-dimensional, i.e., the univariate case, for which we can obtain more concrete result than the multivariate case. We remark here that in the univariate case, by the independence of the sequence $(\mathbf{A}_n)_{n\in\mathbb{Z}^+}$, we have

$$\frac{1}{n}\log \mathbb{E}\,|\mathbf{A}_1 \cdots \mathbf{A}_n|^s = \mathbb{E}|\mathbf{A}|^s$$

and therefore we can simplify $h_{\mathbf{A}}(\cdot)$ into

$$h_{\mathbf{A}}(s) = \mathbb{E}|\mathbf{A}|^s \ . \tag{6.31}$$

Theorem 6.15. *Let $d = 1$.*

(1) There exist constants $C_1, C_2 > 0$ such that, for all $\mathbf{x}_0 \in \mathbb{R}$,

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \ge C_1 R^\alpha - C_2(|\mathbf{x}_0|^\alpha + 1) \ . \tag{6.32}$$

(2) If $\alpha \ge 2$, there exist constants $C_3 > 0$ such that, for all $\mathbf{x}_0 \in \mathbb{R}$,

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \le C_3\left(R^\alpha + 1\right) \ .$$

Before proving this theorem, we establish an elementary inequality.

Lemma 6.16. *For all $\alpha \ge 2$, there exists a constant $C_\alpha > 0$ such that, for all $x, y \in \mathbb{R}$*

$$|x+y|^\alpha \ge |x|^\alpha + \alpha|x|^{\alpha-2}xy + C_\alpha|y|^\alpha \ .$$

Proof. The statement is clear for $y = 0$ or $\alpha = 2$ (for which $C_\alpha = 1$). Hence, assume that $y \ne 0$ and $\alpha > 2$. By renormalize suitably, we can assume that $y = 1$ and hence it suffices to prove that there exists a constant $C_\alpha > 0$ such that for all $x \in \mathbb{R}$,

$$|x+1|^\alpha \ge |x|^\alpha + \alpha|x|^{\alpha-2}x + C_\alpha.$$

Let $f(x) = |x+1|^\alpha - \big(|x|^\alpha + \alpha|x|^{\alpha-2}x\big)$ and then it suffices to check that $\inf_{x\in\mathbb{R}} f(x) > 0$. By a simple computation, one can directly check that $f' < 0$ on $(-\infty, -1]$ and $f' > 0$ on $(0, \infty)$. Hence, it suffices to check that

$$\inf_{x\in[-1,\,0]} f(x) = \inf_{x\in[0,\,1]} f(-x) > 0\,. \tag{6.33}$$

Writing $g(x) = f(-x)$, we have

$$g'(x) = \alpha x^{\alpha-1}\left(-\left(\frac{1}{x}-1\right)^{\alpha-1} - 1 + (\alpha-1)\frac{1}{x}\right)\,.$$

By an elementary computation, we can check that there exists a unique solution $x_0 \in (0,\,1)$ of $g'(x_0) = 0$ and $g' < 0$ on $(0,\,x_0)$ and $g' > 0$ on $(x_0,\,1)$. Hence, g attains its minimum on $(0,\,1)$ at x_0. Furthermore, we can check that

$$g(x_0) = (\alpha-1)x_0^{\alpha-2} > 0$$

and therefore we get (6.33). $\square$

Now we turn to the proof of Theorem 6.15

Proof of Theorem 6.15. (1) Since $h_{\mathbf{A}}(\alpha) = 1$, by (6.31), we have $\mathbb{E}|\mathbf{A}|^\alpha = 1$. Therefore, the proof of Proposition 6.2 works for $\gamma = \alpha$ in univariate case, and hence we get (6.32).
(2) Let us fix $\mathbf{x}_0 \in \mathbb{R}$. We first assume that

$$\mathbb{E}\left[|\mathbf{A}|^{\alpha-2}\mathbf{A}\mathbf{B}\right] = 0\,. \tag{6.34}$$

By this assumption, Lemma 6.16, and the fact that $\mathbb{E}[|\mathbf{A}|^\alpha] = 1$, for each $n \in \mathbb{Z}^+$, we have (regardless of the starting point of the process (2.1)),

$$\begin{aligned}
\mathbb{E}[|\mathbf{X}_{n+1}|^\alpha \mid \mathbf{X}_n] - |\mathbf{X}_n|^\alpha &= \mathbb{E}[|\mathbf{A}_{n+1}\mathbf{X}_n + \mathbf{B}_{n+1}|^\alpha - |\mathbf{A}_{n+1}\mathbf{X}_n|^\alpha \mid \mathbf{X}_n]\\
&\geq \mathbb{E}[\alpha|\mathbf{A}_{n+1}\mathbf{X}_n|^{\alpha-2}(\mathbf{A}_{t+1}\mathbf{X}_n)\mathbf{B}_{n+1} + C_\alpha|\mathbf{B}_{n+1}|^\alpha \mid \mathbf{X}_n]\\
&= \alpha\mathbb{E}\left[|\mathbf{A}|^{\alpha-2}\mathbf{A}\mathbf{B}\right]\mathbf{X}_n + C_\alpha\mathbb{E}|\mathbf{B}|^\alpha\\
&= C_\alpha\mathbb{E}|\mathbf{B}|^\alpha\,.
\end{aligned}$$

Let $\kappa_0 = C_\alpha\mathbb{E}|\mathbf{B}|^\alpha$ so that by the previous observation, the sequence $(M_n)_{n\in\mathbb{Z}^+}$ defined by

$$M_n := \mathbf{X}_n(\mathbf{x}_0) - \kappa_0 n$$

is a $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$-submartingale. By the optional stopping theorem, we get, for all $n \in \mathbb{Z}^+$,

$$\mathbb{E}|\mathbf{X}_{n\wedge\tau_R}(\mathbf{x}_0)|^\alpha - \kappa_0\mathbb{E}[n\wedge\tau_R(\mathbf{x}_0)] \geq |\mathbf{x}_0|^\alpha\,.$$

By letting $n \to \infty$ and apply the same logic with the proof of Proposition 6.1, we get, for some constant $C_0 > 0$,

$$\mathbb{E}[\tau_R(\mathbf{x}_0)] \leq C_0 R^\alpha\,. \tag{6.35}$$

Now we consider the general case without the assumption (6.34). Since $\alpha \geq 2$, by Theorem 2.3 we have

$$\mathbb{E}|\mathbf{A}|^{\alpha-2}\mathbf{A} \leq \mathbb{E}|\mathbf{A}|^{\alpha-1} = h(\alpha-1) < h(\alpha) = \mathbb{E}|\mathbf{A}|^{\alpha} \ .$$

Therefore, we can take $c_0 \in \mathbb{R}$ such that

$$\mathbb{E}\left[|\mathbf{A}|^{\alpha-2}\mathbf{A}\mathbf{B}\right] + c_0\left(\mathbb{E}|\mathbf{A}|^{\alpha} - \mathbb{E}|\mathbf{A}|^{\alpha-2}\mathbf{A}\right) = 0 \ . \tag{6.36}$$

Then, we set

$$\begin{cases} \widetilde{\mathbf{X}}_n := \mathbf{X}_n(\mathbf{x}_0) - c_0 \ , & n \geq 0 \ , \\ \widetilde{\mathbf{A}}_n := \mathbf{A}_n \quad \text{and} \quad \widetilde{\mathbf{B}}_n := \mathbf{B}_n + c_0\mathbf{A}_n - c_0 \ , & n \geq 1 \ , \end{cases}$$

so that we have

$$\widetilde{\mathbf{A}}_{n+1}\widetilde{\mathbf{X}}_n + \widetilde{\mathbf{B}}_{n+1} = \mathbf{A}_{n+1}(\mathbf{X}_n(\mathbf{x}_0) - c_0) + \mathbf{B}_{n+1} + c_0\mathbf{A}_{n+1} - c_0 = \widetilde{\mathbf{X}}_{n+1} \ .$$

Hence, $(\widetilde{\mathbf{X}}_n)_{n\in\mathbb{Z}^+}$ is another process (2.1) associated with $(\widetilde{\mathbf{A}}_n, \widetilde{\mathbf{B}}_n)_{n\in\mathbb{Z}^+}$. Since

$$\mathbb{E}\left[|\widetilde{\mathbf{A}}|^{\alpha-2}\widetilde{\mathbf{A}}\widetilde{\mathbf{B}}\right] = \mathbb{E}\left[|\mathbf{A}|^{\alpha-2}\mathbf{A}\mathbf{B}\right] + c\left(\mathbb{E}|\mathbf{A}|^{\alpha} - \mathbb{E}|\mathbf{A}|^{\alpha-2}\mathbf{A}\right) = 0$$

by (7.3), we get from the first part of the proof (cf. (6.35)) that

$$\mathbb{E}[\widetilde{\tau}_R] \leq C_0 R^{\alpha}$$

where $\widetilde{\tau}_R := \inf\{n \in \mathbb{Z}^+ : |\widetilde{\mathbf{X}}_n| > R\}$. By the definition of $\widetilde{\mathbf{X}}_n$, we have

$$\tau_R(\mathbf{x}_0) \leq \widetilde{\tau}_{R+|c_0|} \ ,$$

and therefore we get

$$\mathbb{E}[\tau_R(\mathbf{x}_0)] \leq \mathbb{E}[\widetilde{\tau}_{R+|c_0|}] \leq C_0(R+c_0)^{\alpha} \ .$$

This completes the proof. □

7. Proofs of Main Results: Explosive Regime

The purpose of the current section is to prove Theorem 3.7. Hence, we assume that the Lyapunov exponent γ_L is positive, namely $\gamma_L > 0$, and moreover assume Assumption 6 throughout this section. The proof of Theorem 3.7 will be divided into the lower and the upper bound. We provide a preliminary results regarding explosive regime in Section 7.1, and then prove the lower and the upper bound in Sections 7.2 and 7.3, respectively.

