

# ONE-DIMENSIONAL SYSTEM ARISING IN STOCHASTIC GRADIENT DESCENT

KONSTANTINOS KARATAPANIS,\* *University of Pennsylvania*

## Abstract

We consider stochastic differential equations of the form  $dX_t = |f(X_t)|/t^\gamma dt + 1/t^\gamma dB_t$ , where  $f(x)$  behaves comparably to  $|x|^k$  in a neighborhood of the origin, for  $k \in [1, \infty)$ . We show that there exists a threshold value  $:= \tilde{\gamma}$  for  $\gamma$ , depending on  $k$ , such that if  $\gamma \in (1/2, \tilde{\gamma})$ , then  $\mathbb{P}(X_t \rightarrow 0) = 0$ , and for the rest of the permissible values of  $\gamma$ ,  $\mathbb{P}(X_t \rightarrow 0) > 0$ . These results extend to discrete processes that satisfy  $X_{n+1} - X_n = f(X_n)/n^\gamma + Y_n/n^\gamma$ . Here,  $Y_{n+1}$  are martingale differences that are almost surely bounded.

This result shows that for a function  $F$  whose second derivative at degenerate saddle points is of polynomial order, it is always possible to escape saddle points via the iteration  $X_{n+1} - X_n = F'(X_n)/n^\gamma + Y_n/n^\gamma$  for a suitable choice of  $\gamma$ .

**Keywords:** Stochastic approximations; gradient descent; saddle points; stochastic differential equations

2010 Mathematics Subject Classification: Primary 60F15  
Secondary 49M15

## 1. Introduction

Let  $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $d \geq 1$ , be a vector field. For much of what follows  $F$  arises as the gradient of a potential function  $V$ , namely  $V: \mathbb{R}^d \rightarrow \mathbb{R}$ , and  $F = -\nabla V$ . Now, we define a system driven by

$$X_{n+1} = X_n + a_n(F(X_n) + \xi_{n+1}). \quad (1)$$

To elaborate on the parameters, let  $\mathcal{F}_n$  be a filtration; then  $a_n$ ,  $\xi_n$  are adapted, and the  $\xi_n$  constitute martingale differences, i.e.  $\mathbb{E}(\xi_{n+1} | \mathcal{F}_n) = 0$ . For the purposes of this introduction we will simplify and assume, without any great loss of abstraction, that  $a_n$  is deterministic and either is a constant number or is converging to zero comparably to  $n^{-\gamma}$  (i.e. it is  $\Theta(n^{-\gamma})$ ), where  $\gamma \in (1/2, 1]$ . Also, some additional assumptions on the noise are usually required: one is a boundedness constraint, that is, we assume the existence of a constant  $M$  such that  $|\xi_n| \leq M$  almost surely (a.s.); and secondly we want  $\xi_n$  to be quasi-isotropic (see [DKLH18]), i.e.,  $\mathbb{P}((\theta \cdot \xi_n)^+ > \delta) > \delta$  for any unit direction  $\theta \in \mathbb{R}^d$ . This condition makes sure that the process gets jiggled in every direction. This versatile system is well studied, and it arises naturally in many different areas. In machine learning and statistics, (1) can be a powerful tool for quick optimization and statistical inference (see [AAZB+17], [LLKC18], [CdITTZ16]), among other uses. Furthermore, many urn models are represented by (1). These processes play

Received 6 April 2018; revision received 16 December 2019.

\* Postal address: University of Pennsylvania, Department of Mathematics, 209 South 33rd Street, United States. Email address: kkarat@sas.upenn.edu

© The Author(s), 2021. Published by Cambridge University Press on behalf of Applied Probability Trust.

a central role in probability theory due to their wide applicability in physics, biology, and social sciences; for a comprehensive exposition on the subject see [Pem07].

In machine learning, processes satisfying (1) appear in stochastic gradient descent (SGD). First, to provide context, let us briefly introduce the gradient descent method (GD) and then see why SGD arises naturally from it. The GD is an optimization technique which finds local minima for a potential function  $V$  via the iteration

$$x_{n+1} - x_n = -\eta_n \nabla V(x_n), \quad (2)$$

where in many applications we take  $\eta_n$  to be positive and constant. Notice that (2) is a specialization of (1), with  $F = -\nabla V$ ,  $\xi_{n+1} \equiv 0$ , and  $a_n = \eta_n$ . The above method when applied to non-convex functions has the shortcoming that it may get stuck near saddle points (i.e. points where the gradient vanishes, that are neither local minima nor local maxima), or may locate local minima instead of global ones. The former issue can be resolved by adding noise into the system, which, consequently, helps in pushing the particle downhill and eventually escaping saddle points (see [Pem90] and [KY03, Section 5.8]). For the latter, in general, avoiding local minima is a difficult problem ([GM91] and [RRT17]); however, fortunately, in many instances finding local minima is satisfactory. Recently there have been several problems of interest where this is indeed the case, either because all local minima are global minima ([GHJY15] and [SQW17]), or, in other cases, because local minima provide equally good results as global minima [CHM15]. Furthermore, in certain applications saddle points lead to highly suboptimal results ([JKN15] and [SL16]), which highlights the importance of escaping saddle points.

As described in the previous paragraph, escaping saddle points when performing SGD is an important problem. The saddle problem is well understood when nondegeneracy conditions are imposed. Results showing that asymptotically SGD will escape saddle points date back to works of Pemantle [Pem91] and, more recently, [LSJR16], where the authors prove that random initialization guarantees almost sure convergence to minimizers. The establishment of asymptotic convergence subsequently led to results on how this can be done efficiently [LSJR16].

Processes satisfying (1), when  $a_n$  goes to zero, are known as stochastic approximations, after [RM51]. These processes have been extensively studied since then [KY03]. An important feature is that the step size  $a_n$  satisfies

$$\sum_{n \geq 1} a_n = \infty \quad \text{and} \quad \sum_{n \geq 1} a_n^2 < \infty.$$

This property balances the effects of the noise in the system, so that there is an implicit averaging that, eventually, eliminates the effects of the noise. The previously described system hence behaves similarly to the mean flow: the ordinary differential equation whose right-hand side corresponds to the expectation of the driving term ( $F(X_t)$ ). This heuristic can help us identify the support  $S$  of the limiting process  $X_\infty := \lim_{n \rightarrow \infty} X_n$  in terms of the topological properties of the dynamical system  $\frac{dX_t}{dt} = F(X_t)$  (see [KY03, Chapter 5]). More specifically, in most instances, one can argue that attractors are in  $S$ , whereas repellers or ‘strict’ saddle points are not (see [KY03, Section 5.8]). However, there has not been a systematic approach to finding when a degenerate saddle point, i.e. a point that is neither an attractor nor a repeller, belongs in  $S$ .

Stochastic approximations arise naturally in many different contexts. Some early results were published by [Rup88] and [PJ92]. There, the authors dealt with averaged stochastic gradient descent (ASGD) arising from a strongly convex potential  $V$  with step size  $n^{-\gamma}$ ,

$\gamma \in (1/2, 1]$ . In their work they proved that one can build, with proper scaling, consistent estimators  $\tilde{x}_n$  (for the arg min  $(V)$ ) whose limiting distribution is Gaussian. In learning problems, a modified version of ASGD [RSS12] provides convergence rates to global minima of order  $n^{-1}$ . Additionally, many classical urn processes can be described via (1), where  $a_n$  is of the order of  $n^{-1}$ . Certain efforts are being made towards understanding the support of the limiting process  $X_\infty$ . In specific instances, the underlying problem boils down to understanding an SGD problem: to characterize the support of  $X_\infty$  in terms of the class of critical points of the corresponding potential  $V$ . For a comprehensive exposition on urn processes see [Pem07].

From the previous discussion, some fundamental questions of interest regarding (1) are the following:

1. Does  $X_n$  converge?
2. When does  $X_n$  converge to (local) minima, consequently avoiding saddle points?
3. When does  $X_n$  converge to global minima?
4. How fast does  $X_n$  converge to local minima?

When  $F$  arises from a potential function  $V$ , the first question is for the most part settled: the limit of the process converges, and it is supported on a subset of the set of critical points of  $V$  (see [KY03, Chapter 5]).

Here, our primary focus will be understanding the second question in a one-dimensional setting. More specifically, we will work with processes that solve

$$X_{n+1} - X_n = \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad \gamma \in (1/2, 1]. \tag{3}$$

To put this in context, the antiderivative of  $-f$  would correspond to the potential function  $-V$ . Therefore, if a point  $p$  has a neighborhood  $\mathcal{N}$  such that  $f$  is positive except for  $f(p) = 0$ , then the point  $p$  would be a saddle point.

**Problem 1.1.** Let  $(X_n)_{\geq 1}$  solve (3). Suppose that  $p$  is a saddle point. Find the threshold value, denoted  $\tilde{\gamma}$ , for  $\gamma$ , should it exist, such that the following hold:

1. When  $\gamma \in (1/2, \tilde{\gamma})$ ,  $\mathbb{P}(X_n \rightarrow p) = 0$ .
2. When  $\gamma \in (\tilde{\gamma}, 1]$ ,  $\mathbb{P}(X_n \rightarrow p) > 0$ .

Part 1 of Problem 1.1 guarantees that the SGD avoids saddle points, and hence converges to local minima. Choosing  $\gamma$  appropriately in the first regime (i.e.  $\gamma \in (1/2, \tilde{\gamma})$ ) enables us to optimize the performance of the SGD. In practice, choosing a small step size can slow the rate of convergence; however, a bigger step size may lead the process to bounce around (see [BR95] and [SL87]). In [EDM01] the authors study the rate of convergence for polynomial step sizes in the context of Q-learning for Markov decision processes, and they experimentally demonstrate that for  $\gamma$  approximately  $\frac{17}{20}$  the rate of convergence is optimal.

In the literature there are many results of this type. However, as already mentioned, the vast majority of them require the saddle points to satisfy certain nondegeneracy conditions. In fact, nondegenerate saddle points will never be in the support of  $X_\infty$ . Interestingly enough, the previous conclusion is not always valid for degenerate ones; see [Pem91], in which the support of  $X_\infty$  for a generalized urn model [HLS80] fitting (3) for  $\gamma = 1$  is characterized in terms of a ‘general’ function  $f$ . However, we show that for any  $V$ , under some mild conditions,

we can find  $\gamma$  such that saddle points do not belong to  $S$  (see Theorems 1.2 and 1.4). Hence, we demonstrate that implementing SGD, by adding enough noise, gives the desired asymptotic behavior even in the degenerate case.

It is the hope of the author that this work is a step towards understanding a broader class of non-convex problems. One prospective application would be analyzing complex systems that can be studied by finding a corresponding simpler one-dimensional system. Although non-convex optimization problems are, generally, NP-hard (for a discussion in the context of escaping saddle points see [AG16]), it would be possible to extend the results of this paper to certain classes of problems in higher dimensions, as we are focusing on the asymptotic behavior of the system. Potentially, such an extension can be achieved by reducing the multidimensional problem to a suitable one-dimensional problem and then applying the results of this paper. For an example where the analysis of the asymptotic behavior of a system of stochastic approximations relies on reducing the problem to a one-dimensional problem, see [Pem90]. Also we are trying to establish that if we understand the underlying dynamical system sufficiently, then by adding enough noise, we can guarantee that the process will wander until it is captured by a downhill path, and thus it will eventually escape the unstable neighborhood. Finally, this paper, and even more so a multidimensional extension of it, can serve as a theoretical guarantee of convergence, much in the spirit of the works of [LSJR16] and [Pem90], which were succeeded by efficient algorithms [JGN<sup>+</sup>17].

To extend results to the multidimensional setting using this paper, one would need to find a suitable corresponding one-dimensional system. One potential path to accomplish this is to use a Łojasiewicz-type inequality; for a reference see [Spr, Theorem 2] and [Son12, Lemma 3.2, p. 315]. Before we state the inequality we will need a definition.

**Definition 1.1.** Suppose that  $V: \mathbb{R}^n \rightarrow \mathbb{R}$ . The zero set of  $V$  is denoted by  $Z_V = \{x \in \mathbb{R}^n: V(x) = 0\}$ .

**Theorem 1.1.** Let  $V$  be defined as before. Let  $Z_V$  denote the zero set of  $V$ . Then there is an open set  $0 \in \mathcal{O}$  such that there is a positive constant  $k \in (1, 2)$  such that the following holds:

$$|\nabla V(x)| \geq c|V(x)|^{k/2} \quad \text{for all } x \in \mathcal{O}.$$

Now suppose that  $X_n$  satisfies (1), where  $F(\cdot) = -V$  where  $a_n = \frac{1}{n^\nu}$ . Furthermore, assume  $V: \mathbb{R}^d \rightarrow \mathbb{R}$  is an analytic function such that  $V(0) = \nabla V(0) = 0$ . Hereby, we assume that 0 is a saddle point and that it is also an isolated critical point.

To study whether  $X_n \rightarrow 0$ , our candidate line of attack consists of three distinct steps.

- We start by studying the process  $(V(X_n))_{n \geq 1}$ . Then Theorem 1.1 should give an upper bound on  $|V(X_n)|$ .
- Then the process  $X_n$  may wander into the realm where  $V(X_n) < 0$  with probability bounded from below.
- Lastly, we show that when  $V(X_n) < 0$ , the process may stay negative with probability bounded from below; hence we conclude that  $\mathbb{P}(X_n \rightarrow 0) = 0$ .

For the second part of the strategy we notice that the path from  $X_n$  to  $z \in Z_V$  along the flow  $x'_t = -\nabla V(x_t)$  has length  $V(X_n)$ . So we should expect that as long as  $V(X_n)$  and the remaining noise in the recursion (1) are comparable, then  $X_n$  may wander into the realm where  $V(X_n) < 0$ .

To expand on the third step we need the following definition.

**Definition 1.2.** Suppose that  $V : \mathbb{R}^n \rightarrow \mathbb{R}$  and let  $x \in \mathbb{R}^n$  such that  $V(x) < 0$ . Denote by  $\mathcal{O}_x$  the connected component of  $\{x \in \mathbb{R}^n : V(x) \leq 0\}$  such that  $x \in \mathcal{O}_x$ .

For the last step of the strategy we ought to understand the geometry of the conical region  $\mathcal{O}_{X_n} \cap Z_V$ . For instance, the surface  $\mathcal{O}_{X_n} \cap Z_V$  may be very steep, so that under the slightest perturbation the iterates  $X_n$  may return to the realm  $V(X_n) > 0$ . It is important to note that in certain instances, depending on the surface  $\mathcal{O}_{X_n} \cap Z_V$ , there could be a degenerate saddle point where the iterates could get stuck. However, if we assume, for example, that  $V$  is a homogeneous polynomial, then the angular size of this surface near the origin is not changing over time, and so we should expect a probability bounded from below that the process crosses into the region with  $V(X_n) < -\delta t^{\frac{1-\gamma}{1-k}}$ , where  $k$  is given by Theorem 1.1. To gain intuition on the previous bound one can look at the proof of Proposition 6.1.

### 1.1. Results for the continuous model

We proceed by transitioning to a continuous model. For that purpose we need a potential, a step size, and a noise. However, it is natural to consider, without the need to contemplate, a process defined by

$$dL_t = \frac{f(L_t)}{t^\gamma} dt + \frac{1}{t^\gamma} dB_t, \quad \gamma \in (1/2, 1]. \tag{4}$$

We assume that  $f(0) = 0$  and that  $f$  is otherwise positive in a neighborhood  $\mathcal{N}$  of zero. What we wish to investigate is whether  $L_t$  will not converge to 0 with probability 1, or will converge there with some positive probability. The answer to these questions depends only on the local behavior of  $f$  on  $\mathcal{N}$ .

The main non-convergence result is the following.

**Theorem 1.2.** *Suppose that  $\mathcal{N}$  is a neighborhood of zero. Let  $(L_t)_{t \geq 1}$  be a solution of (4), where  $f(x)$  is Lipschitz. We distinguish two cases depending on  $f$  and the parameters of the system:*

1.  $k|x| \leq f(x)$ ,  $k > \frac{1}{2}$ , and  $\gamma = 1$  for all  $x \in \mathcal{N}$ .
2.  $|x|^k \leq f(x)$ ,  $\frac{1}{2} + \frac{1}{2k} \geq \gamma$ , and  $k > 1$  for all  $x \in \mathcal{N}$ .

*If either 1 or 2 holds, then  $\mathbb{P}(L_t \rightarrow 0) = 0$ .*

In the first part of the theorem, the result holds even in the case  $k = \frac{1}{2}$ ; however, the proof is omitted to avoid repetition. In Part 1, we have only considered  $\gamma = 1$  since that is the only critical case; that is, for  $\gamma < 1$  the effects of the noise would be overwhelming and for all  $k$  we would obtain  $\mathbb{P}(L_t \rightarrow 0) = 0$ .

We now state the main convergence theorem.

**Theorem 1.3.** *Suppose that  $\mathcal{N}$  is a neighborhood of zero. Let  $(L_t)_{t \geq 1}$  be a solution of (4). We distinguish two cases depending on  $f$  and the parameters of the system:*

1.  $k_1|x| \leq f(x) \leq k_2|x|$ ,  $0 < k_i < 1/2$ , and  $\gamma = 1$  for all  $x \in \mathcal{N} \cap (-\infty, 0]$ .
2.  $0 < c|x|^k \leq f(x) \leq |x|^k$ ,  $\frac{1}{2} + \frac{1}{2k} > \gamma$ , and  $k > 1$  for all  $x \in \mathcal{N} \cap (-\infty, 0]$ .

*If either 1 or 2 holds, then  $\mathbb{P}(L_t \rightarrow 0) > 0$ .*

This is proved by first establishing the previous results for monomials, i.e.  $f(x) = |x|^k$  or  $f(x) = k|x|$ , which is done in Sections 3 and 4. We prove the stated theorems in Section 5, by utilizing the comparison results found in Section 2.

In Section 3 we deal with the linear case, i.e.  $f(x) = k|x|$ . There, the stochastic differential equation (SDE) can be explicitly solved, which simplifies matters to a great extent. Firstly, in Subsection 3.2, we prove that when  $k > 1/2$ , the corresponding process a.s. will not converge to 0, which is accomplished by proving that it will converge to infinity a.s. Secondly, in Subsection 3.3, we show that when  $k < 1/2$ , the process will converge to 0 with some positive probability.

In Section 4 we move on to the higher-order monomials, i.e.  $f(x) = |x|^k$ . Here we show that the process will behave as the ‘mean flow’ process  $h(t)$  infinitely often; this is accomplished by studying the process  $L_t/h(t)$ . In Subsection 4.2, the main theorem is that when  $\frac{1}{2} + \frac{1}{2k} \geq \gamma$ ,  $L_t \rightarrow \infty$  a.s. In Subsection 4.3, we show that when  $\frac{1}{2} + \frac{1}{2k} < \gamma$ , the process may converge to 0 with positive probability.

Qualitatively, the previous constraints on the parameters are in accordance with our intuition. To be more specific, when  $k$  increases,  $f$  becomes steeper, which should indicate it is easier for the process to escape. When  $\gamma$  decreases the remaining variance increases; hence we should expect that the process visits the unstable trajectory with greater ease, due to higher fluctuations.

### 1.2. Results for the discrete model

The asymptotic behavior of the discrete processes is the expected one, depending on the parameters of the problem. Here, we study processes satisfying

$$X_{n+1} - X_n \geq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad \gamma \in (1/2, 1), \quad k \in (1, \infty), \tag{5}$$

or

$$X_{n+1} - X_n \leq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad \gamma \in (1/2, 1), \quad k \in (1, \infty), \tag{6}$$

where  $Y_n$  are a.s. bounded (i.e. there is a constant  $M$  such that  $|Y_n| < M$  a.s.),  $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = 0$ , and  $\mathbb{E}(Y_{n+1}^2|\mathcal{F}_n) \geq l > 0$ . The main non-convergence theorem is the following.

**Theorem 1.4.** *Suppose that  $\mathcal{N}$  is a neighborhood of zero. Let  $(X_n)_{n \geq 1}$  solve (5). If  $|x|^k \leq f(x)$ ,  $\frac{1}{2} + \frac{1}{2k} > \gamma$ , and  $k > 1$  for all  $x \in \mathcal{N}$ , then  $\mathbb{P}(X_n \rightarrow 0) = 0$ . For the convergence result the nondegeneracy condition  $\mathbb{E}(Y_{n+1}^2|\mathcal{F}_n) \geq l$  is replaced with the assumption stated in Part 1 of Theorem 1.5.*

**Theorem 1.5.** *Let  $\mathcal{N} = (-3\epsilon, 3\epsilon)$  be a neighborhood of zero. Suppose  $(X_n)_{n \geq 1}$  solve (6). Assume the following:*

1. *There exist  $-\epsilon_2 > -3\epsilon$  and  $-\epsilon_1 < -\epsilon$  such that for all  $M > 0$ , there exists  $n > M$  such that  $\mathbb{P}(X_n \in (-\epsilon_2, -\epsilon_1)) > 0$ .*
2.  *$0 < f(x) \leq |x|^k$ ,  $\frac{1}{2} + \frac{1}{2k} < \gamma$ , and  $k > 1$  for all  $x \in \mathcal{N}$ .*

Then  $\mathbb{P}(X_n \rightarrow 0) > 0$ .

The assumption imposed on  $X_n$ , Part 1 of Theorem 1.5, says that the process should be able visit a neighborhood of the origin for large enough  $n$ . If this constraint is not imposed on the

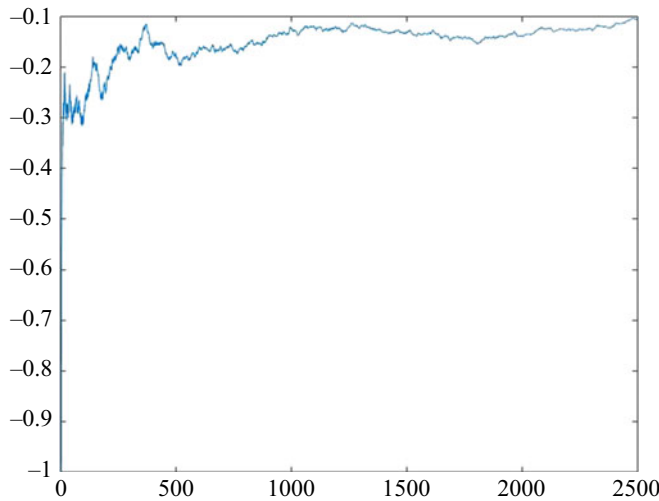


FIGURE 1.  $(X_n)_{n \geq 10}$  and  $X_{10} = -1$ .

process, the previous result need not hold. For instance, the drift could dominate the noise, and consequently the process might never reach a neighborhood of the origin with probability 1. There are processes that naturally satisfy this property; an example is the urn process defined in Section 1 (see [Pem91]).

**Example 1.1.** Suppose that  $X_n$  satisfies

$$X_{n+1} - X_n = \frac{\max(|X_n|^3, 1)}{n^{\frac{3}{4}}} + \frac{U_n}{n^{\frac{3}{4}}},$$

where the  $U_n$  are independent and identically distributed variables, uniformly distributed on  $(-2, 2)$ . As the  $U_n$  dominate the driving term, the assumption 1 is satisfied. And since  $\frac{1}{2} + \frac{1}{2 \cdot 3} < \frac{3}{4}$ , we expect that  $X_n \rightarrow 0$  holds with positive probability. In Figure 1 we can see a typical example where convergence of the iterates occurs.

### 2. Preliminary results

We will now prove two important lemmas that will be needed throughout. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be Lipschitz such that for every  $\epsilon > 0$  there exists  $c$  such that  $f(x) > c > 0$  for all  $x \in \mathbb{R} \setminus (-\epsilon, \epsilon)$ . Also, let  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  be a continuous function such that  $\int_0^\infty g^2(t)dt < \infty$ . Let  $X_t$  satisfy

$$dX_t = f(X_t)dt + g(t)dB_t. \tag{7}$$

**Lemma 2.1.**  $\limsup_{t \rightarrow \infty} X_t \geq 0$  a.s.

*Proof.* We will argue by contradiction. Assume that  $\limsup_{t \rightarrow \infty} X_t < 0$ , and pick  $\delta > 0$  such that  $\limsup_{t \rightarrow \infty} X_t < -\delta$  with positive probability. Then there is a time  $u$  such that  $X_t \leq -\delta$  for all  $t \geq u$ . But this has as an immediate consequence that  $\int_1^t f(X_s)ds \rightarrow \infty$ . However, since the process  $G_t = \int_1^t g(s)dB_s$  has finite quadratic variation, i.e.  $\sup_t \langle G_t \rangle = \int_0^\infty g^2(s)ds < \infty$ ,  $G_t$  stays a.s. finite. The last two observations imply that  $X_t \rightarrow \infty$ , which is a contradiction.



**Lemma 2.2.**  $\liminf_{t \rightarrow \infty} X_t \geq 0$  a.s.

*Proof.* We will again argue by contradiction. Assume that  $\liminf_{t \rightarrow \infty} X_t < 0$  on a set of positive probability. Take an enumeration of the pairs of positive rationals  $(q_n, p_n)$  such that  $q_n > p_n$ . Now, define  $A_n = \{X_t \leq -q_n \text{ i.o.}, X_t \geq -p_n \text{ i.o.}\}$ . Since  $\limsup_{t \rightarrow \infty} X_t \geq 0$ , we have  $\bigcup_{n \geq 0} A_n = \{\liminf_{t \rightarrow \infty} X_t < 0\}$ . Now, for  $t_1 < t_2$  assume that  $X_{t_1} \geq -p_n$  and  $X_{t_2} \leq -q_n$ . Then we see that  $X_{t_2} - X_{t_1} \leq -q_n + p_n$ ; however,

$$\begin{aligned} X_{t_2} - X_{t_1} &= \int_{t_1}^{t_2} f(X_s)ds + \int_{t_1}^{t_2} g(s)dB_s \\ &\geq \int_{t_1}^{t_2} g(s)dB_s. \end{aligned}$$

Hence we conclude that  $\int_{t_1}^{t_2} g(s)dB_s \leq -q_n + p_n$ . By the definition of  $A_n$ , on the event  $A_n$  we can find a sequence of times  $(t_{2k}, t_{2k+1})$  such that  $t_{2k} < t_{2k+1}$  and  $\int_{t_{2k}}^{t_{2k+1}} g(s)dB_s \leq -q_n + p_n$ . Now, if we define  $G_{u,t} = \int_u^t g(s)dB_s$ , we see that  $G_{1,t}$  converges a.s. since it is a martingale of bounded quadratic variation. Hence  $\mathbb{P}(A_n) = 0$ , i.e.  $\mathbb{P}(\liminf_{t \rightarrow \infty} X_t < 0) = 0$ .

The next comparison result is intuitively obvious; however, it will be useful for comparing processes with different drifts.

**Proposition 2.1.** Let  $(C_t)_{t \geq 0}$  and  $(D_t)_{t \geq 0}$  be stochastic processes in the same Wiener space that satisfy

$$dC_t = f_1(C_t)dt + g(t)dB_t, \quad dD_t = f_2(D_t)dt + g(t)dB_t$$

respectively, where  $g, f_1, f_2$  are deterministic real-valued functions. Assume that  $f_1(x) > f_2(x)$  for all  $x \in \mathbb{R}$ , and  $C_{s_0} > D_{s_0}$ . Then  $C_t > D_t$  for every  $t \geq s_0$  a.s.

*Proof.* Define  $\tau = \inf\{t > s_0 | C_t = D_t\}$ , and set  $D_\tau = C_\tau = c$ , for  $\tau < \infty$ . Now, from continuity of  $f_1$  and  $f_2$ , we can find  $\delta$  such that  $f_1(x) > f_2(x)$  for every  $x \in (c - \delta, c]$ . However, for all  $s$  we have

$$C_\tau - D_\tau - (C_s - D_s) = -(C_s - D_s) = \int_s^\tau f_1(C_u) - f_2(D_u)du.$$

Thus, for  $s$  such that  $C_y, D_y \in (c - \delta, c)$  for every  $y \in (s, \tau)$ , we have

$$\begin{aligned} 0 &> -(C_s - D_s) \\ &= \int_s^\tau f_1(C_u) - f_2(D_u)du \\ &> 0. \end{aligned}$$

Therefore  $\{\tau < \infty\}$  has zero probability.

In what follows, we will prove two important lemmas, corresponding to Lemma 2.1 and Lemma 2.2, for the discrete case. We will assume that  $X_n$  satisfies

$$X_{n+1} - X_n \geq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad \gamma \in (1/2, 1), \tag{8}$$

where  $f$  has the property that for every  $\epsilon > 0$  there exists  $c > 0$  such that  $f(x) \geq c$  for every  $x \in (-\infty, -\epsilon)$ , and the  $Y_n$  are defined similarly as in (5).



**Lemma 2.3.**  $\limsup X_n \geq 0$  a.s.

*Proof.* The proof is nearly identical to that of the continuous case (Lemma 2.1).

**Lemma 2.4.**  $\liminf_{t \rightarrow \infty} X_t \geq 0$  a.s.

*Proof.* The proof is identical to that of the continuous case (Lemma 2.2).

We provide a suitable version of the Borel–Cantelli lemma (for a reference see Theorem 5.3.2 in [Dur13]).