7.1. **Preliminary results.**

To derive Theorem 3.7, we need the following technical lemma.

Lemma 7.1. *If* $\mathbf{A}$ *is irreducible (cf. Assumption 6), we almost surely have that*

$$\lim_{n\to\infty} \frac{1}{n} \log \|\Pi_n^{-1}\| = -\gamma_L$$

Proof. This result is a direct application of Oseledet's theorem; see [8, Theorem 4.2. page 39]. □

Now we proof the boundedness of the exit time for the explosive regime.

Proposition 7.2. *We have* $\mathbb{P}(\tau_R(\mathbf{x}_0) < \infty) = 1$ *for all* $\mathbf{x} \in \mathbb{R}^d$ *and* $R > 0$*.*

Proof. We fix $\mathbf{x} \in \mathbb{R}^d$ and $R > 0$, and suppose on the contrary that

$$\mathbb{P}\left[\tau_R(\mathbf{x}_0) = \infty\right] > 0 \ .$$

The same argument as in the proof of proposition 5.5 for contractive regime yields the existence of probability measure μ_∞ satisfying

$$\mathbb{P}_{\mu_\infty}\left[\mathbf{X}_n \in \mathcal{B}_R\right] = 1 \tag{7.1}$$

for all $n \in \mathbb{Z}^+$. That means if we pick two independent random variables $\mathbf{Y}_1, \mathbf{Y}_2 \overset{d}{\sim} \mu_\infty$, we have that

$$\mathbb{P}_{\mu_\infty}\left[|\mathbf{X}_n(\mathbf{Y}_i)| \le R\right] = 1, \quad i = 1, 2,\, n \in \mathbb{Z}^+.$$

Considering coupling method (5.2)

$$|\mathbf{X}_n(\mathbf{Y}_1) - \mathbf{X}_n(\mathbf{Y}_2)| = |\Pi_n(\mathbf{Y}_1 - \mathbf{Y}_2)| \ge \|\Pi_n^{-1}\|^{-1} \cdot |\mathbf{Y}_1 - \mathbf{Y}_2|,$$

and Lemma 7.1

$$\lim_{n\to\infty} \frac{1}{n} \log \|\Pi_n^{-1}\|^{-1} = \gamma_L > 0$$

gives us that

$$\mathbb{P}(\mathbf{Y}_1 = \mathbf{Y}_2) = 1$$

if we send $n \to \infty$. Since we drew $\mathbf{Y}_1, \mathbf{Y}_2$ independently from μ_∞, this happens only if when the support of μ_∞ is a singleton $\{\mathbf{x}_\infty\}$ for some $\mathbf{x}_\infty \in \mathbb{R}^d$. Repeatedly, one step forward from (7.1) gives

$$\mathbb{P}_{\mathbf{x}_\infty}\left[\mathbf{X}_n(\mathbf{A}_1\mathbf{x}_\infty + \mathbf{B}_1) \in \mathcal{B}_R\right] = 1, \quad \forall n \in \mathbb{Z}^+.$$

Again, picking $\mathbf{Y}_1 = \mathbf{x}_\infty$ and $\mathbf{Y}_2 = \mathbf{A}_1'\mathbf{x}_\infty + \mathbf{B}_1'$ where $(\mathbf{A}_1', \mathbf{B}_1')$ is independent of ω, the same argument above gives that

$$\mathbb{P}(\mathbf{A}_1'\mathbf{x}_\infty + \mathbf{B}_1' = \mathbf{x}_\infty) = 1,$$

which is contradiction to the assumption. □

Lemma 7.3. *There exists a constant* $C_1 > 0$ *such that, for all* $R > R_0$ *(where the constant* R_0 *is the one appeared in Assumption 6) and* $\mathbf{x}_0 \in \mathcal{B}_R$*, we have*

$$\mathbb{E} \log |\mathbf{X}_{\tau_R}(\mathbf{x}_0)| \le C_1 + \log R \ .$$

Proof. Let $R > R_0$. For $\mathbf{x} \in \mathcal{B}_R$, by the layer-cake formula and the fact that

$$\lim_{z\to\infty} \log z \cdot \mathbb{P}(|\mathbf{A}\mathbf{x} + \mathbf{B}| > z) = 0$$

which follows from (3.1), we can write

$$\mathbb{E}\left[\log|\mathbf{A}\mathbf{x} + \mathbf{B}| \cdot \mathbf{1}\left\{|\mathbf{A}\mathbf{x} + \mathbf{B}| > R\right\}\right] = \log R \cdot \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > R) + \int_1^\infty \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > Rz) \cdot \frac{1}{z} dz \ . \tag{7.2}$$

By (3.10) of Assumption 6, (remind that $\beta_0 > 1$)

$$\begin{aligned}
&\int_1^\infty \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > Rz)\cdot\frac{1}{z}dz \\
&= \int_1^{z_0} \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > Rz)\cdot\frac{1}{z}dz + \int_{z_0}^\infty \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > Rz)\cdot\frac{1}{z}dz \\
&\le \int_1^{z_0} \mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > R)\cdot\frac{1}{z}dz + \int_{z_0}^\infty \frac{C_0}{(\log z)^{\beta_0}}\cdot\mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > R)\cdot\frac{1}{z}dz \\
&= \left(\log z_0 + \frac{C_0}{(\beta_0-1)(\log z)^{\beta_0-1}}\right)\mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > R)\ . \qquad (7.3)
\end{aligned}$$

Inserting this to (7.2), and dividing both sides by $\mathbb{P}(|\mathbf{A}\mathbf{x}+\mathbf{B}| > R)$, we get, for some constant $C > 0$,

$$\mathbb{E}\left[\log|\mathbf{A}\mathbf{x}+\mathbf{B}|\,\big|\,|\mathbf{A}\mathbf{x}+\mathbf{B}| > R\right] \le C + \log R\ .$$

This along with the strong Markov property proves the lemma since we have

$$\mathbb{E}\log|\mathbf{X}_{\tau_R}(\mathbf{x}_0)| \le \sup_{\mathbf{x}\in\mathcal{B}_R}\mathbb{E}\left[\log|\mathbf{A}\mathbf{x}+\mathbf{B}|\,\big|\,|\mathbf{A}\mathbf{x}+\mathbf{B}| > R\right]\ .$$

□

7.2. **Lower bound.** We first handle a special univariate case which will be used in the proof of the lower bound for the multivariate case. We start from a specific univariate case.

Lemma 7.4. *Let $d = 1$. Suppose that $\mathbf{A} > 0$, $\mathbf{B} \ge 0$ almost surely, $\mathbb{P}(\mathbf{B} = 0) < 1$,*

$$\mathbb{E}|\log\mathbf{A}| < \infty\ , \quad \textit{and} \quad \mathbb{E}|\log\mathbf{B}| < \infty\ .$$

Let $\mathbf{x}_0 \ge 0$ so that by the previous conditions we have $\mathbf{X}_n(\mathbf{x}_0) \ge 0$ for all $n \in \mathbb{Z}^+$.

(1) We almost surely have that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbf{X}_n = \mathbb{E}\log\mathbf{A}\ . \qquad (7.4)$$

(2) We almost surely have that (cf. (3.6))

$$\lim_{R\to\infty}\frac{\tau_R(\mathbf{x}_0)}{\log R} = \frac{1}{\mathbb{E}\log\mathbf{A}} = \frac{1}{\gamma_L}\ . \qquad (7.5)$$

Proof. Fix $\mathbf{x}_0 \ge 0$, and for the convenience of the notation, we write $\mathbf{X}_n = \mathbf{X}_n(\mathbf{x}_0)$ when we do not need to specify $\mathbf{x}_0$.

(1) Let us first consider the case $\mathbf{x}_0 > 0$. Since we assumed that $\mathbf{B} \ge 0$, we have

$$\mathbf{X}_n \ge \mathbf{A}_n\mathbf{X}_{n-1} \ge \cdots \ge \mathbf{A}_n\cdots\mathbf{A}_1\mathbf{x}_0,$$

and therefore we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbf{X}_n \ge \liminf_{n\to\infty}\left(\frac{1}{n}\log\mathbf{x}_0 + \frac{1}{n}\sum_{i=1}^n\log\mathbf{A}_i\right) = \mathbb{E}\log\mathbf{A} \qquad (7.6)$$

almost surely by the strong law of large numbers. Note here that, for univariate model, $\mathbb{E}\log\mathbf{A} = \gamma_L > 0$.