**Lemma 2.5.** Let  $\mathcal{F}_n, n \geq 0$ , be a filtration with  $\mathcal{F}_0 = \{0, \Omega\}$ , and let  $A_n, n \geq 1$ , be a sequence of events with  $A_n \in \mathcal{F}_n$ . Then

$$\{A_n \text{ i.o.}\} = \left\{ \sum_{n \geq 1} \mathbb{P}(A_n | \mathcal{F}_{n-1}) = \infty \right\}.$$

### 3. Continuous model, simplest case

#### 3.1. Introduction

Let  $L_t$  be defined by (4), for  $f(x) = k|x|$  and  $\gamma = 1$ . To simplify, we make a time change and consider  $X_t := L_{e^t}$ , and subsequently we obtain

$$\begin{aligned} X_{t+dt} - X_t &= L_{e^t+e^t dt} - L_{e^t} \\ &= k|L_{e^t}|dt + e^{-t}(B_{t+e^t dt} - B_{e^t}) \\ &= k|X_t|dt + e^{-\frac{t}{2}}dB_t, \end{aligned}$$

which is the model we will study. We begin with some definitions: first,

$$dX_t = k|X_t|dt + e^{-\frac{t}{2}}dB_t. \tag{9}$$

We introduce another SDE closely related to the previous one, which will be useful:

$$dK_t = kK_t dt + e^{-\frac{t}{2}}dB_t. \tag{10}$$

It is easy to see that both of these SDEs admit unique strong solutions; for a reference see Theorem 11.2 in Chapter 6 in [RWW87]. Therefore we can construct  $X_t, K_t$  in the classical Wiener space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The solution for the SDE (10) is given by  $K_t = e^{kt}(e^{-t_0k}K_{t_0} + \int_{t_0}^t e^{-s(k+\frac{1}{2})}dB_s)$ . Indeed, substituting in (10) and using Itô’s formula, we get

$$\begin{aligned} dK_t &= a'(t)(k_0 + \int_{t_0}^t b(s)dB_s) + a(t)b(t)dB_t \\ &= \frac{a'(t)}{a(t)}K_t + a(t)b(t)dB_t, \end{aligned}$$

where  $a(t) = e^{k(t-t_0)}$  and  $b(t) = e^{-t(\frac{1}{2}+k)+kt_0}$ , so that  $\frac{a'(t)}{a(t)} = k$  and  $a(t)b(t) = e^{-\frac{t}{2}}$ .

**Proposition 3.1.** Let  $(X_t)_{t \geq t_0}, (K_t)_{t \geq t_0}$  in the Wiener probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  be the solutions of (9), (10) respectively. We start them at time  $t_0, X_{t_0} \geq K_{t_0} \geq 0$ . Then  $X_t \geq K_t$  for every  $t \geq t_0$ .

*Proof.* This is a direct application of Proposition 2.1.

### 3.2. Analysis of $X_t$ when $k > 1/2$

We start by stating the main result of this section, which we will prove at the end of the subsection.

**Theorem 3.1.** *Let  $(X_t)_{t \geq 1}$  be the solution of (9) for  $k > \frac{1}{2}$ ; then  $X_t \rightarrow \infty$  a.s.*

Now we will show that  $(X_t)_{t \geq 1}$  cannot stay negative for all times. This will be accomplished by a direct computation after solving the SDE.

**Proposition 3.2.** *Let  $(X_t)_{t \geq 1}$  be the solution of (9) for  $k > \frac{1}{2}$ . Assume that at time  $s$ ,  $X_s < 0$ . Then  $X_t$  will reach 0 with probability 1, i.e.  $\mathbb{P}(\sup_{u \geq s} X_u > 0) = 1$ .*

*Proof.* First, note that the solution of the SDE (9), run from time  $s$  with initial condition  $X_s < 0$ , coincides with the solution of the SDE  $dX_t = -kX_t dt + e^{-\frac{k}{2}t} dB_t$  before  $X_t$  hits 0. Formally, we define  $\tau_0 = \inf\{t | t \geq s, X_t = 0\}$ . Using the same method as when solving SDE (10), we obtain  $X_t = e^{-kt}(e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})} dB_u)$  on  $\{t < \tau_0\}$ . Set  $G_t = \int_s^t e^{u(k-\frac{1}{2})} dB_u$ , and calculate the quadratic variation of  $G_t$ , namely  $\langle G_t \rangle = (e^{2t(k-\frac{1}{2})} - e^{2s(k-\frac{1}{2})})/(2k - 1)$ . Next, we compute the probability of never returning to zero:

$$\begin{aligned}
\mathbb{P}(\tau = \infty) &= \mathbb{P}\left(\sup_{s < u < \infty} X_u \leq 0\right) \\
&= \mathbb{P}\left(\sup_{s < u < \infty} G_u \leq -e^{ks}X_s\right) \\
&= 1 - \mathbb{P}\left(\sup_{s < u < \infty} G_u > -e^{ks}X_s\right) \\
&= 1 - \lim_{t \rightarrow \infty} \mathbb{P}\left(\sup_{s < u < t} G_u > -e^{ks}X_s\right) \\
&= 1 - \lim_{t \rightarrow \infty} 2\mathbb{P}\left(G_t > -e^{ks}X_s\right), \text{ from the reflection principle} \\
&= 1 - \lim_{t \rightarrow \infty} 2\mathbb{P}\left(N\left(0, \frac{e^{2t(k-\frac{1}{2})} - e^{2s(k-\frac{1}{2})}}{2k - 1}\right) > -e^{ks}X_s\right) \\
&= 0,
\end{aligned}$$

since

$$\frac{e^{2t(k-\frac{1}{2})} - e^{2s(k-\frac{1}{2})}}{2k - 1} \rightarrow \infty.$$

We will now prove two important lemmas that are true for solutions of (9) for any  $k > 0$ .

**Lemma 3.1.** *Let  $(X_t)_{t \geq 1}$  be the solution of (9). Then on the event  $\{X_t \geq 0 \text{ i.o.}\}$ , there is a positive constant  $c < 1$  such that  $\{X_t \geq ce^{-t/2} \text{ i.o.}\}$  holds a.s.*

*Proof.* Assume we start the SDE at time  $t_i$  with initial condition  $X_{t_i} \geq 0$ . Then we see that

$$X_t \geq \int_{t_i}^t k|X_u|du + \int_{t_i}^t e^{-\frac{u}{2}} dB_u \geq \int_{t_i}^t e^{-\frac{u}{2}} dB_u.$$

Set  $G_t = \int_{t_i}^t e^{-\frac{u}{2}} dB_u$ . The quadratic variation of  $G_t$  is  $\langle G_t \rangle = e^{-t} - e^{-t_i}$ . Fix  $0 < c < 1$ . Now, observe that we can always choose  $t$  big enough so that  $\langle G_t \rangle \geq ce^{-t_i}$  for any  $t_i$ .

Then

$$\begin{aligned} \mathbb{P}\left(\sup_{t_i < u < t} X_u > e^{-t_i/2}\right) &\geq \mathbb{P}\left(\sup_{t_i < u < t} G_u > e^{-t_i/2}\right) \\ &= 2\mathbb{P}(G_t > e^{-t_i/2}) \\ &\geq 2\mathbb{P}(N(0, ce^{-t_i}) > e^{-t_i/2}) \\ &= 2\mathbb{P}(N(0, c) > 1) > \gamma > 0. \end{aligned}$$

Let  $g(x) = \inf\{y | e^{-x} - e^{-y} \geq ce^{-x}\}$ . Now we can formally define the sequence of the stopping times. The first stopping time is  $\tau_1 = \inf\{t | X_t \geq 0\}$ ; then we define recursively  $\tau_{i+1} = \inf\{t | t > \tau_i, t > g(\tau_i), X_t \geq 0\}$ . We also define the associated filtration  $\mathcal{F}_n = \mathcal{F}_{\tau_n}$ , for  $n \geq 1$  and  $\mathcal{F}_0 = \{0, \Omega\}$ . Now let  $A_n = \{\exists t, \tau_{n-1} < t < \tau_n, \text{ s.t. } X_t \geq ce^{-t/2}\}$ . By definition  $A_n \in \mathcal{F}_n$ . We find a lower bound for  $\mathbb{P}(A_n | \mathcal{F}_{n-1})$ :

$$\begin{aligned} \mathbb{P}(A_n | \mathcal{F}_{n-1}) &\geq \mathbb{P}\left(\sup_{\tau_{n-1} < u < \tau_n} X_u > ce^{-\tau_{n-1}/2} | \mathcal{F}_{n-1}\right) \\ &\geq \mathbb{P}\left(\sup_{\tau_{n-1} < u < g(\tau_{n-1})} X_u > ce^{-\tau_{n-1}/2} | \mathcal{F}_{n-1}\right) \\ &> \gamma. \end{aligned}$$

On  $\{X_t \geq 0 \text{ i.o.}\}$  the sum  $\sum_{n \geq 1} \mathbb{P}(A_n | \mathcal{F}_{n-1})$  has infinitely many nonzero terms bigger than  $\gamma$ ; hence  $\sum_{n \geq 1} \mathbb{P}(A_n | \mathcal{F}_{n-1}) = \infty$  a.s. Finally, by Lemma 2.5 (Borel–Cantelli) we conclude.

The next lemma uses the previous lemma to establish that on  $\{X_t \geq 0 \text{ i.o.}\}$  we have  $\liminf_{t \rightarrow \infty} X_t > 0$ .

**Lemma 3.2.** *Let  $(X_t)_{t \geq 1}$  be the solution of (9). Then on the event  $\{X_t \geq 0 \text{ i.o.}\}$  we have that  $\{\liminf_{t \rightarrow \infty} X_t > 0\}$  holds a.s.*

*Proof.* Indeed, if we start the process at time  $s$  with initial condition  $X_s \geq ce^{-\frac{s}{2}}$ , then the solution of (9), before hitting 0, is given by

$$X_t = e^{kt} \left( e^{-ks} X_s + \int_s^t e^{-u(k+\frac{1}{2})} dB_u \right) \geq e^{kt} \left( ce^{-s(k+\frac{1}{2})} + \int_s^t e^{-u(k+\frac{1}{2})} dB_u \right).$$

Define  $G_t = \int_s^t e^{-s(k+\frac{1}{2})} dB_s$ . We calculate its quadratic variation:

$$\langle G_t \rangle = \frac{e^{-2tk-t}}{-2k-1} + \frac{e^{-2sk-s}}{2k+1}.$$

Taking  $t \rightarrow \infty$  shows  $\langle G_\infty \rangle = \frac{e^{-2sk-s}}{2k+1}$ . Therefore,

$$\begin{aligned}
 \mathbb{P}\left(\inf_{s \leq u < \infty} X_u > \frac{c}{2}e^{-\frac{s}{2}}\right) &= \mathbb{P}\left(\inf_{s \leq u < \infty} e^{ku} \left(ce^{-s(k+\frac{1}{2})} + G_u\right) > \frac{c}{2}e^{-\frac{s}{2}}\right) \\
 &\geq \mathbb{P}\left(\inf_{s \leq u < \infty} e^{ks} \left(ce^{-s(k+\frac{1}{2})} + G_u\right) > \frac{c}{2}e^{-\frac{s}{2}}\right) \\
 &= \mathbb{P}\left(\inf_{s \leq u < \infty} ce^{-s(k+\frac{1}{2})} + G_u > \frac{c}{2}e^{-s(k+\frac{1}{2})}\right) \\
 &= \mathbb{P}\left(\inf_{s \leq u < \infty} G_u > -\frac{c}{2}e^{-s(k+\frac{1}{2})}\right) \tag{11} \\
 &= 1 - \mathbb{P}\left(\sup_{s \leq u < \infty} G_u > \frac{c}{2}e^{-s(k+\frac{1}{2})}\right) \\
 &= 1 - 2 \lim_{t \rightarrow \infty} \mathbb{P}\left(G_t > -\frac{c}{2}e^{-s(k+\frac{1}{2})}\right), \text{ by the reflection principle} \\
 &= 1 - 2\mathbb{P}\left(N\left(0, \frac{e^{-s(2k+1)}}{2k+1}\right) > \frac{c}{2}e^{-s(k+\frac{1}{2})}\right) \\
 &= 1 - 2\mathbb{P}\left(N\left(0, \frac{1}{k+1}\right) > \frac{c}{2}\right) > \delta > 0.
 \end{aligned}$$

We know that on  $\{X_t \geq 0 \text{ i.o.}\}$  the event  $\{X_t \geq ce^{-\frac{t}{2}} \text{ i.o.}\}$  holds a.s. Therefore, on  $\{X_t \geq 0 \text{ i.o.}\}$ , if we define  $\tau_0 = 0$  and  $\tau_{n+1} = \{t > \tau_n + 1 | X_t \geq ce^{-\frac{t}{2}}\}$ , we see that  $\tau_n < \infty$  a.s., and  $\tau_n \rightarrow \infty$  a.s. Also, we define the corresponding filtration, namely  $\mathcal{F}_n = \sigma(\tau_n)$ .

To show that on the event  $\{X_t \geq ce^{-\frac{t}{2}} \text{ i.o.}\}$  the event  $A = \{\liminf_{t \rightarrow \infty} X_t \leq 0\}$  has probability zero, it suffices to argue that there is a  $\delta$  such that  $\mathbb{P}(A|\mathcal{F}_n) < 1 - \delta$  a.s. for all  $n \geq 1$ . This is immediate from the previous calculation. Indeed,

$$\begin{aligned}
 \mathbb{P}(A|\mathcal{F}_n) &\leq 1 - \mathbb{P}\left(\inf_{\tau_n \leq u < \infty} X_u > \frac{c}{2}e^{-\frac{\tau_n}{2}} | \mathcal{F}_n\right) \\
 &< 1 - \delta.
 \end{aligned}$$

Now we can prove Theorem 3.1.

*Proof of Theorem 3.1.* From Proposition 3.2 we know that  $\{X_t \geq 0 \text{ i.o.}\}$  has probability 1. Therefore from Lemma 3.2 we deduce  $\liminf_{t \rightarrow \infty} X_t > 0$  a.s. Consequently,  $\int_0^\infty |X_u|du \rightarrow \infty$  a.s. At the same time  $\limsup_{t \rightarrow \infty} \int_0^t e^{-\frac{u}{2}} dB_u < \infty$  a.s.; hence  $X_t \rightarrow \infty$  a.s.

**3.3. Analysis of  $X_t$  when  $k < 1/2$**

As before,  $(X_t)_{t \geq 1}$  is the solution of the stochastic differential equation  $dX_t = k|X_t|dt + e^{-\frac{1}{2}t}dB_t$ .

The behavior of  $X_t$  when  $k < 1/2$  is different. The process in this regime can converge to 0 with positive probability. More specifically, we have the following theorem.

**Theorem 3.2.** Let  $(X_t)_{t \geq 1}$  solve (9) with  $k < \frac{1}{2}$ , and define  $A = \{X_t \rightarrow 0\}$ ,  $B = \{X_t \rightarrow \infty\}$ . Then the following hold:

1. Let  $A, B$  be as above. Then  $\mathbb{P}(A \cup B) = 1$ .
2. Both  $A$  and  $B$  are nontrivial, i.e.,  $\mathbb{P}(A) > 0$  and  $\mathbb{P}(B) > 0$ .
3. On  $\{X_t \geq 0 \text{ i.o.}\}$  we get  $X_t \rightarrow \infty$ .

Before proving the theorem, we first need to prove a proposition. We will show that the process, starting from a negative value, will never cross 0 with positive probability.

**Proposition 3.3.** Let  $(X_t)_{t \geq 1}$  solve (9) with  $k < \frac{1}{2}$ . Assume that at time  $s$ ,  $X_s < 0$ . Then  $(X_t)_{t \geq 1}$  will hit 0 with probability  $\alpha$ , where  $0 < \alpha < 1$ .