For the other direction, we observe first that, for $n \in \mathbb{Z}^+$,

$$\begin{aligned}\log \mathbf{X}_n &= \log(\mathbf{A}_n \mathbf{X}_{n-1} + \mathbf{B}_n) \\ &= \log \mathbf{X}_{n-1} + \log \mathbf{A}_n + \log\left(1 + \frac{\mathbf{B}_n}{\mathbf{A}_n \mathbf{X}_{n-1}}\right) \\ &\leq \log \mathbf{X}_{n-1} + \log \mathbf{A}_n + \frac{\mathbf{B}_n}{\mathbf{A}_n \mathbf{X}_{n-1}} \ ,\end{aligned}$$

where the last line used the inequality $\log(1+x) \leq x$. Hence, we get

$$\frac{1}{n}\log \mathbf{X}_n \leq \frac{1}{n}\log \mathbf{x}_0 + \frac{1}{n}\sum_{i=1}^{n}\log \mathbf{A}_i + \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{B}_i}{\mathbf{A}_i \mathbf{X}_{i-1}} \ .$$

Therefore, by the same computation with (7.6), we get

$$\limsup_{n\to\infty}\frac{1}{n}\log \mathbf{X}_n \leq \mathbb{E}\log \mathbf{A} + \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{B}_i}{\mathbf{A}_i \mathbf{X}_{i-1}} \ . \tag{7.7}$$

Therefore, it suffices to prove that (as, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}_n \geq 0$ for all $n \in \mathbb{Z}^+$)

$$\sum_{k=0}^{\infty}\frac{\mathbf{B}_{k+1}}{\mathbf{A}_{k+1}\mathbf{X}_k} < \infty \quad \text{almost surely.} \tag{7.8}$$

To that end, observe first that, for i.i.d random variables $U_1, U_2, \ldots$ with $\mathbb{E}|U_1| < \infty$, we have

$$\lim_{n\to\infty}\frac{U_n}{n} = 0 \quad \text{almost surely,} \tag{7.9}$$

as we have

$$\mathbb{P}\left[\limsup_{n\to\infty}\frac{|U_n|}{n} > 0\right] \leq \lim_{\delta\to 0}\mathbb{P}\left[\limsup_{n\to\infty}\frac{|U_n|}{n} \geq \delta\right] = 0$$

where the last equality follows from the Markov inequality, the Borel-Cantelli lemma and the layer-cake formula, as

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{|U_n|}{n} \geq \delta\right) = \sum_{n=1}^{\infty}\mathbb{P}\left(\frac{|U_1|}{\delta} \geq n\right) \leq \frac{\mathbb{E}\left[|U_1|\right]}{\delta} < \infty \ .$$

Then, we have

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\left(\frac{\mathbf{B}_{n+1}}{\mathbf{A}_{n+1}\mathbf{X}_n}\right) &\leq \limsup_{n\to\infty}\frac{1}{n}\log\left(\frac{\mathbf{B}_{n+1}}{\mathbf{A}_{n+1}}\right) - \liminf_{n\to\infty}\frac{1}{n}\log \mathbf{X}_n \\ &= -\liminf_{n\to\infty}\frac{1}{n}\log \mathbf{X}_n \leq -\mathbb{E}\log \mathbf{A} < 0\end{aligned}$$

where the second line follows from (7.9), (7.6), and (3.6). The last bound along with Cauchy's criterion proves (7.8) and we are done.

Now let us consider the case $\mathbf{x}_0 = 0$. In view of coupling in (5.2), we have that

$$x \leq x' \ \text{ implies that } \ \mathbf{X}_n(x) \leq \mathbf{X}_n(x') \ \text{ holds for all } n \in \mathbb{Z}^+ \ . \tag{7.10}$$

Therefore, we immediately have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{X}_n(0) \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{X}_n(1) = \mathbb{E} \log \mathbf{A} \ , \tag{7.11}$$

where the last equality holds from the previous part.

For the lower bound in the case $\mathbf{x}_0 = 0$, we need some notation.

- Write $\omega = (\mathbf{A}_n,\, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$ and denote by Ω the sample space of ω which is a countable product of the copy of sample space of $(\mathbf{A},\, \mathbf{B})$.
- Write
$$\zeta = \inf\{n \ge 1 : \mathbf{B}_n > 0\} = \inf\{n \ge 1 : \mathbf{X}_n > 0\}$$
so that ζ is a geometric random variable with success probability $\mathbb{P}(\mathbf{B} \neq 0) > 0$, and hence $\mathbb{P}(\zeta < \infty) = 1$.
- For $k,\, \ell \in \mathbb{Z}^+$, we define the event $\Omega_{k,\,\ell}$ as
$$\Omega_{k,\,\ell} = \left\{ \zeta = k,\, \mathbf{B}_k \in \left[\frac{1}{\ell},\, \frac{1}{\ell - 1} \right) \right\}$$
where in the case $\ell = 1$, the last line reads as $\Omega_{k,\,1} = \{\zeta = k,\, \mathbf{B}_k \in [1,\, \infty)\}$. Then, it is clear that $(\Omega_{k,\,\ell})_{k,\,\ell\in\mathbb{Z}^+}$ is a decomposition of the sample space Ω.
- For $k \in \mathbb{Z}^+$ and $x > 0$, we write $(\mathbf{X}_n^{(k)}(x))_{n\in\mathbb{Z}_0^+}$ the process (2.1) defined by $\mathbf{X}_0^{(k)}(x) = x$ and
$$\mathbf{X}_{n+1}^{(k)}(x) = \mathbf{A}_{n+k+1}\mathbf{X}_n^{(k)}(x) + \mathbf{B}_{n+k+1} \quad ;\ n \in \mathbb{Z}^+ \ .$$
Note that $(\mathbf{X}_n^{(k)}(x))_{n\in\mathbb{Z}_0^+}$ has the same law with $(\mathbf{X}_n(x))_{n\in\mathbb{Z}_0^+}$ and moreover is independent of $(\mathbf{A}_n,\, \mathbf{B}_n)_{n=1}^k$.

Let us fix $k,\, \ell \in \mathbb{Z}^+$, and we claim now that, conditioning on $\Omega_{k,\,\ell}$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{X}_n(0) \ge \mathbb{E} \log \mathbf{A} \tag{7.12}$$

almost surely. Proving this claim along with (7.11) completes the proof of part (1).

To prove this claim, first observe that, conditioning on $\Omega_{k,\,\ell}$, by (7.10), we have

$$\mathbf{X}_n(0) \ge \mathbf{X}_{n-k}^{(k)}\left(\frac{1}{\ell}\right) \quad \text{for all } n \ge k \ . \tag{7.13}$$

Furthermore, the event $\Omega_{k,\,\ell}$ which is completely determined by $(\mathbf{A}_n,\, \mathbf{B}_n)_{n=1}^k$ is independent of $(\mathbf{X}_n^{(k)}(x))_{n\in\mathbb{Z}_0^+}$, and therefore, conditioning on $\Omega_{k,\,\ell}$ does not affect the evolution of the process $(\mathbf{X}_n^{(k)}(x))_{n\in\mathbb{Z}_0^+}$ and therefore, conditioning on $\Omega_{k,\,\ell}$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{X}_n^{(k)}\left(\frac{1}{\ell}\right) = \mathbb{E} \log \mathbf{A} \tag{7.14}$$

almost surely in $(\mathbf{A}_n,\, \mathbf{B}_n)_{n=k+1}^\infty$, and hence in $(\mathbf{A}_n,\, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$, by the previous step. Summing up (7.13) and (7.14), we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{X}_n(0) \ge \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{X}_{n-k}^{(k)}\left(\frac{1}{\ell}\right) = \liminf_{n\to\infty} \frac{1}{n-k} \log \mathbf{X}_{n-k}^{(k)}\left(\frac{1}{\ell}\right) = \mathbb{E} \log \mathbf{A}$$

almost surely. This completes the proof of (7.12).

(2) Fix an arbitrary $\epsilon \in (0, \gamma_L)$. By (7.4), there exists $N = N(\epsilon) \in \mathbb{Z}^+$ such that, for all $n > N$,

$$\frac{1}{n} \log \mathbf{X}_n \in (\gamma_L - \epsilon, \gamma_L + \epsilon) \ . \tag{7.15}$$

Let $R > \max\left\{\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_N, e^{N(\gamma_L+\epsilon)}\right\}$. Then, for all $n \le \left\lfloor \frac{1}{\gamma_L+\epsilon} \log R \right\rfloor$ so that $e^{n(\gamma_L+\epsilon)} < R$, we have

$$\mathbf{X}_n \le \max\left\{\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_N, e^{n(\gamma_L+\epsilon)}\right\} < R$$

and therefore $\tau_R(\mathbf{x}_0) \ge \left\lfloor \frac{1}{\gamma_L+\epsilon} \log R \right\rfloor$. On the other hand, for $n = \left\lceil \frac{1}{\gamma_L-\epsilon} \log R \right\rceil > N$ (where the last inequality holds since $R > e^{N(\gamma_L+\epsilon)}$), by (7.15) it holds that

$$\mathbf{X}_n > e^{n(\gamma_L-\epsilon)} \ge R \ ,$$

and therefore, $\tau_R(\mathbf{x}_0) \le \left\lceil \frac{1}{\gamma_L-\epsilon} \log R \right\rceil$. Summing up, we have

$$\left\lfloor \frac{1}{\gamma_L+\epsilon} \log R \right\rfloor \le \tau_R(\mathbf{x}_0) \le \left\lceil \frac{1}{\gamma_L-\epsilon} \log R \right\rceil \ ,$$

and thus we get

$$\frac{1}{\gamma_L+\epsilon} \le \liminf_{R\to\infty} \frac{\tau_R(\mathbf{x}_0)}{\log R} \le \limsup_{R\to\infty} \frac{\tau_R(\mathbf{x}_0)}{\log R} \le \frac{1}{\gamma_L+\epsilon} \ .$$

Letting $\epsilon \searrow 0$ completes the proof. $\square$

Now we are ready to prove the lower bound for multivariate case.

Proposition 7.5. *For all $\mathbf{x}_0 \in \mathbb{R}^d$, we almost surely have that*

$$\liminf_{R\to\infty} \frac{\tau_R(\mathbf{x}_0)}{\log R} \ge \frac{1}{\gamma_L} \ .$$

Proof. We fix $\mathbf{x}_0 \in \mathbb{R}^d$. Then, we fix a positive integer N_0. Then, by the submultiplicativeness of the matrix norm, we have

$$\frac{1}{N_0} \mathbb{E} \log \|\mathbf{A}_1 \cdots \mathbf{A}_{N_0}\| \ge \frac{1}{kN_0} \mathbb{E} \log \|\mathbf{A}_1 \cdots \mathbf{A}_{kN_0}\|$$

for all $k \ge 1$. Hence, by letting $k \to \infty$ and then recalling the definition of the Lyapunov exponent $\gamma_L > 0$, we get

$$\mathbb{E} \log \|\mathbf{A}_1 \cdots \mathbf{A}_{N_0}\| \ge N_0 \gamma_L \ .$$

Observe that, for all $n \in \mathbb{Z}^+$, we can write

$$\mathbf{X}_{n+N_0}(\mathbf{x}_0) = \mathbf{A}_{n+N_0} \cdots \mathbf{A}_{n+1} \mathbf{X}_n(\mathbf{x}_0) + (\mathbf{A}_{n+N_0} \cdots \mathbf{A}_{n+2} \mathbf{B}_{n+1} + \cdots + \mathbf{B}_{n+N_0}) \ .$$

and therefore, we have

$$|\mathbf{X}_{n+N_0}(\mathbf{x}_0)| \le \|\mathbf{A}_{n+N_0} \cdots \mathbf{A}_{n+1}\| \cdot |\mathbf{X}_n(\mathbf{x}_0)| + |\mathbf{A}_{n+N_0} \cdots \mathbf{A}_{n+2} \mathbf{B}_{n+1} + \cdots + \mathbf{B}_{n+N_0}|. \tag{7.16}$$

We fix $s \in [\![1,\, N_0]\!]$ and let, for $k \in \mathbb{Z}^+$,

$$a_k = \|\mathbf{A}_{kN_0+s} \cdots \mathbf{A}_{(k-1)N_0+s+1}\|$$
$$b_k = |\mathbf{A}_{kN_0+s} \cdots \mathbf{A}_{(k-1)N_0+s+2}\mathbf{B}_{(k-1)N_0+s+1} + \cdots + \mathbf{B}_{kN_0+s}| \ .$$

We next consider an univariate process $(Y_n^{(s)})_{n\in\mathbb{Z}_0^+}$ defined by $Y_0^{(s)} = |\mathbf{X}_s(\mathbf{x}_0)|$ and

$$Y_k^{(s)} = a_k Y_{k-1}^{(s)} + b_k \quad ; \ k \in \mathbb{Z}^+.$$

Then, by (7.16), for all $k \in \mathbb{Z}^+$, we have

$$|\mathbf{X}_{kN_0+s}(\mathbf{x}_0)| \le |Y_k^{(s)}| \ . \tag{7.17}$$

Define $\zeta_R^{(s)} = \inf\left\{k : |Y_k^{(s)}| > R\right\}$ so that by Lemma 7.4, we almost surely have

$$\lim_{R\to\infty} \frac{\zeta_R^{(s)}}{\log R} = \frac{1}{\mathbb{E}\log a_1} \ . \tag{7.18}$$

Note here that the condition $\mathbb{P}(b_k = 0) < 1$ of Lemma 7.4 is a consequence of Lemma 5.4-(2) with $\mathbf{x} = \mathbf{0}$, as b_k has the same distribution of $\mathbf{X}_{N_0}(\mathbf{0})$.

We next claim that

$$\tau_R(\mathbf{x}_0) \ge \min_{1\le s\le N_0} \left\{N_0\zeta_R^{(s)} + s\right\} \ . \tag{7.19}$$

This claim completes the proof since this and (7.18) together implies that

$$\liminf_{R\to\infty} \frac{\tau_R(\mathbf{x}_0)}{\log R} \ge \frac{N_0}{\mathbb{E}\log\|\mathbf{A}_{N_0}\cdots\mathbf{A}_1\|} \ ,$$

and therefore by letting $N_0 \to \infty$ we can get the desired bound.

Now we turn to the proof of (7.19). We can find unique integers $k_0 \ge 0$ and $1 \le s_0 \le N_0$ such that

$$\tau_R(\mathbf{x}_0) = k_0N_0 + s_0 \ .$$

Then, the comparison (7.17) implies that $\zeta_R^{(s_0)} \le k_0$ and therefore we get

$$\min_{1\le s\le N_0} \left\{N_0\zeta_R^{(s_0)} + s\right\} \le N_0\zeta_R^{(s_0)} + s_0 \le N_0k_0 + s_0 = \tau_R(\mathbf{x}_0)$$

as desired. □

7.3. **Upper bound.** We now turn to the upper bound. The purpose is to prove the following bound.

Proposition 7.6. *For all $\mathbf{x}_0 \in \mathbb{R}^d$, there exist constants $\kappa_1,\, \kappa_2 > 0$ which depend only on the distribution of $(\mathbf{A},\, \mathbf{B})$ such that*

$$\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right] \le \kappa_1(1 + \log R) - \kappa_2 \log^+ |\mathbf{x}_0| \ .$$

We note that the constants appeared in the previous proposition can be computed explicitly in the course of proof. The proof of the upper bound heavily relies on the construction of a submartingale. In this construction, we heavily use the function

$$\log^+ x = \log(x \vee 1) = (\log x) \vee 0 \quad ; \ x \geq 0 \ ,$$

where we regard $\log^+ 0 = 0$. For this function, we have useful elementary inequalities.

Lemma 7.7. *For all x, $y \geq 0$, we have*

$$\log^+(x+y) \leq \log^+ x + \log^+ y + \log 2 \quad \textit{and}$$
$$\log^+ xy \leq \log^+ x + \log^+ y \ .$$

Proof. We look at the first inequality. Thanks to the symmetry, it suffices to consider following three cases separately.

- $0 \leq x,\ y \leq 1$: the inequality is immediate from $x + y \leq 2$ and that $\log^+(\cdot)$ is an increasing function.
- $0 \leq x \leq 1 \leq y$: in this case, we have

$$\log^+(x+y) = \log(x+y) \leq \log(2y) = \log y + \log 2 = \log^+ x + \log^+ y + \log 2$$

- $x,\ y \geq 1$: in this case, we have $\log^+ = \log$ and hence it suffices to check that $2xy \geq x + y$. This is immediate since $xy \geq x$ and $xy \geq y$.

The proof of the second one is immediate from

$$\log^+ xy = (\log xy) \vee 0 = (\log x + \log y) \vee 0 \leq (\log^+ x + \log^+ y) \vee 0 = \log^+ x + \log^+ y \ .$$

□

We from now on fix $\mathbf{x}_0$ and construct a sub-martingales to complete the proof of Proposition 7.6. By lemma 7.1, we are able to take N_0 such that for all $n \geq N_0$

$$\frac{1}{n}\mathbb{E} \log \|\Pi_n^{-1}\|^{-1} \geq \frac{2}{3}\gamma_L \ . \tag{7.20}$$

The first step is to construct a sequence of stopping times $(\sigma_j)_{j=0}^{\infty}$ along which we construct the submartingale. The following lemma is used in the construction. Recall that $(\mathcal{F}_n)_{n\in\mathbb{Z}_0^+}$ is the filtration associated with the process $(\mathbf{X}_n(\cdot))_{n\in\mathbb{Z}_0^+}$.

Lemma 7.8. *Let τ be a $(\mathcal{F}_n)_{n\in\mathbb{Z}_0^+}$-stopping time and $f : \mathbb{R}^d \to \mathbb{Z}^+$ be a Lebesgue measurable function. Then, for all $\mathbf{x}_0 \in \mathbb{R}^d$,*

$$\sigma = \tau + f(\mathbf{X}_\tau(\mathbf{x}_0))$$

is also a $(\mathcal{F}_n)_{n\in\mathbb{Z}_0^+}$-stopping time.

Proof. Since

$$\{\sigma = n\} = \bigcup_{k,\, \ell\in\mathbb{Z}^+ : k+\ell=n} (\{\tau = k\} \cap \{f(\mathbf{X}_k(\mathbf{x}_0)) = \ell\})$$

and since both $\{\tau = k\}$ and $\{f(\mathbf{X}_k(\mathbf{x}_0)) = \ell\}$ are $\mathcal{F}_k$-measurable set, $\{\sigma = n\} \in \mathcal{F}_n$. □

We let $\sigma_0 = 0$ and suppose that the sequence $\sigma_0,\ \dots,\ \sigma_j$ has been constructed. Then, for all $n \in \mathbb{Z}^+$, by Lemma 7.7, we have

$$\begin{aligned}
&\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_j+n+1}(\mathbf{x}_0)| + \log^+|\mathbf{X}_{\sigma_j+n}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] \\
&= \mathbb{E}\left[\log^+|\mathbf{X}'_n(\mathbf{X}_{\sigma_j+1}(\mathbf{x}_0))| + \log^+|\mathbf{X}'_n(\mathbf{X}_{\sigma_j}(\mathbf{x}_0))| \mid \mathcal{F}_{\sigma_j}\right] \\
&= \mathbb{E}\left[\log^+|\mathbf{X}'_n(\mathbf{X}_{\sigma_j+1}(\mathbf{x}_0)) - \mathbf{X}'_n(\mathbf{X}_{\sigma_j}(\mathbf{x}_0))| \mid \mathcal{F}_{\sigma_j}\right] - \log 2 \\
&\geq \mathbb{E}\log\|\Pi_n^{-1}\|^{-1} + \mathbb{E}\left[\log|\mathbf{X}_{\sigma_j+1}(\mathbf{x}_0) - \mathbf{X}_{\sigma_j}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] - \log 2
\end{aligned}$$

where the last line follows from (5.2) and the fact that $\log^+ \geq \log$. Here, we note that $(\mathbf{X}'_n(\cdot))_{n\in\mathbb{Z}^+}$ denotes an process (2.1) which has the same law with $(\mathbf{X}_n(\cdot))_{n\in\mathbb{Z}^+}$ but uses the random matrices/vectors $(\mathbf{A}_n,\, \mathbf{B}_n)_{n\in\mathbb{Z}^+}$ independent to the ones associated with $(\mathbf{X}_n(\cdot))_{n\in\mathbb{Z}^+}$. Then, by (7.20), (3.9) of Assumption 6, and the strong Markov property, there exists a constant $C_1 > 0$ (which will be fixed throughout the current subsection) such that, for all $n \geq N_0$,

$$\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_j+n+1}(\mathbf{x}_0)| + \log^+|\mathbf{X}_{\sigma_j+n}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] \geq \frac{2n}{3}\gamma_L - C_1\ . \tag{7.21}$$

Therefore, we can define a random time σ_{j+1} which is finite almost surely by

$$\sigma_{j+1} = \sigma_j + \inf\left\{n \geq 1 : \mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_j+n}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] \geq \frac{n}{4}\gamma_L + \log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)|\right\}\ . \tag{7.22}$$

Then, since we have assumed that σ_j is a stopping time, the random time σ_{j+1} is again a stopping time by Lemma 7.8. Note that this recursive construction guarantees that $\sigma_j < \infty$ almost surely for all $j \in \mathbb{Z}^+$.