*Proof.* Define the stopping time  $\tau_1 = \inf\{t \geq s | X_t = 0\}$ . As in Proposition 3.2, the solution for  $X_t$  started at time  $s$  up to time  $\tau_1$  is given by  $X_t = e^{-kt}(e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})} dB_u)$ . We have

$$\begin{aligned} \mathbb{P}(\tau = \infty) &= \mathbb{P}\left(\sup_{s < u < \infty} X_u \leq 0\right) \\ &= 1 - \lim_{t \rightarrow \infty} 2\mathbb{P}\left(N\left(0, \frac{e^{2t(k-\frac{1}{2})} - e^{2t(k-\frac{1}{2})}}{2k-1}\right) > -e^{ks}X_s\right), \text{ as in Proposition 3.2} \\ &= 1 - 2\mathbb{P}(N(0, -e^{2s(k-\frac{1}{2})}/(2k-1)) > -e^{ks}X_s) \\ &= 1 - \alpha. \end{aligned}$$

Therefore  $0 < \alpha < 1$ .

*Proof of Theorem 3.2.*

1. Define the events  $N = \{\exists s \text{ t. } X_t < 0 \forall t \geq s\}$  and  $P = \{X_t \geq 0 \text{ i.o.}\}$ . Of course  $N$  and  $P$  are disjoint and  $\mathbb{P}(P \cup N) = 1$ . To prove Part 1, we will show that  $N \subset \{X_t \rightarrow 0\}$  up to a null set and  $P = \{X_t \rightarrow \infty\}$ . From Lemma 2.2 we know that  $\liminf_{t \rightarrow \infty} X_t \geq 0$  a.s.; therefore  $N \subset \{X_t \rightarrow 0\}$  up to a null set.

To show that  $P = \{X_t \rightarrow \infty\}$ , note that Lemma 3.2 shows that on  $\{X_t \geq 0 \text{ i.o.}\}$ ,  $\liminf_{t \rightarrow \infty} X_t > 0$  a.s. Consequently, on  $\{X_t \geq 0 \text{ i.o.}\}$  we have  $X_t \rightarrow \infty$ , as  $\int_0^\infty |X_u| du \rightarrow \infty$  and  $\limsup_{t \rightarrow \infty} \int_0^t e^{-\frac{u}{2}} dB_u < \infty$  a.s. Therefore,  $P = \{X_t \rightarrow \infty\}$ , which concludes Part 1.

2. The fact that  $\mathbb{P}(A) > 0$  follows immediately from Proposition 3.3. Now, we will prove that  $\mathbb{P}(B) > 0$ . Define the stopping time  $\tau_0 = \inf\{t | X_t = 0\}$ . Also, define  $Y(t, \omega) = 1$  if  $X_s \geq 0$  for all  $s \geq t + 1$ . Observe that  $\{Y_{\tau_0} = 1, \tau_0 < \infty\} \subset P$ . Hence, using the strong Markov property,

$$\begin{aligned} \mathbb{P}(Y_t = 1, \tau < \infty) &= \int_0^\infty \mathbb{P}(\tau = u) \mathbb{P}_0(X_t \geq 0, \forall t \geq 1) du \\ &\geq \int_0^\infty \mathbb{P}(\tau = u) \mathbb{P}_0(K_t \geq 0, \forall t \geq 1) du \quad \text{since } X_t \geq K_t \\ &= \alpha \mathbb{P}_0(K_t \geq 0, \forall t \geq 1) \\ &> 0. \end{aligned}$$

3. This follows immediately from the proof of *I*.

Lastly, we prove a proposition that will be used in Section 5.

**Proposition 3.4.** *Suppose  $(X_t)_{t \geq 1}$ ,  $(Y_t)_{t \geq 1}$  solve (9), with constants  $k$  and  $k_1$  respectively. Suppose that  $0 < k_1 < k < 1/2$ . Let  $\epsilon > 0$ . If  $X_s, Y_s \in (-2\epsilon, -\epsilon)$  a.s., then there is an event  $A$  with positive probability, such that  $X_t, Y_t \in (-3\epsilon, 0)$  for every  $t > s$ .*

*Proof.* Solving the SDE before it hits zero, we find  $X_t = e^{-kt}(e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})}dB_u)$  and  $Y_t = e^{-k_1t}(e^{k_1s}Y_s + \int_s^t e^{u(k_1-\frac{1}{2})}dB_u)$ . Let  $\epsilon > 0$ . Since the process  $G_t = \int_s^t e^{u(k-\frac{1}{2})}dB_u$  has finite quadratic variation, the event  $A = \{G_t \in (-\epsilon, \epsilon) \forall t > s\}$  has positive probability. Set  $\tilde{G}_t = \int_s^t e^{u(k_1-\frac{1}{2})}dB_u$ , and define  $N_t = G_t e^{t(k_1-k)}$ . Using Itô's formula, we find  $dN_t = e^{t(k-\frac{1}{2})}e^{(k_1-k)t}dB_t + (k_1-k)e^{(k_1-k)t}G_t dt$ . Therefore,

$$G_t e^{(k_1-k)t} = \tilde{G}_t + \int_s^t (k_1-k)e^{(k_1-k)u}G_u du.$$

So

$$G_t e^{(k_1-k)t} - \int_s^t (k_1-k)e^{(k_1-k)u}G_u du = \tilde{G}_t.$$

To bound  $|\tilde{G}_t|$  observe that

$$\begin{aligned} -\int_s^t (k_1-k)e^{(k_1-k)u}G_u du &\leq -\epsilon \int_s^t (k_1-k)e^{(k_1-k)u} du \\ &= -\epsilon \left( e^{(k_1-k)t} - e^{(k_1-k)s} \right). \end{aligned}$$

Similarly we obtain  $-\int_s^t (k_1-k)e^{(k_1-k)u}G_u du \geq \epsilon(e^{(k_1-k)t} - e^{(k_1-k)s})$ . Thus on  $A$ , we obtain the following inequalities:

$$-\epsilon e^{(k_1-k)t} + \epsilon \left( e^{(k_1-k)t} - e^{(k_1-k)s} \right) \leq \tilde{G}_t \leq \epsilon e^{(k_1-k)t} - \epsilon \left( e^{(k_1-k)t} - e^{(k_1-k)s} \right).$$

Simplifying, we obtain  $|\tilde{G}_t| \leq \epsilon e^{(k_1-k)s} \leq \epsilon$ . Now we will estimate  $X_t$  on  $A$ . Using that  $\epsilon < |e^{ks}X_s|$  we obtain the upper bound

$$\begin{aligned} X_t &= e^{-kt} \left( e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})}dB_u \right) \\ &\leq e^{-kt}(e^{ks}X_s + \epsilon) \\ &< 0 \end{aligned}$$

and the lower bound

$$\begin{aligned} X_t &= e^{-kt} \left( e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})}dB_u \right) \\ &\geq e^{-kt}(-2e^{ks}\epsilon - \epsilon) \\ &\geq -3\epsilon. \end{aligned}$$

Doing similarly for  $Y_t$ , we conclude.

### 4. Analysis of $dL_t = \frac{|L_t|^k}{t^\gamma} dt + \frac{1}{t^\gamma} dB_t$

#### 4.1. Introduction

As in the previous section, to simplify matters, we will work with reparametrizing  $L_t$ . Set  $\theta(t) = t^{\frac{1}{1-\gamma}}$ , and let  $X_t = L_{\theta(t)}$ . To obtain the SDE that  $X_t$  obeys, notice that  $dB_{\theta(t)} = \sqrt{\theta'(t)}dB_t$ . Therefore

$$\begin{aligned} dX_t &= \frac{|X_t|^k}{\theta(t)^\gamma} \theta'(t)dt + \frac{1}{\theta(t)^\gamma} \sqrt{\theta'(t)}dB_t \\ &= c_1 |X_t|^k dt + c_2 t^{-\frac{\gamma}{1-\gamma}} \sqrt{\theta'(t)}dB_t \\ &= c_1 |X_t|^k dt + c_2 t^{-\frac{\gamma}{2(1-\gamma)}} dB_t, \end{aligned}$$

where  $c_2^2 = c_1 = 1/(1 - \gamma)$ . By abusing the notation we set  $X_t = X_t/c_2$ , which satisfies an SDE of the form

$$dX_t = c |X_t|^k dt + t^{-\frac{\gamma}{2(1-\gamma)}} dB_t, \tag{12}$$

where  $k > 1$ ,  $\gamma \in (1/2, 1)$ , and  $c \in (0, \infty)$ . By a time scaling, we may assume that  $X_t$  solves

$$dX_t = |X_t|^k dt + t^{-\frac{\gamma}{2(1-\gamma)}} dB_t, \tag{13}$$

where  $k > 1$  and  $\gamma \in (1/2, 1)$ . Notice that the noise is scaled differently. However, it will be evident that only the order of the noise is relevant. The SDE (13) will be the primary focus of the next subsection, and the results will apply to solutions of (12) as well.

We define another process that will be fundamental for our analysis, namely  $Z_t = -\frac{X_t}{h(t)}$ , where  $h(t) = -t^{\frac{1}{1-k}}$ . Next, we find the SDE that  $Z_t$  satisfies.

**Proposition 4.1.** *Suppose that  $(X_t)_{t \geq 1}$  solve (12), and set  $C(c) = \frac{1}{c(k-1)}$ ,  $h(t) = -t^{\frac{1}{1-k}}$ . Then the process  $Z_t = -\frac{X_t}{h(t)}$  satisfies*

$$Z_t - Z_s = \int_s^t c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u. \tag{14}$$

Also, before  $X_t$  hits zero we get a solution purely in terms of  $Z_t$ :

$$Z_t - Z_s = \int_s^t c |h(u)|^{k-1} Z_u \left( C - (-Z_u)^{k-1} \right) du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u. \tag{15}$$

*Proof.* Recall that since  $h(t)$  is a continuous function, the covariance  $\langle h(t), Z_t \rangle$  is 0. Using Itô's formula we obtain

$$dZ_t = -\frac{1}{h(t)} dX_t + X_t d\left(-\frac{1}{h(t)}\right).$$



Thus,

$$\begin{aligned} Z_t - Z_s &= \int_s^t -\frac{1}{h(u)} c |X_u|^k du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u + \int_s^t X_u \frac{h'(u)}{h(u)^2} du \\ &= \int_s^t X_u \frac{h'(u)}{h(u)^2} - \frac{1}{h(u)} c |X_u|^k du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u \\ &= \int_s^t c \frac{X_u}{h(u)} \left( \frac{h'(u)}{ch(u)} - \frac{|X_u|^k}{X_u} \right) du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u \\ &= \int_s^t c \frac{X_u}{h(u)} \left( \frac{1}{c(k-1)} \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du + \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u. \end{aligned}$$

The SDE (15) is an immediate consequence of the last line of the above calculation.

In the next proposition, we describe some properties of the noise for the process  $Z_t$ , and we give a very important inequality for Subsection 4.2, which relates the order of the deterministic system converging to zero and the order of the remaining noise for  $X_t$ , i.e. the order of  $\left\langle \int_s^\infty u^{-\frac{\gamma}{2(1-\gamma)}} dB_u \right\rangle$ .

**Proposition 4.2.** *Set  $G'_{s,t} = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u$ , the noise term of (15) and (14).*

1. *In the regime  $\frac{1}{2} + \frac{k}{2} \geq \gamma$ ,  $\langle G'_{s,\infty} \rangle = \infty$ .*
2. *In the regime  $\frac{1}{2} + \frac{k}{2} < \gamma$ ,  $\langle G'_{s,\infty} \rangle < \infty$ .*
3. *Also, given the same conditions as in Part 1 for the pair  $(k, \gamma)$ , the following inequality is true:*

$$\frac{1}{k-1} \geq \frac{2\gamma-1}{2(1-\gamma)}.$$

*Proof.* We calculate its quadratic variation at time  $t$ , namely,

$$\langle G'_{s,t} \rangle = \int_s^t \frac{1}{h(u)^2} u^{-\frac{\gamma}{1-\gamma}} du.$$

Notice that by the definition of  $h(t)$ , we have  $h(t)^{-1} = \Theta \left( t^{\frac{1}{k-1}} \right)$ ; therefore

$$\frac{1}{h(u)^2} u^{-\frac{\gamma}{1-\gamma}} = \Theta \left( u^{\frac{2}{k-1} - \frac{\gamma}{1-\gamma}} \right).$$

Consequently,  $\langle G'_{s,\infty} \rangle = \infty$  when

$$\frac{2}{k-1} - \frac{\gamma}{1-\gamma} \geq -1 \iff \frac{2}{k-1} + \frac{1}{\gamma-1} \geq -2.$$

In the first regime we have

$$\begin{aligned} \frac{1}{2} + \frac{1}{2k} \geq \gamma &\iff \frac{k-1}{2k} \leq 1-\gamma \iff \frac{2k}{k-1} \geq \frac{1}{1-\gamma} \\ &\iff \frac{2}{k-1} + 2 \geq \frac{1}{1-\gamma} \iff \frac{2}{k-1} + \frac{1}{1-\gamma} \geq -2. \end{aligned} \tag{16}$$

So, indeed, when  $\frac{1}{2} + \frac{1}{2k} \geq \gamma$ ,  $\langle G'_{s,\infty} \rangle = \infty$ . Also, from the previous calculation we see that when  $\frac{1}{2} + \frac{1}{2k} < \gamma$ ,

$$\frac{2}{k-1} - \frac{\gamma}{1-\gamma} < -1; \tag{17}$$

therefore when  $\frac{1}{2} + \frac{1}{2k} < \gamma$ ,  $\langle G'_{s,t} \rangle < \infty$ . Finally, rearranging the first inequality of (16), we obtain

$$\frac{1}{k-1} \geq \frac{2\gamma-1}{2(1-\gamma)}.$$

The solution of the SDE (13), when  $X_t$  is positive, explodes in finite time. However, since we are interested in the behavior of  $X_t$  when  $X_t < M$  for a positive constant  $M$ , we may change the drift when  $X_t$  surpasses the value  $M$ , which in turn would imply that the SDE (13) admits strong solutions. One way to do this is by studying the SDE whose drift term is equal to  $|x|^k$  when  $x < M$  and  $M^k$  when  $x > M$ . This SDE can be seen to admit strong solutions for infinite time. The reason is that this process is a.s. bounded from below, and the drift is positive. Also,  $X_t$  cannot explode to plus infinity in finite time since the drift is bounded from above when  $X_t$  is positive. However, for simplicity, we will use the form shown in (13).

**4.2. Analysis of  $X_t$  when  $\frac{1}{2} + \frac{1}{2k} \geq \gamma$ ,  $k > 1$ , and  $\gamma \in (\frac{1}{2}, 1)$**

The main result of this section is the following theorem.

**Theorem 4.1.** *Let  $(X_t)_{t \geq 1}$  solve (13). When  $1/2 + 1/2k \geq \gamma$ ,  $X_t \rightarrow \infty$  a.s.*

We will see its proof at the end of this subsection. Now we will prove an important proposition which shows that  $X_t$  cannot stay far away from the left of the origin.

**Proposition 4.3.** *Let  $(X_t)_{t \geq 1}$  solve (12) for  $c = 1$ . Then, for some  $\beta < 0$ , the event  $\{X_t \geq \beta t^{\frac{1-2\gamma}{2(1-\gamma)}} \text{ i.o.}\}$  has probability 1.*

*Proof.* Set

$$G'_t = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u,$$

which corresponds to the noise term of (15). First we will prove that  $\{X_t \geq C'h(t) \text{ i.o.}\}$  a.s., where  $C' > C^{\frac{1}{k-1}}$  and  $C = C(1) = \frac{1}{k-1}$ . To do so, we will argue by contradiction. Assume that  $A = \{\exists s, X_t < C' \cdot h(t) \forall t > s\}$  has positive measure. Take  $\omega \in A$ , and find  $s(\omega)$  such that  $X_t < C' \cdot h(t)$  for all  $t > s$ . Notice that this implies that  $Z_t < -C'$  for  $t > s$ . Take  $u > s$ ; since  $\frac{|x|^k}{x}$  is increasing we see that

$$\frac{|X_u|^k}{X_u} < C'^{k-1} \frac{|h(u)|^k}{h(u)} < C \frac{|h(u)|^k}{h(u)}.$$

This in turn gives

$$C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} > 0.$$

Therefore

$$\int_s^t \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du > 0$$

for all  $t > 0$ . However, since the process

$$G'_{w,t} := \int_w^s -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u$$

for any fixed  $w$  has infinite quadratic variation, we may find  $G'_{s,t} > -Z_s$ . Now, from (14) we get

$$Z_t = \int_s^t \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du + Z_s + G'_{s,t} > 0.$$

This contradicts the fact that  $Z_t < -C'$ . Therefore  $\{X_t > C'h(t) \text{ i.o.}\}$  a.s.

Finally, in Proposition 4.2 Part 3 we have shown that  $\frac{1}{k-1} \geq \frac{2\gamma-1}{2(1-\gamma)}$ ; therefore

$$-t^{\frac{1}{1-k}} \geq -t^{\frac{1-2\gamma}{2(1-\gamma)}}.$$

So we conclude that there exists a constant  $\beta < 0$  such that  $\{X_t \geq \beta t^{\frac{1-2\gamma}{2(1-\gamma)}} \text{ i.o.}\}$  holds a.s.

**Corollary 4.1.** *Let  $(X_t)_{t \geq 1}$  solve (12) for  $c = 1$ . Then  $\liminf_{t \rightarrow \infty} X_t > 0$  a.s.*

*Proof.* Set  $G_{s,t} = \int_s^t u^{-\frac{\gamma}{2(1-\gamma)}} dB_u$ , and note that  $\langle G_{s,\infty} \rangle = \Theta(s^{\frac{1-2\gamma}{1-\gamma}})$ . Fix  $\gamma > 0$ ; since  $\langle G_{s,\infty} \rangle = \Theta(s^{\frac{1-2\gamma}{1-\gamma}})$  for any  $u > 0$ , it is possible to find  $W(u) > u > 0$  such that

$$\mathbb{P} \left( \sup_{u < t < W(u)} G_{u,t} > \gamma u^{\frac{1-2\gamma}{2(1-\gamma)}} \right) > \delta, \tag{18}$$

for  $\delta$  independent of  $u$ . Take  $\gamma > -\beta$ , where  $\beta$  is such that  $\{X_t \geq \beta t^{\frac{1-2\gamma}{2(1-\gamma)}} \text{ i.o.}\}$  (as in Proposition 4.3). Now, using the lower bound  $X_t - X_s \geq G_{s,t}$ , we obtain

$$\mathbb{P} \left( \sup_{s < t < W(s)} X_t - X_s > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}} \right) \geq \mathbb{P} \left( \sup_{s < t < W(s)} G_{s,t} > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}} \right) > \delta. \tag{19}$$

When  $X_s \geq \beta s^{\frac{1-2\gamma}{2(1-\gamma)}}$ , observe that on the event

$$\left\{ \sup_{s < t < W(s)} X_t - X_s > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}} \right\}$$

there is  $\tau_s$  such that

$$X_{\tau_s} \geq X_s + \beta s^{\frac{1-2\gamma}{2(1-\gamma)}} = (\gamma + \beta) s^{\frac{1-2\gamma}{2(1-\gamma)}} \geq 0.$$

Hence, if we choose a sequence of stopping times such that  $X_{\tau_n} \geq \beta \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}}$  and  $\tau_{n+1} > W(\tau_n)$ , we have

$$\mathbb{P} \left( \sup_{\tau_n < t < \tau_{n+1}} G_{\tau_n,t} > \gamma \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}} | \mathcal{F}_{\tau_n} \right) > \delta.$$

So, by Borel–Cantelli (Lemma 2.5), on the events

$$\left\{ \sup_{\tau_n < t < \tau_{n+1}} X_t - X_{\tau_n} > \gamma \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}} \right\}$$

we may conclude that  $\{X_t \geq 0 \text{ i.o.}\}$  has probability 1. Define  $\tau_n$  as before, except that instead of  $X_{\tau_n} \geq \beta \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}}$  we set  $X_{\tau_n} \geq 0$ ; by Borel–Cantelli we obtain that  $\{X_t \geq \gamma t^{\frac{1-2\gamma}{2(1-\gamma)}} \text{ i.o.}\}$  has probability 1.

Since  $G_{s,t}$  is symmetric and  $\langle G_{s,\infty} \rangle = \Theta \left( s^{\frac{1-2\gamma}{1-\gamma}} \right)$ , we have

$$\begin{aligned} \mathbb{P} \left( \inf_{s < t < \infty} X_t - X_s > -\frac{\gamma}{2} s^{\frac{1-2\gamma}{2(1-\gamma)}} \right) &\geq \mathbb{P} \left( \inf_{s < t < \infty} G_{s,t} > -\frac{\gamma}{2} s^{\frac{1-2\gamma}{2(1-\gamma)}} \right) \\ &= 1 - \mathbb{P} \left( \sup_{s < t < \infty} G_{s,t} > \frac{\gamma}{2} s^{\frac{1-2\gamma}{2(1-\gamma)}} \right) \\ &> \delta' > 0. \end{aligned}$$

for some  $\delta'$  independent of  $s$ .

Define  $\tau_n$  such that  $X_{\tau_n} \geq \gamma \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}}$ , and set  $\mathcal{F}_{\tau_n} = \mathcal{F}_n$ . To show that  $A = \{\liminf_{t \rightarrow \infty} X_t \leq 0\}$  has probability zero, it suffices to argue that there is a  $\delta$  such that  $\mathbb{P}(A|\mathcal{F}_n) < 1 - \delta$  a.s. for all  $n \geq 1$ . This is immediate from the previous calculation. Indeed,

$$\begin{aligned} \mathbb{P}(A|\mathcal{F}_n) &\leq 1 - \mathbb{P} \left( \inf_{\tau_n \leq u < \infty} X_u - X_{\tau_n} > -\frac{\gamma}{2} \tau_n^{\frac{1-2\gamma}{2(1-\gamma)}} | \mathcal{F}_n \right) \\ &< 1 - \delta'. \end{aligned}$$

*Proof of Theorem 4.1.* Since  $X_t$  is a solution of (13), we have  $X_t - X_1 = \int_1^t |X_u|^k du + G_{1,t}$ . From Corollary 4.1 we know that  $\liminf_{t \rightarrow \infty} X_t > 0$  a.s.; therefore  $\int_1^t |X_u|^k du \rightarrow \infty$  a.s. However, since  $\langle G_{1,\infty} \rangle < \infty$ , we have that  $\limsup_{t \rightarrow \infty} |G_{1,t}| < \infty$  a.s. Therefore,  $X_t \rightarrow \infty$  a.s.

### 4.3 Analysis of $X_t$ when $\frac{1}{2} + \frac{1}{2k} < \gamma$ and $k > 1$

We now state the main theorem of this section, which we will prove at the end.

**Theorem 4.2.** *The process  $(X_t)_{t \geq 1}$ , the solution of (12), converges to zero with positive probability, when  $X_1 < 0$ .*

*We prove a technical lemma first.*

**Lemma 4.1.** *Let  $(Z_t)_{t \geq s}$  solve (14), and set*

$$G'_t = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} dB_u.$$

*Suppose that  $Z_s > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$ . Define*

$$A = \left\{ G'_t \in (-\epsilon, \epsilon) \forall t \in (s, s + \delta), \text{ and } G'_t \in \left(-2\epsilon, -\frac{9}{10}\epsilon\right) \forall t \in (s + \delta, \infty) \right\}.$$

*Then the following hold:*

1.  $\mathbb{P}(A) > 0$  for every  $\epsilon, \delta > 0$ .
2. For all  $\epsilon > 0$  small enough, there is  $\delta > 0$  such that  $Z_t < -\frac{5\epsilon}{3}$  for every  $t \in (s, s + \delta)$  on  $A$ .

3. Define  $\tau_C = \inf \left\{ t > s \mid Z_t = -2 \left( \frac{C}{k} \right)^{\frac{1}{k-1}} \right\}$ . Then  $\tau_C > s + \delta$  a.s., where  $\delta$  is the same as in Part 2.

*Proof.*

1. This is immediate, since in Proposition 4.2 Part 2 we have shown that  $\langle G'_\infty \rangle < \infty$ .
2. The first restriction on  $\epsilon$  is such that  $Z_s < -3\epsilon$ . Next, we begin by defining  $f_1$  and  $f_2$  on  $(s, s + \delta)$  satisfying

$$f'(x) = c|h(x)|^{k-1}f(x)(C - (-f(x)^{k-1})), \tag{20}$$

where  $c$  and  $C$  are the same as the parameters of the SDE (14), with initial conditions satisfying

$$-\left(\frac{C}{k}\right)^{\frac{1}{k-1}} < Z_s + \epsilon < f_1(s) < -\frac{5\epsilon}{3}$$

and

$$-\left(\frac{C}{k}\right)^{\frac{1}{k-1}} < f_2(s) < Z_s - \epsilon.$$

Also, we define the function  $q(x) = x(C - (-x)^{k-1})$ , whose derivative is  $q'(x) = C - k(-x)^{k-1}$ , which implies that  $q(x)$  is increasing on  $\left(-\left(\frac{C}{k}\right)^{\frac{1}{k-1}}, 0\right)$ . This function will be important later. We should also note that  $f$  is decreasing in intervals where  $f(x) \in \left(-\left(\frac{C}{k}\right)^{\frac{1}{k-1}}, 0\right)$ , since there  $f'(x) < 0$ .

We can pick the  $\delta > 0$  so that  $f_2(t) > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$  for every  $t \in (s, s + \delta)$ . We will show that  $Z_t > f_2(t)$  on  $(s, s + \delta)$  by contradiction. Using the SDE (15) for  $Z_t$ , we get that

$$Z_t - Z_s = \int_s^t c|h(u)|^{k-1}Z_u \left( C - (-Z_u)^{k-1} \right) du + g(t), \tag{21}$$

where  $g(t)$  is a continuous function such that  $\sup_{t \in (s, s + \delta)} |g(t)| \leq \epsilon$ . Assume that  $f_2$  and  $Z$  become equal at some point, and choose  $t$  to be the first time. Using the integral form of (20), and subtracting it from (15), we get

$$\begin{aligned} 0 &= Z_t - f_2(t) \\ &= \int_s^t c|h(u)|^{k-1}Z_u(C - (-Z_u)^{k-1}) - c|h(u)|^{k-1}f_2(u)(C - (-f_2(u))^{k-1})du \\ &\quad + Z_s - f_2(s) + g(t) \\ &= (t - s)(c|h(\xi)|^{k-1}Z_\xi(C - (-Z_\xi)^{k-1}) - c|h(\xi)|^{k-1}f_2(\xi)(C - (-f_2(\xi))^{k-1})) \\ &\quad + Z_s + g(t) - f_2(s) \\ &> (t - s)(c|h(\xi)|^{k-1}Z_\xi(C - (-Z_\xi)^{k-1}) - c|h(\xi)|^{k-1}f_2(\xi)(C - (-f_2(\xi))^{k-1})), \end{aligned}$$

where in the last line we used that  $Z_s + g(t) - f_2(s) > 0$ . Since  $\xi < t$ , we have that  $Z_\xi > f_2(\xi) > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$ , and consequently  $q(Z_\xi) > q(f_2(\xi))$ , so

$$|h(\xi)|^{k-1}q(Z_\xi) > |h(\xi)|^{k-1}q(f_2(\xi)).$$

Therefore,

$$0 < c|h(\xi)|^{k-1}Z_\xi \left( C - (-Z_\xi)^{k-1} \right) - c|h(\xi)|^{k-1}f_2(\xi) \left( C - (-f_2(\xi))^{k-1} \right),$$

which gives a contradiction.

Arguing similarly we can show that  $f_1(t) > Z_t$  on  $(s, s + \delta)$ , which completes Part 2.

3. Finally, for Part 3 we observe that  $Z_t > f_2(t)$  for  $t \in (s, s + \delta)$ ; hence  $\tau_C > s + \delta$  a.s.

Before proving the theorem we will need the following proposition.

**Proposition 4.4.** *Let  $(X_t)_{t \geq s}$  solve (12). Assume that at time  $s$ ,  $X_s < 0$ , and  $Z_s > -\left(\frac{c}{k}\right)^{\frac{1}{k-1}}$ . Then with positive probability the process never returns to 0.*

*Proof.* The condition  $1/2 + 1/2k < \gamma$ , as has already been shown in Section 4.2, implies that  $\langle G'_\infty \rangle < \infty$ . On the event  $A$  as defined in Lemma 2.3, using (14), we get the following upper and lower bounds for all  $t \geq s + \delta$ :

$$-\frac{X_t}{h(t)} \leq -\frac{X_s}{h(s)} + \int_s^t c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du - \frac{9}{10}\epsilon, \tag{22}$$

$$-\frac{X_t}{h(t)} \geq -\frac{X_s}{h(s)} + \int_s^t c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du - 2\epsilon. \tag{23}$$

**Claim:** On the event  $A$ ,  $X_t < 0$  for all  $t > s$ .

**Proof:** We will argue by contradiction. Assume that  $\mathbb{P}(\{\tau_0 < \infty\} \cap A) > 0$ . We choose  $\epsilon$  such that  $\frac{3\epsilon}{2} < C^{\frac{1}{k-1}}$ . Now, define  $\tau_l = \sup\{t \leq \tau_0 \mid -\frac{X_t}{h(t)} = -\frac{3\epsilon}{2}\}$  and notice that Lemma 4.1 implies that  $\tau_l \epsilon > s + \delta$ , since  $Z_t < -\frac{5\epsilon}{3}$  on  $(s, s + \delta)$ . Also, on  $\{\tau_0 < \infty\} \cap A$  we have  $\tau_l < \infty$ . Then from (23) we see that

$$\int_s^{\tau_l} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du \leq \frac{X_s}{h(s)} + \frac{\epsilon}{2}.$$

Therefore,

$$-\frac{X_s}{h(s)} + \int_s^{\tau_l} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du - \frac{9}{10}\epsilon \leq -\frac{2\epsilon}{5}. \tag{24}$$

Now, notice that  $X_t > \frac{3}{2}\epsilon h(t)$  for every  $t \in (\tau_l, \tau_0)$ , so if  $w \in (\tau_l, \tau_0)$ , we get

$$C \frac{|h(w)|^k}{h(w)} - \frac{|X_w|^k}{X_w} < C \frac{|h(w)|^k}{h(w)} - C \frac{|h(w)|^k}{h(w)} = 0,$$

and of course  $\frac{X_w}{h(w)} > 0$ . So we conclude that

$$\int_{\tau_l}^{\tau_0} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du < 0. \tag{25}$$

Combining (24) and (25), we get that

$$\begin{aligned} 0 &= -\frac{X_{\tau_0}}{h(\tau_0)} \\ &\leq -\frac{X_s}{h(s)} + \int_s^{\tau_0} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du - \frac{9}{10}\epsilon \\ &= -\frac{X_s}{h(s)} + \int_s^{\tau_1} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du - \frac{9}{10}\epsilon \\ &\quad + \int_{\tau_1}^{\tau_0} c \frac{X_u}{h(u)} \left( C \frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u} \right) du \\ &\leq -\frac{2\epsilon}{5}, \end{aligned}$$

a contradiction.