Summing up, we constructed an increasing sequence of finite stopping times $(\sigma_j)_{j=0}^{\infty}$ with $\sigma_0 = 0$ that satisfies, for all $j \geq 0$,

$$\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_{j+1}}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] \geq \frac{\gamma_L}{4}(\sigma_{j+1} - \sigma_j) + \log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)|\ . \tag{7.23}$$

Lemma 7.9. *With the notation above, for all $k \in \mathbb{Z}^+$, we have $\mathbb{E}[\sigma_k] < \infty$ and $\mathbb{E}[\log^+|\mathbf{X}_{\sigma_k}(\mathbf{x}_0)|] < \infty$.*

Proof. We proceed with induction. First, the statement is clear for $k = 0$ as $\sigma_k = 0$.

Next we assume that the statement of the lemma holds for $k = j$ and look at the case $k = j + 1$. We first observe from (7.21) that, with

$$n_j := \left\lceil \frac{6}{\gamma_L}\left(2\log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)| + C_1\right)\right\rceil\ ,$$

we have

$$\begin{aligned}
\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_j+n_j+1}(\mathbf{x}_0)| + \log^+|\mathbf{X}_{\sigma_j+n_{j0}}(\mathbf{x}_0)| \mid \mathcal{F}_{\sigma_j}\right] &\geq \frac{2n_0}{3}\gamma_L - C_1\ . \\
&\geq \frac{n_0}{2}\gamma_L + 2\log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)|
\end{aligned}$$

and therefore we can conclude from (7.22) that

$$\sigma_{j+1} - \sigma_j \leq n_j + 1 \leq \frac{6}{\gamma_L}\left(2\log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)| + C_1\right) + 2\ . \tag{7.24}$$

Hence, by the induction hypothesis, we get

$$\mathbb{E}\sigma_{j+1} \leq \mathbb{E}\sigma_j + \frac{12}{\gamma_L}\mathbb{E}\log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)| + \frac{6C_1}{\gamma_L} + 2 < \infty \ .$$

Next look at $\mathbb{E}\log^+|\mathbf{X}_{\sigma_{j+1}}(\mathbf{x}_0)|$. By Lemma 7.7, we have

$$\begin{aligned}\mathbb{E}\left[\log^+|\mathbf{X}_{n+1}(\mathbf{x}_0)| \mid \mathcal{F}_n\right] &\leq \mathbb{E}\left[\log^+|\mathbf{A}_{n+1}\mathbf{X}_n(\mathbf{x}_0)| + \log^+|\mathbf{B}_{n+1}| \mid \mathcal{F}_n\right] + \log 2 \\ &\leq \log^+|\mathbf{X}_n(\mathbf{x}_0)| + \mathbb{E}\log^+\|\mathbf{A}\| + \mathbb{E}\log^+|\mathbf{B}| + \log 2 \ .\end{aligned}$$

Hence, by letting $C_2 := \mathbb{E}\log^+\|\mathbf{A}\| + \mathbb{E}\log^+|\mathbf{B}| + \log 2 > 0$ and defining

$$Z_n = \log^+|\mathbf{X}_n(\mathbf{x}_0)| - C_2 n \quad ; \, n \in \mathbb{Z}^+ \ ,$$

we can notice that $(Z_n)_{n\in\mathbb{Z}_0^+}$ is a $(\mathcal{F}_n)_{n\in\mathbb{Z}_0^+}$-supermartingale. Thus, by the optimal stopping theorem, we have

$$\mathbb{E}\log^+|\mathbf{X}_{n\wedge\sigma_{j+1}}(\mathbf{x}_0)| \leq C_2\mathbb{E}\left[n\wedge\sigma_{j+1}\right] + \log^+|\mathbf{x}_0| \tag{7.25}$$

for all $n \in \mathbb{Z}^+$. Since $\sigma_j < \infty$ almost surely, by Fatou's lemma (at the left-hand side) and dominated convergence theorem (at the right-hand side, as we already verified that $\mathbb{E}(\sigma_{j+1}) < \infty$), we get

$$\mathbb{E}\log^+|\mathbf{X}_{\sigma_{j+1}}(\mathbf{x}_0)| \leq C_2\mathbb{E}\sigma_{j+1} + \log^+|\mathbf{x}_0| < \infty \ . \tag{7.26}$$

This completes the induction step. □

Define a new filtration $(\mathcal{G}_j)_{j=0}^\infty$ by $\mathcal{G}_j = \mathcal{F}_{\sigma_j}$ and then we define, for $j \geq 0$,

$$M_j := \log^+|\mathbf{X}_{\sigma_j}(\mathbf{x}_0)| - \frac{\gamma_L}{4}\sigma_j \ . \tag{7.27}$$

Note that by Lemma 7.9, we have $\mathbb{E}|M_j| < \infty$ for all $j \in \mathbb{Z}^+$. Then, by (7.23), the sequence $(M_j)_{j=0}^\infty$ is a $(\mathcal{G}_j)_{j=0}^\infty$-submartingale. Next we define

$$\theta = \theta_R := \inf\left\{j \geq 0 : \max\left\{|\mathbf{X}_0(\mathbf{x}_0)|,\, |\mathbf{X}_1(\mathbf{x}_0)|,\, \dots,\, |\mathbf{X}_{\sigma_j}(\mathbf{x}_0)|\right\} \geq R\right\} \ .$$

Here, θ is a $(\mathcal{G}_j)_{j=0}^\infty$-stopping time since

$$\{\theta \leq n\} = \bigcup_{k=1}^{\sigma_n} \{\, |\mathbf{X}_k(\mathbf{x}_0)| \geq R \,\} \ .$$

Note that $\theta < \infty$ almost surely by Proposition 7.2.

Lemma 7.10. *With the notation above, there exists a constant $C_3 > 0$ such that*

$$\mathbb{E}\log^+|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)| \leq C_3\left(1 + \log R\right) \ .$$

Proof. For $0 \leq n \leq \sigma_\theta$, by the same argument with the one from (7.25) to (7.26), we get

$$\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)| \mid \mathcal{F}_n\right] \leq C_2(\sigma_\theta - n) + \log^+|\mathbf{X}_n(\mathbf{x}_0)|$$

Therefore, we can deduce

$$\begin{aligned}
\mathbb{E}\log^+|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)| &= \sum_{n=0}^{\infty}\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)|\cdot\mathbf{1}\{\tau_R(\mathbf{x}_0)=n\}\right]\\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[\mathbb{E}\left[\log^+|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)| \mid \mathcal{F}_n\right]\cdot\mathbf{1}\{\tau_R(\mathbf{x}_0)=n\}\right]\\
&\le \sum_{n=0}^{\infty}\mathbb{E}\left[\left(C_2(\sigma_\theta-n)+\log^+|\mathbf{X}_n(\mathbf{x}_0)|\right)\cdot\mathbf{1}\{\tau_R(\mathbf{x}_0)=n\}\right]\\
&= C_2\mathbb{E}\left[\sigma_\theta-\tau_R(\mathbf{x}_0)\right]+\mathbb{E}\log^+|\mathbf{X}_{\tau_R}(\mathbf{x}_0)|\ .
\end{aligned}\tag{7.28}$$

Since $\sigma_{\theta-1}<\tau_R(\mathbf{x}_0)\le\sigma_\theta$ by definition, we obtain from (7.24) and the fact $|\mathbf{X}_{\sigma_{\theta-1}}(\mathbf{x}_0)|\le R$ that

$$\sigma_\theta-\sigma_{\theta-1}\le\frac{6}{\gamma_L}\left(2\log^+|\mathbf{X}_{\sigma_{\theta-1}}(\mathbf{x}_0)|+C_1\right)+2\le\frac{6}{\gamma_L}\left(2R+C_1\right)+2\ ,$$

so we get

$$\mathbb{E}\left[\sigma_\theta-\tau_R(\mathbf{x}_0)\right]\le\mathbb{E}\left[\sigma_\theta-\sigma_{\theta-1}\right]\le\frac{6}{\gamma_L}\left(2R+C_1\right)+2\ .$$

Injecting this and Lemma 7.3 to 7.28 completes the proof. □

Now we complete the proof of Proposition 7.6.

Proof of Proposition 7.6. Since $(M_j)_{j=0}^\infty$ is a $(\mathcal{G}_j)_{j=0}^\infty$-submartingale, and θ is a $(\mathcal{G}_j)_{j=0}^\infty$ stopping time, by the optional stopping theorem, we get, for all $j\in\mathbb{Z}^+$,

$$\mathbb{E}M_{\theta\wedge j}\ge\mathbb{E}M_0=\log^+|\mathbf{x}|\ .\tag{7.29}$$

On the other hand, by definition (7.27) of M_j, the fact that $|\mathbf{X}_{\sigma_{\theta\wedge j}}(\mathbf{x}_0)|\le|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)|\wedge R$, and Lemma 7.10, we obtain, for all $j\in\mathbb{Z}^+$,

$$\begin{aligned}
\mathbb{E}M_{\theta\wedge j}&\le\mathbb{E}\log^+(|\mathbf{X}_{\sigma_\theta}(\mathbf{x}_0)|\wedge R)-\frac{\gamma_L}{4}\mathbb{E}\left[\sigma_{\theta\wedge j}\right]\\
&\le C(1+\log R)-\frac{\gamma_L}{4}\mathbb{E}\left[\sigma_{\theta\wedge j}\right]\ .
\end{aligned}$$

Combining this bound with (7.29), and then letting $j\to\infty$, we get

$$\mathbb{E}\left[\sigma_\theta\right]\le C(1+\log R)-C'\log^+|\mathbf{x}|$$

Since $\tau_R(\mathbf{x}_0)\le\sigma_\theta$, the proof is completed. □

7.4. **Proof of Theorem 3.7.** We are now finally ready to prove Theorem 3.7.

Proof of Theorem 3.7. By Proposition 7.5 and Fatou's theorem, we get

$$\liminf_{R\to\infty}\frac{\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right]}{\log R}\ge\mathbb{E}\left[\liminf_{R\to\infty}\frac{\tau_R(\mathbf{x}_0)}{\log R}\right]=\frac{1}{\gamma_L}\ .\tag{7.30}$$

On the other hand, by Proposition 7.6, we have

$$\limsup_{R\to\infty}\frac{\mathbb{E}\left[\tau_R(\mathbf{x}_0)\right]}{\log R}\le\kappa_1\tag{7.31}$$

where κ_1 is the constant appeared in the statement of Proposition 7.6. Combining (7.30) and (7.31) completes the proof. $\square$

Appendix A. Verifications of Assumptions for Example Models

In this appendix, we confirm that the time series model (namely, the ARCH and the GARCH models) introduced in Section 4.2 satisfy the assumptions of Section 3 so that we can apply Theorems 3.4 and 3.7 according to the sign of the Lyapunov exponent γ_L.

A.1. **Model review.** We recall from Section 4.2 that ARCH(p) and GARCH(1, p) models can be viewed as the process $\mathbf{X}_n$ defined in Definition 2.1 with

$$\mathbf{A}_n = \begin{pmatrix} \alpha_1 W_n^2 + \beta_1 & \alpha_2 W_n^2 + \beta_2 & \cdots & \alpha_{p-1} W_n^2 + \beta_{p-1} & \alpha_p W_n^2 + \beta_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \mathbf{B}_n = \begin{pmatrix} \alpha_0 W_n^2 + \beta_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{A.1}$$

where $(W_n)_{n\geq 0}$ are i.i.d. normally distributed random variables (with mean m and variance σ^2) and α_i, $\beta_i \geq 0$ for all $0 \leq i \leq p$. In particular, ARCH and GARCH models satisfy the following assumptions.

- ARCH(p): $\beta_0 = \beta_1 = \cdots = \beta_p = 0$ and α_0, $\alpha_p > 0$.
- GARCH(1, p): $\alpha_2 = \cdots = \alpha_p = 0$, $\alpha_0 = 0$, and α_1, β_0, $\beta_p > 0$.

In addition, we assume $\alpha_1 \neq 0$ for the ARCH model for technical reason, as without this condition the Assumption 5 is invalid for e.g., ARCH(2) model. We note that this assumption is satisfied in the usual real-world time series model. We focus this section only on the ARCH model, as the proof for the GARCH model is very similar to that for the ARCH model as they share the expression (A.1).

Notation A.1. We denote in this appendix by W the normal random variable with mean m and variance σ^2 (so that W has the same distribution with W_n, $n \geq 0$, appeared above) and denote by $\mathbf{A}$ and $\mathbf{B}$ the matrix and vector obtained from $\mathbf{A}_n$ and $\mathbf{B}_n$ in (A.1) by replacing W_n with W. Since the model depends only on the square of W_n, we can assume without generality that $m \geq 0$.

A.2. **Lyapunov exponent.** In the lemma below, we prove that the Lyapunov exponent γ_L can be either positive or negative, so that both contractive and explosive regimes are required to be analyzed.

Lemma A.2. *For the ARCH(p) model, we have that*

$$\lim_{\alpha_1 \to \infty} \gamma_L = \infty \quad \textit{and} \quad \lim_{(\alpha_1, \ldots, \alpha_p) \to (0, \ldots, 0)} \gamma_L = -\infty\ .$$

Proof. Since every elements of $\mathbf{A}$ are non-negative, the same holds for $\mathbf{A}_1 \cdots \mathbf{A}_n$ and hence we have

$$\|\mathbf{A}_1 \cdots \mathbf{A}_n\| \geq [\mathbf{A}_1 \cdots \mathbf{A}_n]_{1,1} \geq [\mathbf{A}_1]_{1,1} \cdots [\mathbf{A}_n]_{1,1} \geq \alpha_1^n \prod_{t=1}^{n} W_t^2 , \tag{A.2}$$

where $[\mathbf{U}]_{i,j}$denotes the (i, j)-th component of the matrix $\mathbf{U}$. Therefore, we immediately have

$$\gamma_L = \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \log \|\mathbf{A}_1 \cdots \mathbf{A}_n\| \geq \log \alpha_1 + \log \mathbb{E} W^2 .$$

This proves the first assertion of the Lemma.

For the second assertion, we write $\mathbf{A}_n = W_n^2 \mathbf{D} + \mathbf{E}$ where

$$\mathbf{D} := \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \ \mathbf{E} := \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} . \tag{A.3}$$

Then, since $\mathbf{E}^p = \mathbf{O}$, we can apply the triangle inequality and the submultiplicativity of the matrix norm at the right-hand side (after expanding) of

$$\mathbf{A}_1 \cdots \mathbf{A}_p = (W_1^2 \mathbf{D} + \mathbf{E}) \cdots (W_p^2 \mathbf{D} + \mathbf{E})$$

to deduce (since $\|\mathbf{E}\| = 1$)

$$\|\mathbf{A}_1 \cdots \mathbf{A}_p\| \leq (W_1^2 \|\mathbf{D}\| + 1) \cdots (W_p^2 \|\mathbf{D}\| + 1) - 1 .$$

Since $\|\mathbf{D}\| \leq \sum_{i=1}^{p} \alpha_i$, we have

$$\limsup_{(\alpha_1, \ldots, \alpha_p) \to \mathbf{0}} \|\mathbf{A}_1 \cdots \mathbf{A}_p\| = 0 .$$

Then, we can apply Jensen's inequality and dominated convergence theorem to get

$$\limsup_{(\alpha_1, \ldots, \alpha_p) \to \mathbf{0}} \mathbb{E} \log \|\mathbf{A}_1 \cdots \mathbf{A}_p\| \leq \limsup_{(\alpha_1, \ldots, \alpha_p) \to \mathbf{0}} \log \mathbb{E} \|\mathbf{A}_1 \cdots \mathbf{A}_p\| = -\infty .$$

This proves the second assertion since we have the following upper bound on γ_L thanks to the submultiplicativity of the matrix norm and independence of the matrices $\mathbf{A}_n$:

$$\gamma_L = \lim_{m\to\infty} \frac{1}{mp} \log \mathbb{E} \|\mathbf{A}_1 \cdots \mathbf{A}_{mp}\| \leq \frac{1}{p} \log \mathbb{E} \|\mathbf{A}_1 \cdots \mathbf{A}_p\| .$$

□

A.3. **Assumptions for contractive regime.** We first assume that $\gamma_L < 0$ so that we are in the contractive regime. In this regime, we shall verify that Assumptions 1, 2, 3, 4 and 5 are valid for the ARCH model.

Verification of Assumption 1. Since we have assumed $\gamma_L < 0$ in this subsection, it suffices to verify (2), (3), and (4) of Assumption 1. First, (2) and (3) are direct from the fact that the entries of $\mathbf{A}$ and $\mathbf{B}$ are normal and therefore have exponentially small tail.

For (4), suppose that $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ satisfies $\mathbf{Ax} + \mathbf{B} = \mathbf{x}$ almost surely. Then, by comparing the 2nd to the pth coordinates of each side, we get $x_1 = x_2 = \cdots = x_p$. Furtheremore, by comparing the first coordinate, we get

$$\sum_{i=1}^{p} (\alpha_i W^2 + \beta_i) x_1 + \alpha_0 W^2 + \beta_0 = x_1$$

almost surely. Thus, we have

$$\left(\sum_{i=1}^{p} \alpha_i \right) x_1 + \alpha_0 = 0 \quad \text{and} \quad \left(\sum_{i=1}^{p} \beta_i \right) x_1 + \beta_0 = x_1 \ .$$

For the ARCH model, this implies $x_1 = 0$ as we have $\beta_0 = \beta_1 = \cdots = \beta_p = 0$. Summing up, we must have $\mathbf{x} = \mathbf{0}$ and thus, from $\mathbf{Ax} + \mathbf{B} = \mathbf{x}$ we get $\mathbf{B} = \mathbf{0}$ almost surely. This is a contradiction and therefore we verified all the requirements of Assumption 1.