We have developed all the tools necessary to prove the theorem.

*Proof of Theorem 4.2.* Define a stopping time  $\sigma = \inf\{t|Z_t > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\}$ . If the event  $\{\sigma < \infty\}$  has positive probability, then Proposition 4.4 implies that  $X_t$  converges to zero with positive probability. Indeed, recall from Lemma 1 that  $\liminf_{t \rightarrow \infty} X_t \geq 0$  a.s. Therefore, since on the event  $A$  as in Lemma 4.1 we have  $\limsup_{t \rightarrow \infty} X_t \leq 0$ , we deduce that  $\lim_{t \rightarrow \infty} X_t = 0$ . To finish the proof, it suffices to show that when  $\{\sigma < \infty\}$  has zero probability,  $X_t \rightarrow 0$  with positive probability. This is easy to see since  $\mathbb{P}(\{\sigma < \infty\}) = 0$  implies that  $Z_t$  never hits zero; therefore  $\limsup_{t \rightarrow \infty} X_t \leq 0$  on  $\{\sigma < \infty\}$ .

We now prove a proposition that will be used in the next section.

**Proposition 4.5.** *Let  $(X_t)_{t \geq s}$  solve (12). Take the event  $A$  such that Lemma 4.1 holds, with  $\epsilon < \left(\frac{C}{k}\right)^{\frac{1}{k-1}}$ , where  $C(c) = \frac{1}{c(k-1)}$  as in the parameter  $C$  of the SDE (14). Then, on  $A$ , the process  $X_t$  stays within a region of the origin. More specifically,  $Z_t > -2\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$ .*

*Proof.* Let  $\tau_C = \inf\left\{t > s | Z_t = -2\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\right\}$ , and define

$$\sigma = \sup\left\{\tau_C > t > s | Z_t = -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\right\}.$$

We will show that  $\tau_C = \infty$  a.s. We assume otherwise, and reach a contradiction. From Lemma 4.1 Part 3, we know that  $\tau_C > s + \delta$ . Therefore,

$$\begin{aligned} Z_{\tau_C} &\geq Z_s + \int_s^{\tau_C} c|h(u)|^{k-1}Z_u \left( C - (-Z_u)^{k-1} \right) du - 2\epsilon \\ &\geq Z_{\sigma} - Z_{\sigma} + Z_s + \int_s^{\tau_C} c|h(u)|^{k-1}Z_u \left( C - (-Z_u)^{k-1} \right) du - 2\epsilon \\ &\geq Z_{\sigma} + \frac{9\epsilon}{10} - 2\epsilon > -2\left(\frac{C}{k}\right)^{\frac{1}{k-1}}, \end{aligned}$$

the desired contradiction.



**5. Analysis of  $dL_t = \frac{f(L_t)}{t^\gamma} dt + \frac{1}{t^\gamma} dB_t$**

For this section, we assume that  $f$  is globally Lipschitz. For  $f$  as before, we define

$$dL_t = \frac{f(L_t)}{t^\gamma} dt + \frac{1}{t^\gamma} dB_t, \quad \gamma \in \left(\frac{1}{2}, 1\right]. \tag{26}$$

By our assumptions on  $f$ , the SDE (26) admits strong solutions. Also, we define a more general SDE, namely

$$dX_t = f(X_t)dt + g(t)dB_t, \tag{27}$$

where  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{> 0}$  is continuous, and  $T = \int_0^\infty g^2(t)dt$  is possibly infinite.

**Proposition 5.1.** *Let  $(X_t)_{t \geq 1}$  be a solution of (27). Then for every  $t, c > 0$  and  $x \in \mathbb{R}$ ,  $\mathbb{P}(X_t \in (x - c, x + c)) > 0$ .*

*Proof.* Firstly, we change time. Let  $\xi(t) = \int_0^t g^2(t)dt$ , and define  $\tilde{X}_t = X_{\xi^{-1}(t)}$ . Then

$$d\tilde{X}_t = \frac{f(\tilde{X}_t)}{g^2(\xi^{-1}(t))} dt + dB_t. \tag{28}$$

This gives a well-defined SDE whose solution is defined on  $[0, T']$  for  $T' \in \mathbb{R}$ ,  $\xi(t) < T' < T$ . The path-space measure of  $\tilde{X}_t$  is mutually absolutely continuous to the one induced from the Brownian motion. Since the Brownian motion satisfies the property described in the proposition, so does  $X_t$ .

We give the proofs of Theorems 1.2 and 1.3. For the proofs, we use that the theorems hold if and only if they hold for their corresponding reparametrizations.

*Proof of Theorem 1.2 Parts 1 & 2.* Both parts can be proved simultaneously. Let  $\tau = \inf\{t | X_t \in (-\epsilon, \epsilon)\}$  and  $\tau' = \inf\{t > \tau | X_t \in \{-3\epsilon, 3\epsilon\}\}$ . Now, define a stochastic process  $(U_t)_{t \geq \tau}$  started on  $\mathcal{F}_\tau$  that satisfies (13) or (9) with  $U_\tau = -2\epsilon$ . From Proposition 2.1, we see that  $U_t < X_t$  for all  $t \in (\tau, \tau')$ . Now we can see that  $\mathbb{P}(\tau' = \infty) = 0$ . Indeed,

$$\mathbb{P}(\tau' = \infty) \leq \mathbb{P}(U_t \leq 3\epsilon \text{ for all } t \geq \tau) \leq 1 - \mathbb{P}(U_t \rightarrow \infty) = 0.$$

*Proof of Theorem 1.3 Part 1.* Suppose  $\mathcal{N} = (-3\epsilon, 3\epsilon)$  for  $\epsilon > 0$ . Without loss of generality and for the purposes of this proof, assume that

$$\epsilon < \min \left( \left( \frac{C(1)}{k} \right)^{\frac{1}{k-1}}, \left( \frac{C(c)}{k} \right)^{\frac{1}{k-1}} \right).$$

Pick a time  $z$  such that

$$h(t) \geq -\frac{3}{2} \left( \frac{k}{C(c)} \right)^{\frac{1}{k-1}} \epsilon$$

for all  $t \geq z$ , and define  $\tau = \inf\{t \geq z | X_t \in (-\epsilon, \epsilon)\}$  and  $\tau' = \inf\{t > \tau | X_t \in \{-3\epsilon, 3\epsilon\}\}$ . From Proposition 5.1,  $\tau < \infty$  with positive probability. Now we define two stochastic processes  $(Y_t)_{t \geq \tau}, (Y'_t)_{t \geq \tau}$  in the same probability space as  $X_t$  started on  $\mathcal{F}_\tau$ , that satisfy (12) with drift constants 1 and  $c$  respectively. From Proposition 2.1, we see that if  $Y_\tau > X_\tau > Y'_\tau$ , then  $Y_t > X_t > Y'_t$  for all  $t \in (\tau, \tau')$ . We set  $Y'_\tau$  such that  $X_\tau > Y'_\tau$ , and

$$Z'_t = -\frac{Y'_t}{h(t)} > \max \left( -\left( \frac{C(1)}{k} \right)^{\frac{1}{k-1}}, -\left( \frac{C(c)}{k} \right)^{\frac{1}{k-1}} \right).$$

Now we need to show that  $\{\tau' = \infty\} \cap \{Y_t \rightarrow 0\} \cap \{Y'_t \rightarrow 0\}$  is nontrivial. Take  $\epsilon_1$  and  $\epsilon_c$ , both less than  $\epsilon$ , as in the statement of Lemma 4.1 for  $Y_t$  and  $Y'_t$  respectively, and pick  $\epsilon' = \min(\epsilon_1, \epsilon_c)$ . For  $\epsilon'$ , using Lemma 4.1, we know we can find  $\delta_1$  and  $\delta_c$  such that on

$$A_1 = \{G_t \in (-\epsilon', \epsilon') \text{ for all } t \in (s, s + \delta_1), \text{ and } G_t \in \left(-2\epsilon', -\frac{9}{10}\epsilon'\right) \text{ for all } t \in (s + \delta_1, \infty)\}$$

we have  $Y_s \rightarrow 0$ , and on

$$A_c = \{G_t \in (-\epsilon', \epsilon') \text{ for all } t \in (s, s + \delta_c), \text{ and } G_t \in \left(-2\epsilon', -\frac{9}{10}\epsilon'\right) \text{ for all } t \in (s + \delta_c, \infty)\}$$

we have  $Y'_s \rightarrow 0$ . From here, since  $A_1 \cap A_c$  is nontrivial, we only need to argue that  $\{\tau' = \infty\} \supset A \cap A_c$ . From the remark of Lemma 4.1 we see that  $Y_t$  and  $Y'_t$  always stay below 0 on  $A_1 \cap A_c$ . Also, from Proposition 4.5, we see that

$$Z_t^{Y'} > -2 \left(\frac{C(c)}{k}\right)^{\frac{1}{k-1}}.$$

Equivalently, and using that

$$h(t) \geq -\frac{3}{2} \left(\frac{k}{C(c)}\right)^{\frac{1}{k-1}} \epsilon,$$

we have

$$\begin{aligned} Y'_t &> 2h(t) \left(\frac{C(c)}{k}\right)^{\frac{1}{k-1}} \\ &\geq -3\epsilon. \end{aligned}$$

We now prove the second part of Theorem 1.3.

*Proof of Theorem 1.3 Part 2.* Let  $\mathcal{N} = (-3\epsilon, 0)$ . Define  $\tau = \inf\{t \geq 1 | X_t \in (-\frac{3\epsilon}{2}, -\frac{5\epsilon}{4})\}$ , and let the exit time from  $\mathcal{N}$  be  $\tau_e = \inf\{t | X_t \notin (-3\epsilon, 0)\}$ . From Proposition 5.1, we have that  $\tau < \infty$  holds with positive probability. Define  $(Y_t)_{\tau \leq t \leq \tau_e}, (Y'_t)_{\tau \leq t \leq \tau_e}$  to be two processes that satisfy (9) with constants  $k_1, k_2$  respectively. Suppose that  $Y_\tau < X_\tau < Y'_\tau$  and  $Y_\tau, Y'_\tau \in (-2\epsilon, -\epsilon)$ . Then by Proposition 2.1, we get  $Y_t < X_t < Y'_t$  for all  $t \in (\tau, \tau_e)$ . Now, using Proposition 3.4, there is an event  $A$  such that  $Y_t, Y'_t \in (-3\epsilon, 0)$  for all  $t \geq \tau$ . Consequently,  $X_t \in (-3\epsilon, 0)$  for all  $t \geq \tau$ , since  $\tau_e = \infty$  on  $A$ . Finally, using Lemma 2.2 we conclude that  $Y_t \rightarrow 0$  on  $A$ ; hence also  $X_t \rightarrow 0$  on  $A$ .

### 6. The discrete model

#### 6.1. Analysis of $X_t$ when $\frac{1}{2} + \frac{1}{2k} > \gamma, k > 1$ , and $\gamma \in (1/2, 1)$

Before proving Theorem 1.4, as described in Section 1.2, we assume that  $X_n$  satisfies

$$X_{n+1} - X_n \geq \frac{|X_n|^k}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad k > 1 \text{ and } \gamma \in (1/2, 1), \tag{29}$$

where the  $Y_n$  are a.s. bounded and  $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = 0$ . In this section we additionally require  $Y_n$  to satisfy  $\mathbb{E}(Y_{n+1}^2 | \mathcal{F}_n) \geq l > 0$ .

**Theorem 6.1.** *Let  $(X_n)_{n \geq 1}$  solve (29). When  $1/2 + 1/2k > \gamma, X_t \rightarrow \infty$  a.s. Now we develop the necessary tools to prove this theorem.*

**Proposition 6.1.** *The process  $(X_n)_{n \geq 1}$  gets close to the origin infinitely often. More specifically, for  $\beta < 0$  the event  $\left\{X_n \geq \beta n^{\frac{1-2\gamma}{2}} \text{ i.o.} \right\}$  has probability 1.*

*Proof.* From the restrictions on  $k$  we obtain

$$\frac{1}{2} + \frac{1}{2k} \geq \gamma \iff \frac{k-1}{2k} \leq 1 - \gamma.$$

Set  $h(t) = -t^{\frac{1-\gamma}{1-k}}$ , and define  $Z_n = -\frac{X_n}{h_n}$ . From here, on the event  $\{X_m < 0 \text{ for all } m \geq n\}$ , we get the following recursion:

$$Z_{n+1} - Z_n \geq -\frac{X_{n+1}}{h(n+1)} + \frac{X_n}{h(n)} \tag{30}$$

$$\begin{aligned} &\geq -X_n \left( \frac{1}{h(n+1)} - \frac{1}{h(n)} \right) - \frac{|X_n|^k}{n^\gamma h(n+1)} - \frac{Y_{n+1}}{n^\gamma h(n+1)} \\ &\geq X_n \frac{1-\gamma}{k-1} \xi_n^{-\frac{1-\gamma}{1-k}-1} - \frac{|X_n|^k}{n^\gamma h(n+1)} - \frac{Y_{n+1}}{n^\gamma h(n+1)}, \text{ where } \xi_n \in (n, n+1) \\ &\geq \frac{X_n}{h(n+1)n^\gamma} \left( \frac{1-\gamma}{k-1} \xi_n^{-\frac{1-\gamma}{1-k}-1} h(n+1)n^\gamma - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} \\ &\geq \frac{X_n}{h(n+1)n^\gamma} \left( -a_n \frac{1-\gamma}{k-1} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} \end{aligned} \tag{31}$$

$$\geq \frac{X_n}{h(n+1)n^\gamma} \left( -\frac{2(1-\gamma)}{k-1} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)}, \tag{32}$$

where

$$a_n = \frac{\xi_n^{-\frac{(1-\gamma)}{1-k}-1} h(n+1)n^\gamma}{-|h(n)|^{k-1}}. \tag{33}$$

To justify the inequality (32), for large enough  $n$ , notice that  $a_n \rightarrow 1$ .

Define

$$G'_{s,n} = \sum_{i=s}^{n-1} \frac{Y_i + 1}{i^\gamma h(i+1)}.$$

We will see that  $G'_{1,n}$  grows big enough so that  $Z_n$  must, at certain times, get close enough to the origin so that  $X_n$  surpasses a constant multiple of  $h(n)$ . To this end, we have the following lemma.

**Lemma 6.1.**  $\limsup_{n \rightarrow \infty} G'_{1,n} = \infty$  a.s.

We use the following theorem; for a reference see [Fis92, Theorem 1, p. 676].

**Theorem 6.2.** *Let  $X_n$  be a martingale difference such that  $\mathbb{E}(X_i^2 | \mathcal{F}_{i-1}) < \infty$ . Set  $s_n^2 = \sum_{i=1}^n \mathbb{E}(X_i^2 | \mathcal{F}_{i-1})$ , and define  $\phi(x) = (2 \log_2(x^2 \vee e^2))^{\frac{1}{2}}$ . We assume that  $s_n \rightarrow \infty$  a.s. and that  $|X_i| \leq \frac{K_i s_i}{\phi(s_i)}$  a.s., where  $K_i$  is  $\mathcal{F}_{i-1}$ -measurable with  $\limsup_{i \rightarrow \infty} K_i < K$  for some constant  $K$ . Then there is a positive constant  $\epsilon(K)$  such that*

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^n \frac{X_i}{s_n \phi(s_n)} \geq \epsilon(K) \text{ a.s.}$$

It is clear that  $G'_{1,n}$  satisfies all the hypotheses required for the aforementioned theorem to hold.

From Lemma 6.1, it is immediate that for any random time  $s$  (not necessarily a stopping time),  $\limsup_{n \rightarrow \infty} G'_{s,n} = \infty$  a.s.

Now, we return to the proof of Proposition 6.1. Assume that there is  $n_0$  such that

$$X_n < - \left( \frac{3(1-\gamma)}{k-1} \right)^{\frac{1}{k-1}} n^{\frac{1-\gamma}{1-k}}$$

for all  $n \geq n_0$ . Then, since  $\frac{|x|^k}{x}$  is increasing, we get that

$$\frac{|X_n|^k}{X_n} < - \frac{3(1-\gamma)}{k-1} n^{-1+\gamma}.$$

Therefore,

$$- \frac{2(1-\gamma)}{k-1} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} > - \frac{2(1-\gamma)}{k-1} n^{-1+\gamma} + \frac{3(1-\gamma)}{k-1} n^{-1+\gamma} = \frac{(1-\gamma)}{k-1} n^{-1+\gamma} > 0$$

So

$$\begin{aligned} Z_n &\geq Z_{n_0} + \sum_{i=n_0}^n \frac{X_i}{h(i+1)i^\gamma} \left( \frac{2(1-\gamma)}{k-1} |h(i)|^{k-1} - \frac{|X_i|^k}{X_i} \right) + G'_{n_0,n} \\ &> Z_{n_0} + G'_{n_0,n}, \end{aligned}$$

which gives  $\limsup_{n \rightarrow \infty} Z_n = \infty$ . This is a contradiction, since it would imply  $X_n \geq 0$  infinitely often.

Since  $n^{\frac{1-\gamma}{1-k}} = o(n^{\frac{1-2\gamma}{2}})$ , for every constant  $\beta < 0$  the event  $\{X_n \geq \beta n^{\frac{1-2\gamma}{2}} \text{ i.o.}\}$  holds a.s.

We define  $G_{n,u} = \sum_{i=n}^{u-1} \frac{Y_{i+1}}{i^\gamma}$ ; this is an important quantity for the next lemma and the remainder of the section.

**Lemma 6.2.** *For any  $n$ , we can find  $a_1 > 0, \delta > 0$  such that  $\mathbb{P}\left(\sup_{u \geq n} G_{n,u} \geq a_1 n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) > \delta$  and  $\mathbb{P}\left(G_{n,\infty} \geq a_1 n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) > \delta$ .*

*Proof.* Define  $\tau = \inf\{u \geq n \mid G_{n,u} \notin (-a_2 n^{\frac{1-2\gamma}{2}}, a_2 n^{\frac{1-2\gamma}{2}})\}$ . We calculate the stopped variance of  $G_\tau := G_{n,\tau}$ . We will do so recursively; fix  $m \geq n$  and calculate as follows:

$$\begin{aligned} \mathbb{E}((G_{\tau \wedge m+1})^2 \mid \mathcal{F}_n) - \mathbb{E}(G_{\tau \wedge m})^2 \mid \mathcal{F}_n &= \mathbb{E}\left(1_{\tau > m} \left(2 \frac{Y_{m+1}}{m^\gamma} G_m + \frac{Y_{m+1}^2}{m^{2\gamma}}\right) \mid \mathcal{F}_n\right) \\ &= \mathbb{E}\left(1_{\tau > m} 2 \frac{Y_{m+1}}{m^\gamma} G_m \mid \mathcal{F}_n\right) + \mathbb{E}\left(1_{\tau > m} \frac{Y_{m+1}^2}{m^{2\gamma}} \mid \mathcal{F}_n\right) \\ &= 0 + \mathbb{E}\left(1_{[\tau > m]} \mathbb{E}\left(\frac{Y_{m+1}^2}{m^{2\gamma}} \mid \mathcal{F}_m\right) \mid \mathcal{F}_n\right) \\ &\geq \epsilon \frac{1}{m^{2\gamma}} \mathbb{E}(1_{[\tau > m]} \mid \mathcal{F}_n) \\ &\geq \epsilon \frac{1}{m^{2\gamma}} \mathbb{P}(\tau = \infty \mid \mathcal{F}_n). \end{aligned}$$

Therefore,

$$\begin{aligned} E((G_{\tau \wedge m})^2 | \mathcal{F}_n) &\geq E((G_{\tau \wedge n})^2 | \mathcal{F}_n) + c\mathbb{P}(\tau = \infty | \mathcal{F}_n)(n^{1-2\gamma} - (m-1)^{1-2\gamma}) \\ &= c\mathbb{P}(\tau = \infty | \mathcal{F}_n)(n^{1-2\gamma} - (m-1)^{1-2\gamma}). \end{aligned} \tag{34}$$

Notice that since the  $Y_n$  are a.s. bounded,  $|G_\tau| \leq a_2 n^{\frac{1-2\gamma}{2}} + \frac{M}{n^\gamma}$ , and since  $\gamma > \gamma - 1/2$ , we get that  $|G_\tau| \leq 2a_2 n^{\frac{1-2\gamma}{2}}$  for  $n$  large enough. For  $m$  large, we can find a constant  $c'$  such that  $n^{1-2\gamma} - (m-1)^{1-2\gamma} \geq c'n^{1-2\gamma}$ . Using (34), we obtain

$$\frac{2a_2 n^{1-2\gamma}}{c'n^{1-2\gamma}} = \frac{2a_2}{c'} \geq \mathbb{P}(\tau = \infty | \mathcal{F}_n).$$

Choosing  $a_2$  small enough, we may conclude  $\mathbb{P}(\tau < \infty | \mathcal{F}_n) > 1/2$  for all  $n$  large enough.

Now we take any martingale  $M_n$  starting at 0 such that it exits an interval  $(-2a, 2a)$  with probability at least  $p$ , and  $|M_{n+1} - M_n| < a$  a.s. Then we stop the martingale upon exiting the interval  $(-2a, 2a)$ ; that is, define  $\tau_-$  to be the first time  $M_n$  goes below  $-2a$  and  $\tau_+$  to be the first time that  $M_n$  surpasses  $2a$ , and set  $\tau = \tau_- \wedge \tau_+$ . Using the optimal stopping theorem for the bounded martingale  $M_{\tau \wedge n}$  and taking  $n$  to infinity, we obtain

$$\begin{aligned} 0 = \mathbb{E}(M_\tau) &\leq -2a\mathbb{P}(\tau_- < \tau_+) + 3a\mathbb{P}(\tau_- > \tau_+) + 2a\mathbb{P}(\tau = \infty) \\ &= -2ap + 2a(1-p) + 5a\mathbb{P}(\tau_- > \tau_+). \end{aligned}$$

So  $\mathbb{P}(\tau_- > \tau_+) \geq \frac{4p-2}{5}$ , which implies that  $\mathbb{P}(\sup_n M_n \geq 2a) \geq \frac{4p-2}{5}$ .

The previous argument applied to  $G_{n,u}$ , given  $\mathcal{F}_n$ , concludes the proof of the first part of the lemma. Indeed, since the probability  $p$  of exiting the interval is bigger than  $1/2$ , we may deduce that  $\frac{4p-2}{5} > 0$ .

For the second part of the lemma, we use the following inequality: let  $M_n$  be a martingale such that  $M_0 = 0$  and  $\mathbb{E}(M_n^2) < \infty$ . Then

$$\mathbb{P}\left(\max_{n \geq u \geq 0} M_u \geq \lambda\right) \leq \frac{\mathbb{E}(M_n^2)}{\mathbb{E}(M_n^2) + \lambda^2}$$

(for a reference see [Dur13, Exercise 5.4.5, p. 213]). Let  $\tau$  be the first time  $G_{n,u}$  surpasses  $a_2 n^{1-2\gamma}$ . Condition on  $[\tau < \infty]$ , and notice that  $G_{n,\infty} > \frac{a_2}{2} n^{1-2\gamma}$  when  $\inf_{u \geq \tau} G_{\tau,u} > -\frac{a_2}{2} n^{1-2\gamma}$ . Using the previous inequality and the fact that  $\frac{x}{x+1}$  is increasing, we obtain

$$\begin{aligned} \mathbb{P}\left(G_{n,\infty} \leq \frac{a_2}{2} n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_\tau, [\tau < \infty]\right) &\leq \mathbb{P}\left(\inf_{u \geq \tau} G_{\tau,u} \leq -\frac{a_2}{2} n^{1-2\gamma} \mid \mathcal{F}_\tau, [\tau < \infty]\right) \\ &\leq \frac{\mathbb{E}((G_{\tau,\infty})^2 | \mathcal{F}_\tau, [\tau < \infty])}{\mathbb{E}((G_{\tau,\infty})^2 | \mathcal{F}_\tau, [\tau < \infty]) + \frac{a_2^2}{4} n^{1-2\gamma}} \\ &\leq \frac{c\tau^{1-2\gamma}}{c\tau^{1-2\gamma} + \frac{a_2^2}{4} n^{1-2\gamma}} \\ &\leq \frac{c}{c + \frac{a_2^2}{4}}. \end{aligned}$$

Therefore,

$$\mathbb{P}\left(G_{n,\infty} \geq \frac{a_2}{2} n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) \geq \mathbb{P}(\tau < \infty) \frac{\frac{a_2^2}{4}}{c + \frac{a_2^2}{4}},$$

which concludes the proof.

For any stopping time  $\sigma$ , we get the following version of the previous lemma.

**Lemma 6.3.** *For any  $n$ , we can find  $a_1 > 0, \delta_1 > 0, \delta_2 > 0$  such that  $\mathbb{P}(\sup_{u \geq \sigma} G_{\sigma,u} \geq a_1 \sigma^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_\sigma) > \delta_1$  and  $\mathbb{P}(G_{\sigma,\infty} \geq a_1 \sigma^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_\sigma) > \delta_2$ .*

**Corollary 6.1.** *The event  $\{X_n \geq 0 \text{ i.o.}\}$  holds a.s.*

*Proof.* For any  $m, n$  we get the lower bound  $X_m - X_n \geq G_{n,m}$ . Now, we define an increasing sequence of stopping times  $\tau_n$ , going to infinity a.s., such that  $X_{\tau_n} \geq \beta \tau_n^{\frac{1-2\gamma}{2}}$  for  $|\beta| < a_1$ , where  $a_1$  is such that  $\mathbb{P}(\sup_{u \geq \sigma} G_{\tau_n,u} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_{\tau_n}) > \delta_1$ , whose existence is guaranteed by Lemma 6.3. From Proposition 6.1, we can do so with all  $\tau_n$  a.s. finite. Hence,

$$\begin{aligned} \mathbb{P}\left(\sup_{\infty \geq u \geq \tau_n} X_u - X_{\tau_n} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_{\tau_n}\right) &\geq \mathbb{P}\left(\sup_{\infty \geq u \geq \tau_n} G_{\tau_n,u} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_{\tau_n}\right) \\ &> \delta_1 > 0. \end{aligned}$$

Therefore, by Borel–Cantelli on the event  $\{X_{\tau_n} \geq \beta \tau_n^{\frac{1-2\gamma}{2}} \text{ i.o.}\}$ , we get  $\{X_{\tau_n} \geq 0 \text{ i.o.}\}$ . Therefore  $\{X_n \geq 0 \text{ i.o.}\}$  holds a.s.

*Proof of Theorem 6.1.* Again we define an increasing sequence of stopping times  $\tau_n$ , going to infinity a.s., such that  $X_{\tau_n} \geq 0$  this time. Since  $\mathbb{P}(G_{\tau_n,\infty} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_{\tau_n}) > \delta_2$ , an application of Borel–Cantelli shows that  $\{X_n \geq \frac{a_1}{2} n^{\frac{1-2\gamma}{2}} \text{ i.o.}\}$  holds a.s. We claim that a.s. there are constants  $c(\omega) > 0$  and  $m(\omega)$  such that  $\{X_n > c \text{ for all } n \geq m\} = \{\liminf_{n \rightarrow \infty} X_n > 0\}$ . Indeed, if we define  $\tau_0 = 0$  and  $\tau_{n+1} = \inf\{m > \tau_n + 1 \mid X_m \geq \frac{a_1}{2} m^{\frac{1-2\gamma}{2}}\}$ , we see that  $\tau_n < \infty$  a.s. and  $\tau_n \rightarrow \infty$ . This gives a corresponding filtration, namely  $\mathcal{F}_n = \sigma(\tau_n)$ .

To finish the claim, we show that  $A = \{\liminf_{n \rightarrow \infty} X_n \leq 0\}$  has probability zero. To do so, it is sufficient to argue that there is a  $\delta$  such that  $\mathbb{P}(A \mid \mathcal{F}_n) < 1 - \delta$  a.s. for all  $n \geq 1$ . This is immediate from the previous calculation. Indeed,

$$\begin{aligned} \mathbb{P}(A \mid \mathcal{F}_n) &\leq 1 - \mathbb{P}\left(\liminf_n X_n \geq \frac{3a_1}{2} \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) \\ &= 1 - \mathbb{P}\left(\liminf_n X_n - \frac{a_1}{2} \tau_n^{\frac{1-2\gamma}{2}} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) \\ &\leq 1 - \mathbb{P}\left(\liminf_n G_{\tau_n,n} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} \mid \mathcal{F}_n\right) \\ &< 1 - \delta_2. \end{aligned}$$

The process  $G_{m,\infty}$  is a.s. finite, and since the drift term  $\sum_{i \geq n} \frac{|X_i|^k}{i^\gamma} \rightarrow \infty$ , we get that  $X_n \rightarrow \infty$ .