Verification of Assumption 2. It suffices to check $\lim_{s \to \infty} h_{\mathbf{A}}(s) = \infty$. By (A.2), we have

$$(\mathbb{E}\|\mathbf{A}_1 \cdots \mathbf{A}_n\|^s)^{1/n} \ge \alpha_1^s \mathbb{E} W^{2s} \ ,$$

and therefore $h_{\mathbf{A}}(s) \ge \alpha_1^s \mathbb{E} W^{2s}$. On the other hand, since $W \sim \mathcal{N}(m, \sigma^2)$, it follows from the well-known identity that

$$\mathbb{E} W^{2s} \ge \mathbb{E} \left| W - m \right|^{2s} = \frac{(2s)!}{2^s s!} \sigma^{2s} \quad ; \ s \in \mathbb{Z}^+ \ ,$$

and hence we can conclude that

$$h_{\mathbf{A}}(s) \ge \alpha_1^s \mathbb{E} W^{2s} \ge \frac{(2s)!}{s!} \left(\frac{\alpha \sigma_1^2}{2} \right)^s \ge \left(\frac{\alpha \sigma_1^2}{2} s \right)^s \ .$$

This completes the proof of $\lim_{s \to \infty} h_{\mathbf{A}}(s) = \infty$ and hence the confirmation of Assumption 2.

Verification of Assumption 3. For Assumption 3, the unboundedness of $\mathbf{B}$ (thanks to the unboundedness of the normal random variable) directly implies that of the support of ν_∞.

Verification of Assumption 4. We start from the following elementary fact.

Lemma A.3. *For all $s > 0$, there exists $u_0 = u_0(s) > 0$ such that, for all u_1, u_2 such that $u_0 \le u_1 \le u_2$, we have*

$$\frac{\mathbb{P}(W > u_2)}{\mathbb{P}(W > u_1)} \le 2 \left(\frac{u_1}{u_2} \right)^s \ .$$

Proof. First, we assume that $m = 0$, $\sigma = 1$ so that W is a standard normal random variable. Then, using basic inequality

$$\frac{a}{a^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \le \mathbb{P}(W > a) \le \frac{1}{a} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \quad ; \ a > 0 \ , \tag{A.4}$$

we get for any $u_1, u_2 > 1$,

$$\frac{\mathbb{P}(W > u_2)}{\mathbb{P}(W > u_1)} \le \frac{\frac{1}{u_2} e^{-u_2^2/2}}{\frac{u_1}{u_1^2+1} e^{-u_1^2/2}} \le \frac{\frac{1}{u_2} e^{-u_2^2/2}}{\frac{1}{2u_1} e^{-u_1^2/2}} = \left(\frac{2u_1}{u_2} \right) e^{-\frac{u_2^2 - u_1^2}{2}} \ . \tag{A.5}$$

Observing that if $u_1 \geq \sqrt{s}$, using elementary inequality $x \leq e^{x-1}$, we have that

$$\left(\frac{u_2}{u_1}\right)^{s-1} \leq \left(\frac{u_2}{u_1}\right)^{s} \leq e^{s \cdot \frac{u_2-u_1}{u_1}} \leq e^{u_1(u_2-u_1)} \leq e^{\frac{(u_2+u_1)(u_2-u_1)}{2}} ,$$

hence combining with (A.5) yields the desired conclusion with $u_0 = \max\{1, \sqrt{s}\}$.

For general m, σ, we can deduce from the previous conclusion that, for all u_1, u_2 such that

$$\max\left\{m+\sigma,\, m\sqrt{s}+\sigma\right\} \leq u_1 \leq u_2 ,$$

we have that

$$\frac{\mathbb{P}(W > u_2)}{\mathbb{P}(W > u_1)} \leq 2\left(\frac{(u_1-m)/\sigma}{(u_2-m)/\sigma}\right)^{s} \leq 2\left(\frac{u_1-m}{u_2-m}\right)^{s} \leq 2\left(\frac{u_1}{u_2}\right)^{s} , \tag{A.6}$$

where the last inequality holds as $m < u_1 \leq u_2$. □

Now we are ready to check Assumption 4. Let $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ with $|\mathbf{x}| \leq R$ and set $x_0 := 1$. Then, we can write

$$|\mathbf{A}\mathbf{x} + \mathbf{B}|^2 = \left(\left(\sum_{k=0}^{p} \alpha_k x_k\right) W^2 + \left(\sum_{k=0}^{p} \beta_k x_k\right)\right)^2 + \sum_{k=1}^{p-1} x_k^2 . \tag{A.7}$$

Therefore, Assumption 4 is immediate if $\sum_{k=0}^{p} \alpha_k x_k = 0$. Suppose from now on that $\sum_{k=0}^{p} \alpha_k x_k \neq 0$.

Then, since $|\mathbf{x}| \leq R$, we get from (A.7) that, for any z, $R > 0$,

$$\begin{aligned}\mathbb{P}(|\mathbf{A}\mathbf{x} + \mathbf{B}| > R) &\geq \mathbb{P}\left(\left|\left(\sum_{k=0}^{p} \alpha_k x_k\right) W^2 + \left(\sum_{k=0}^{p} \beta_k x_k\right)\right| > R\right) \\ &\geq \mathbb{P}\left(\left|\sum_{k=0}^{p} \alpha_k x_k\right| W^2 > \left(1 + \sum_{k=0}^{p} |\beta_k|\right) R\right) .\end{aligned}$$

On the other hand, again by (A.7) again, we get

$$\begin{aligned}\mathbb{P}(|\mathbf{A}\mathbf{x} + \mathbf{B}| > zR) &\leq \mathbb{P}\left(\left(\left(\sum_{k=0}^{p} \alpha_k x_k\right) W^2 + \left(\sum_{k=0}^{p} \beta_k x_k\right)\right)^2 + (p-1)R^2 > z^2R^2\right) \\ &\leq \mathbb{P}\left(2\left(\sum_{k=0}^{p} \alpha_k x_k\right)^2 W^4 > (z^2 - C_1)R^2\right)\end{aligned}$$

where $C_1 := p - 1 + 2\left(\sum_{k=0}^{p} |\beta_k|\right)^2$. With these estimates and Lemma A.3 with $s := 2(\alpha+1)$ where α is from Assumption 2, we get, for any $R > 0$ and $z > \left\{2\left(1 + \sum_{k=0}^{p} |\beta_k|\right)^2 + C_1\right\}^{1/2}$,

$$\frac{\mathbb{P}(|\mathbf{A}\mathbf{x} + \mathbf{B}| > zR)}{\mathbb{P}(|\mathbf{A}\mathbf{x} + \mathbf{B}| > R)} \leq \frac{C_2}{(z^2 - C_1)^{(\alpha+1)/2}}$$

for some constant $C_2 > 0$ independent of $\mathbf{x}$, z, and R. Assumption 4 immediately follows from this bound.

Verification of Assumption 5. We shall now prove that, for any $\mathbf{x}_0,\, \mathbf{y}_0 \in \mathbb{R}^p \setminus \{\mathbf{0}\}$,

$$\mathbb{P}(\mathbf{x}_0 \cdot \mathbf{A}_1 \cdots \mathbf{A}_{2p-1}\mathbf{y}_0 = 0) < 1 \ , \tag{A.8}$$

so that Assumption 5 holds. Suppose on the contrary that we have

$$\mathbb{P}(\mathbf{x}_0 \cdot \mathbf{A}_1 \cdots \mathbf{A}_{2p-1}\mathbf{y}_0 = 0) = 1 \ , \tag{A.9}$$

for some $\mathbf{x}_0,\, \mathbf{y}_0 \in \mathbb{R}^p \setminus \{\mathbf{0}\}$.

Recall the matrices $\mathbf{D}$ and $\mathbf{E}$ defined in (A.3) so that we can write $\mathbf{A}_n = W_n^2\mathbf{D}+\mathbf{E}$. Thanks to this expression, if random vectors $\mathbf{X}$, $\mathbf{Y}$ independent of $\mathbf{A}_i$ satisfy $\mathbf{X} \cdot \mathbf{A}_i\mathbf{Y} = 0$ almost surely, then we can deduce that $\mathbf{X}\cdot\mathbf{D}\mathbf{Y} = \mathbf{X}\cdot\mathbf{E}\mathbf{Y} = 0$ almost surely as well. This observation enable us to deduce from (A.9) that

$$\mathbf{x}_0 \cdot \mathbf{M}_1 \cdots \mathbf{M}_{2p-1}\mathbf{y}_0 = 0, \quad \forall \mathbf{M}_i \in \{\mathbf{D},\, \mathbf{E}\}\, , \quad \forall i \in [\![1,\, 2p-1]\!] \ . \tag{A.10}$$

Since $\mathbf{x}_0$, $\mathbf{y}_0$ are nonzero vectors, there exist indices $r,\, s \in [\![1,\, p]\!]$ such that

$$\mathbf{e}_r \cdot \mathbf{x}_0 \neq 0,\;\; \mathbf{e}_s \cdot \mathbf{y}_0 \neq 0 \ , \tag{A.11}$$

where $\{\mathbf{e}_i : i \in [\![1,\, p]\!]\}$ is the standard orthonormal basis of $\mathbb{R}^p$. Then, observing that the first coordinate of

$$\mathbf{x}_0 \cdot \mathbf{E}^{r-1}\mathbf{D}^{2p-r-s+1}\mathbf{E}^{s-1}\mathbf{y}_0$$

equals to $\alpha_1^{2p-r-s}(\mathbf{e}_r\cdot\mathbf{x}_0)(\mathbf{e}_s\cdot\mathbf{y}_0)$ which is nonzero by our assumpition that $\alpha_1 \neq 0$ and (A.11). This contradicts to (A.10) and we are done.

A.4. **Assumptions for explosive regime.** In this section, we assume that $\gamma_L > 0$ and verify Assumption 6. Thanks to (4) of Remark 3.8, we only need to verify (2) of the assumption. To prove (2), we need the following elementary lemma.

Lemma A.4. *For a normal random variable W, for all $\delta > 0$, we have*

$$\inf_{0<\epsilon\le 1,\, y>\delta} \mathbb{E}\log|\epsilon W^2 - y| > -\infty \quad \text{and} \quad \inf_{y\in\mathbb{R}} \mathbb{E}\log|W^2 - y| > -\infty \ .$$

Proof. Recall that $W \sim N\left(m,\, \sigma^2\right)$. We shall assume for simplicity that $m = 0$ and $\sigma = 1$, since the proof for the general case is essentially same (with a slightly more involved computation). Also, it suffices to show only for $\delta < 1$ and $y > 2\delta$ for the first inequality since the infimum is taken on the set $\{0 < \epsilon \le 1,\, y > \delta\}$ which is decreasing respect to δ.

For the first bound, fix $0 < \epsilon \le 1$, $y > 2\delta$. Then, since $y > 2\delta$, we have

$$\begin{aligned}\mathbb{E}\log|\epsilon W^2 - y| &= \mathbb{E}\log|\epsilon W^2 - y|\mathbf{1}\left\{|\epsilon W^2 - y| \ge \delta\right\} + \mathbb{E}\log|\epsilon W^2 - y|\mathbf{1}\left\{|\epsilon W^2 - y| < \delta\right\} \\ &\ge \log\delta + \mathbb{E}\left(\log|\epsilon W^2 - y|\mathbf{1}\left\{y-\delta < \epsilon W^2 < y+\delta\right\}\right) \ .\end{aligned}$$

The last expectation can be bounded from below by

$$\begin{aligned}
&\mathbb{E}\log|\epsilon W^2-y|\mathbf{1}\left\{y-\delta<\epsilon W^2<y+\delta\right\}\\
&=2\int_{\sqrt{(y-\delta)/\epsilon}}^{\sqrt{(y+\delta)/\epsilon}}\frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}\log|\epsilon w^2-y|dw\\
&\geq\sqrt{\frac{2}{\pi}}\exp\left(-\frac{y-\delta}{2\epsilon}\right)\int_{\sqrt{(y-\delta)/\epsilon}}^{\sqrt{(y+\delta)/\epsilon}}\log|\epsilon w^2-y|dw\\
&=\sqrt{\frac{2}{\pi}}\exp\left(-\frac{y-\delta}{2\epsilon}\right)\int_{\sqrt{(y-\delta)/y}}^{\sqrt{(y+\delta)/y}}\sqrt{\frac{y}{\epsilon}}\left(\log y+\log|w^2-1|\right)dw\\
&\geq\sqrt{\frac{2}{\pi}}\exp\left(-\frac{y-\delta}{2\epsilon}\right)\left(\frac{\sqrt{y+\delta}-\sqrt{y-\delta}}{\sqrt{\epsilon}}\log y+\int_{\sqrt{(y-\delta)/y}}^{\sqrt{(y+\delta)/y}}\log\left|w^2-1\right|dw\right)\;. \qquad \text{(A.12)}
\end{aligned}$$

Since

$$\int_{\sqrt{(y-\delta)/y}}^{\sqrt{(y+\delta)/y}}\log\left|w^2-1\right|dw\geq\int_{-\sqrt{2}}^{\sqrt{2}}\log\left|w^2-1\right|dw>-\infty$$

considering the range of w where $\log|w^2-1|$ is negative, and the last line of (A.12) is a continuous function of $(\epsilon,\, y)$ in the region $(0,\,1]\times(2\delta,\,\infty)$ which converges to 0 if $\epsilon\to 0$ or $y\to\infty$, the desired infimum is a finite value.

For the second bound of the lemma, since

$$\inf_{|y|>1}\mathbb{E}\log|W^2-y|>-\infty$$

by the first part, and since $\log|W^2-y|\geq\log|W^2|$ if $y<0$, it suffices to show that

$$\inf_{0\leq y\leq 1}\mathbb{E}\left[\log|W^2-y|\mathbf{1}\left\{|W^2-y|<1\right\}\right]>-\infty\;.$$

Since

$$\begin{aligned}
\int_{|w^2-y|<1}\log|w^2-y|dw&=\int_{-\sqrt{y+1}}^{\sqrt{y+1}}(\log|w-\sqrt{y}|+\log|w+\sqrt{y}|)dw\\
&=2\int_{-\sqrt{y+1}}^{\sqrt{y+1}}\log|w+\sqrt{y}|dw\geq 2\int_{-1}^{1}\log|w|dw>-\infty\;,
\end{aligned}$$

and the probability density function of W has a maximum value $1/\sqrt{2\pi}$, we get [note that $\log|w^2-y|<0$ on the interval of integration]

$$\inf_{0\leq y\leq 1}\mathbb{E}\log|W^2-y|\geq\frac{2}{\sqrt{2\pi}}\int_{-1}^{1}\log|w|dw>-\infty\;,$$

hence the proof is completed. □

Now, we are ready to verify (2) of Assumption 6. Inspired by the previous lemma, let us write

$$C_W:=\inf_{y\in\mathbb{R}}\mathbb{E}\log|W^2-y|>-\infty\;,$$

Fix $\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$. If $\mathbf{x}$ satisfies $\sum_{i=1}^p \alpha_i x_i + \alpha_0 \neq 0$, it holds that

$$\begin{aligned}\mathbb{E}\log|\mathbf{A}\mathbf{x}+\mathbf{B}-\mathbf{x}| &= \frac{1}{2}\mathbb{E}\log\left(\left(\left(\sum_{i=1}^p \alpha_i x_i + \alpha_0\right)W^2 - x_1\right)^2 + \sum_{i=1}^{p-1}(x_i - x_{i+1})^2\right) \\ &\geq \max\left\{\mathbb{E}\log\left|\left(\sum_{i=1}^p \alpha_i x_i + \alpha_0\right)W^2 - x_1\right|,\ \frac{1}{2}\log\left(\sum_{i=1}^{p-1}(x_i - x_{i+1})^2\right)\right\} \\ &\geq \max\left\{\log\left|\sum_{i=1}^p \alpha_i x_i + \alpha_0\right| + C_W,\ \frac{1}{2}\log\left(\sum_{i=1}^{p-1}(x_i - x_{i+1})^2\right)\right\}\ ,\end{aligned}$$

defining $\log 0 := -\infty$. If $\sum_{i=1}^p \alpha_i x_i + \alpha_0 = 0$, it holds that

$$\mathbb{E}\log|\mathbf{A}\mathbf{x}+\mathbf{B}-\mathbf{x}| \geq \inf_{\sum_{i=1}^p \alpha_i x_i + \alpha_0 = 0} \frac{1}{2}\mathbb{E}\log\left(x_1^2 + \sum_{i=1}^{p-1}(x_i - x_{i+1})^2\right) > -\infty\ ,$$

since $x_1^2 + \sum_{i=1}^{p-1}(x_i - x_{i+1})^2 \to 0$ is equivalent to $\mathbf{x} \to \mathbf{0}$. Hence, if some sequence of vectors $(\mathbf{x}_n)_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty} \mathbb{E}\log|\mathbf{A}\mathbf{x}_n + \mathbf{B} - \mathbf{x}_n| = -\infty \tag{A.13}$$

then we must have

$$\lim_{n\to\infty} \mathbf{x}_n = -(\kappa, \kappa, \ldots, \kappa)\ ,$$

where $\kappa := \alpha_0/(\alpha_1 + \cdots + \alpha_p)$ by the observations above.

On the other hand, if $\mathbf{x}$ is close enough to $-(\kappa, \kappa, \ldots, \kappa)$ so that

$$\left|\sum_{i=1}^p \alpha_i x_i + \alpha_0\right| < 1 \text{ and } x_1 < -\frac{\kappa}{2}\ ,$$

it holds that

$$\begin{aligned}\mathbb{E}\log|\mathbf{A}\mathbf{x}+\mathbf{B}-\mathbf{x}| &\geq \mathbb{E}\log\left|\left(\sum_{i=1}^p \alpha_i x_i + \alpha_0\right)W^2 - x_1\right| \\ &\geq \inf_{|\epsilon|<1,\,|y|>\kappa/2} \mathbb{E}\log|\epsilon W^2 - y|\ .\end{aligned}$$

This contradicts with (A.13), and (2) of Assumption 6 is verified.

References

[1] D. Buraczewski, E. Damek, T. Mikosch, et al. *Stochastic models with power-law tails: The equation $X = AX + B$*. Springer, 2016. 2.1, 3a, 2.1, 2.1, 1, 2, 4.2

[2] E. Damek and S. Mentemeier. Analysing heavy-tail properties of stochastic gradient descent by means of stochastic recurrence equations. *arXiv preprint arXiv:2403.13868*, 2024. 1

[3] H. Furstenberg and H. Kesten. Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469, 1960. 2

[4] M. Gurbuzbalaban, U. Simsekli, and L. Zhu. The heavy-tail phenomenon in SGD. In *International Conference on Machine Learning*, pages 3964–3975. PMLR, 2021. 1, 4.1, 4.1

[5] L. Hodgkinson and M. Mahoney. Multiplicative noise and heavy tails in stochastic optimization. In *International Conference on Machine Learning*, pages 4262–4274. PMLR, 2021. 1

[6] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973. 1, 2.1, 2.1, 3b, 2.1, 2.1, 2.1, 3.1, 1, 3.1, 1, 5.1

[7] W. Rudin. *Real and complex analysis*. Springer, 2016. 5.3

[8] M. Viana. *Lectures on Lyapunov exponents*, volume 145. Cambridge University Press, 2014. 7.1