*Proof of Theorem 1.4.* Define  $\tau = \inf\{n|X_n \in (-\epsilon, \epsilon)\}$  and  $\tau' = \inf\{n > \tau|X_n \notin (-\epsilon', \epsilon')\}$ , where  $\epsilon < \epsilon'$ . When  $\epsilon$  is small enough, we may assume that  $\tau < \infty$  with positive probability; otherwise we have nothing to prove. On  $\{\tau < \infty\}$ , couple  $X_n$  with  $X'_n$  so that  $\mathbb{P}(X_n = X'_n, \tau \leq n \leq \tau' | [\tau < \infty]) = 1$ , where  $X'_n$  is a process that solves (29). Since  $X'_n \rightarrow \infty$  a.s., we have that  $\tau' < \infty$  a.s. Thus, on  $\{\lim_{n \rightarrow \infty} X_n = 0\}$ , we have that  $\{X_n > \epsilon' \text{ i.o.}\}$ . Therefore,  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = 0) = 0$ .

**6.2. Analysis of  $X_t$  when  $\frac{1}{2} + \frac{1}{2k} < \gamma, k > 1$ , and  $\gamma \in (1/2, 1)$**

Before proving the main Theorem 1.5, as described in Section 1.2 we will study a process  $(X_n)_{n \geq 1}$  that satisfies

$$X_{n+1} - X_n \leq \frac{f(x)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \quad \gamma \in (1/2, 1), k \in (1, \infty), \tag{35}$$

where  $0 < f(x) \leq |x|^k$  when  $x \in (-\epsilon, \epsilon)$ , and  $f(x) = |x|^k$  when  $x \in \mathbb{R} \setminus (-\epsilon, \epsilon)$ .

The analysis will again rely on studying the process  $Z_n = -\frac{X_n}{h(n)}$ , where  $h(t) = -t^{\frac{1-\gamma}{1-k}}$ . An important quantity related to the process  $Z_n$  will be  $G'_{s,n} = \sum_{i=s}^{n-1} \frac{Y_{i+1}}{i^\gamma h(i+1)}$ .

Recall that the  $Y_n$  constitute martingale differences satisfying  $|Y_n| < M$  almost surely. Furthermore, we find  $x_0 < 0$  such that  $f(x) > M$  for every  $x \leq x_0$ . We will make use of  $x_0$  in the next lemma.

**Lemma 6.4.** *Take  $C = \max(M, |X_1|, |x_0|)$ . Then  $X_n > -2C$  for all  $n$ , a.s.*

*Proof.* We can show this by induction. Of course  $X_1 > -2C$ . For the inductive step, we distinguish two cases. First, assume that  $-2C < X_n < -C$ . Then

$$\begin{aligned} X_{n+1} &= X_n + \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma} \\ &\geq -2C + \frac{f(X_n)}{n^\gamma} - \frac{M}{n^\gamma} \\ &> -2C. \end{aligned}$$

Now, assume that  $X_n \geq -C$ . Then

$$\begin{aligned} X_{n+1} &= X_n + \frac{|X_n|^k}{n^\gamma} + \frac{Y_n}{n^\gamma} \\ &\geq -C + 0 - \frac{M}{n^\gamma} \\ &> -2C. \end{aligned}$$

Pick  $\epsilon > 0$  such that

$$\epsilon \leq \min \left( \frac{1}{4}, \frac{1}{2} \left( \frac{1-\gamma}{3(k-1)} \right)^{\frac{1}{k-1}} \right).$$

Let  $a_n$  be defined as in Section 6.1, first appearing in (31) and defined in (33).

**Claim:** *We can find  $n_0$  that satisfies the following properties:*

1.  $a_n > 1/2$  for  $n \geq n_0$  a.s.



- 2. If  $-\frac{X_{n+1}}{h(n+1)} > -2\epsilon$  and  $-\frac{X_n}{h(n)} \leq -2\epsilon$ , then  $-\frac{X_{n+1}}{h(n+1)} < -\epsilon$ , when  $n \geq n_0$ .
- 3.  $\mathbb{P}(G'_{n_0,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq n_0 | \mathcal{F}_{n_0}) > 0$ .

*Proof.*

- 1. This is trivial.
- 2. Recall that  $h(t) = -t^{\frac{1-\gamma}{1-k}}$ . Since  $|Y_n| < M$ , and  $X_n > -2C$  a.s., whenever  $X_n < 0$  we have  $|X_{n+1} - X_n| = O(n^{-\gamma})$ . Also,  $n^{-\gamma} = o(h(n))$ , since  $\gamma > \frac{1-\gamma}{k-1}$ . Indeed,  $\gamma > \frac{1-\gamma}{k-1}$  is equivalent to  $\gamma > 1/k = 1/2k + 1/2k$ ; however,  $1/2 > 1/2k$ , and since  $\gamma > 1/2 + 1/2k$ , we conclude. Furthermore, notice that  $\frac{h(n)}{h(n+1)} \rightarrow 1$ . Calculate

$$\begin{aligned}
 -\frac{X_{n+1}}{h(n+1)} &= -\frac{X_{n+1} - X_n}{h(n+1)} - \frac{X_n}{h(n)} \cdot \frac{h(n)}{h(n+1)} \\
 &\geq o(1) - 2\epsilon \frac{h(n)}{h(n+1)}.
 \end{aligned}$$

Since the  $o(1)$  term and  $\frac{h(n)}{h(n+1)}$  depend only on  $n$ , we conclude Part 2.

- 3. Using the inequality (17) we find that  $\frac{1-\gamma}{k-1} - \gamma < \frac{-1-\delta}{2}$  for some  $\delta > 0$ , so

$$(m^\gamma h(m+1))^{-1} \sim m^{\frac{1-\gamma}{k-1} - \gamma} \leq m^{\frac{1-\gamma}{k-1} - \gamma} \leq m^{\frac{-1-\delta}{2}}.$$

Therefore, by Doob’s inequality we have

$$\begin{aligned}
 \mathbb{P}\left(\sup_{u \geq n_0} (G'_u)^2 \geq \frac{\epsilon^2}{4}\right) &\leq \sum_{m \geq n_0} \frac{\mathbb{E}(Y_{m+1}^2 | \mathcal{F}_{n_0})}{m^{2\gamma} h^2(m+1)} \\
 &\leq C \sum_{m \geq n_0} \frac{1}{m^{2\gamma} h^2(m+1)} \\
 &= \sum_{m \geq n_0} \Theta\left(m^{2\left(\frac{1-\gamma}{k-1} - \gamma\right)}\right) \\
 &= \sum_{m \geq n_0} \Theta\left(m^{-1-\delta}\right) \\
 &= \Theta\left(n_0^{-\delta}\right) \rightarrow 0.
 \end{aligned}$$

Notice that the previous claim holds for any stopping time  $\tau$ , in place of  $n$ . So we obtain a version of the previous lemma for stopping times.

**Lemma 6.5.** *Let  $\tau$  be a stopping time such that  $\tau \geq n_0$ , where  $n_0$  is the same as in the previous claim. Then  $\mathbb{P}(G'_{\tau,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq \tau | \mathcal{F}_\tau) > 0$ .*

Let

$$\epsilon \leq \min\left(\frac{1}{4}, \frac{1}{2} \left(\frac{1-\gamma}{3(k-1)}\right)^{\frac{1}{k-1}}\right),$$

and define a stopping time  $\tau = \inf\{n \geq n_0 | Z_n < -2\epsilon\}$ .

**Proposition 6.2** Let  $(X_n)_{n \geq 1}$  satisfy (35). If  $\tau < \infty$  holds with positive probability, then  $\mathbb{P}(X_n \rightarrow 0) > 0$ . More specifically, the process  $(X_n : n \geq \tau)$  converges to zero with positive probability.

*Proof.* On the event  $\{X_m < 0 \text{ for all } m \geq n\}$  we use the expression for  $Z_n = -\frac{X_n}{h(n)}$  and obtain, as in (31) (except now the inequalities are reversed),

$$\begin{aligned} Z_{n+1} - Z_n &\leq \frac{X_n}{h(n+1)n^\gamma} \left( -a_n \frac{1-\gamma}{k-1} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} \\ &< \frac{X_n}{h(n+1)n^\gamma} \left( -\frac{1-\gamma}{2(k-1)} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)}. \end{aligned}$$

Set

$$D_n = \frac{X_n}{h(n+1)n^\gamma} \left( -\frac{1-\gamma}{2(k-1)} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right).$$

Then we have

$$Z_m - Z_\tau \leq \sum_{i=\tau}^{m-1} D_i + G'_{\tau,m}. \tag{36}$$

Now we will show by contradiction that on the event  $A = \{G'_{\tau,n} \in (-\frac{\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq \tau\}$  the process satisfies  $X_n < 0$  for all  $n \geq \tau$ . Define  $\tau_0 = \inf\{n \geq \tau | Z_n \geq 0\}$  and  $\sigma = \sup\{\tau \leq n < \tau_0 | Z_{n-1} \leq -2\epsilon, Z_n > -2\epsilon\}$ . Also, when  $Z_n \geq -2\epsilon$  we have  $X_n \geq 2\epsilon h(n) = -2\epsilon n^{\frac{1-\gamma}{1-k}}$ . So  $\frac{|X_n|^k}{X_n} \geq -(2\epsilon)^{k-1} n^{-1+\gamma}$ . Therefore, by the definition of  $\epsilon$ , we get

$$-\frac{1-\gamma}{2(k-1)} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} < \left( -\frac{1-\gamma}{2(k-1)} + \frac{1-\gamma}{3(k-1)} \right) n^{-1+\gamma} = -\frac{1-\gamma}{6(k-1)} n^{-1+\gamma} < 0.$$

Hence  $D_n < 0$  whenever  $Z_n \geq -2\epsilon$ . If  $\{\tau_0 < \infty\} \cap A$  has positive probability, then  $\{\sigma < \infty\} \cap A$  does also. Thus, on  $\{\tau_0 < \infty\} \cap A$ ,

$$\begin{aligned} 0 \leq Z_{\tau_0} &= Z_\tau + \sum_{i=\tau}^{\tau_0-1} D_i + G'_{\tau,\tau_0} \\ &= Z_\tau - Z_\sigma + Z_\sigma + \sum_{i=\tau}^{\tau_0-1} D_i + G'_{\tau,\tau_0} \\ &= Z_\sigma - G'_{\tau,\sigma} + G'_{\tau,\tau_0} + \sum_{i=\sigma}^{\tau_0-1} D_i \\ &< -\epsilon + \frac{\epsilon}{2} + \frac{\epsilon}{2} + 0 = 0, \end{aligned}$$

which is a contradiction.

Now we can complete the proof of the proposition. On the event  $A$  the process satisfies  $X_n < 0$  for all  $n > \tau$ ; therefore  $\limsup_{n \rightarrow \infty} X_n \leq 0$  on  $A$ . However, by Lemma 2.4 we have  $\limsup_{n \rightarrow \infty} X_n \geq 0$  a.s. Therefore, on  $A$ ,  $X_n \rightarrow 0$ .

**Remark 6.1** On  $A$  we showed that  $X_n$  converges to zero, since for all  $n \geq \tau$ ,  $X_n < 0$ , and the only place to converge is the origin.

*Proof of Theorem 1.5.* We define  $\tau = \inf\{n \geq n_0 | X_n \in (-\epsilon_2, -\epsilon_1)\}$ , where  $n_0$  is the same as in Lemma 6.5, and  $\tau_e = \inf\{n | X_n \notin (-3\epsilon, 3\epsilon)\}$ . Let  $(X'_n : n \geq \tau)$  be a process that satisfies (35). Then we couple  $(X_n)$  with  $(X'_n)$  on  $\{\tau < \infty\}$  so that  $\mathbb{P}(X_n = X'_n, \tau \leq n \leq \tau_e | \{\tau < \infty\}) = 1$ . To show that  $X'_n$  converges to zero with positive probability, first we need to verify that the conditions for Proposition 6.2, are met. The only thing we need to check is that  $Z'_\tau = -\frac{X'_\tau}{h(\tau)} < -2\epsilon$ . However, since  $h(t) \rightarrow 0$  this is always possible by choosing  $n_0$  large enough. Furthermore, by Proposition 6.2, we see that there is an event of positive probability such that  $X'_n \rightarrow 0$ , where  $\tau_e$  is infinite conditioned on this event. Therefore,  $X_n$  converges to 0 with positive probability.

### Acknowledgements

The author would like to thank Marcus Michelen, Albert Chen, and Josh Rosenberg for constructive criticism of the manuscript.

### References

- [AAZB+17] AGARWAL, N. et al. (2017). Finding approximate local minima faster than gradient descent. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, Association for Computing Machinery, New York, pp. 1195–1199.
- [AG16] ANANDKUMAR, A. AND GE, R. (2016). Efficient approaches for escaping higher order saddle points in non-convex optimization. In *Proceedings of the 29th Annual Conference on Learning Theory* (Proceedings of Machine Learning Research **49**), PMLR, New York, pp. 81–102.
- [BR95] BRENNAN, R. W. AND ROGERS, P. (1995). Stochastic optimization applied to a manufacturing system operation problem. In *Proceedings of the 27th Conference on Winter Simulation*, IEEE Computer Society, Washington, DC, pp. 857–864.
- [CdITZ16] CHEN, X., LEE, J. D., TONG, X. T. AND ZHANG, Y. (2016). Statistical inference for model parameters in stochastic gradient descent. *Ann. Statist.* **48**, 251–273.
- [CHM+15] CHOROMANSKA, A. et al. (2015). The loss surfaces of multilayer networks. *J. Mach. Learn. Res.* **38**, 192–204.
- [DKLH18] DANESHMAND, H., KOHLER, J., LUCCHI, A. AND HOFMANN, T. (2018). Escaping saddles with stochastic gradients. In *Proceedings of the 35th International Conference on Machine Learning* (Proceedings of Machine Learning Research **80**), PMLR, Stockholm, pp. 1155–1164.
- [Dur13] DURRETT, R. (2013). *Probability: Theory and Examples*. Duxbury Press, Belmont, CA.
- [EDM01] EVEN-DAR, E. AND MANSOUR, Y. (2001). Learning rates for q-learning. *J. Mach. Learn. Res.* **5**, 1–25, 2001.
- [Fis92] FISHER, E. (1992). On the law of the iterated logarithm for martingales. *Ann. Prob.* **20**, 675–680.
- [GHJY15] GE, R., HUANG, F., JIN, C. AND YUAN, Y. (2015). Escaping from saddle points: online stochastic gradient for tensor decomposition. In *Proceedings of the 28th Conference on Learning Theory* (Proceedings of Machine Learning Research **40**), PMLR, Paris, pp. 797–842.
- [GM91] GELFAND, S. B. AND MITTER, S. K. (1991). Recursive stochastic algorithms for global optimization in  $\mathbb{R}^d$ . *SIAM J. Control Optimization* **29**, 999–1018.
- [HLS80] HILL, B. M., LANE, D. AND SUDDERTH, W. (1980). A strong law for some generalized urn processes. *Ann. Prob.* **8**, 214–226.
- [JGN<sup>+</sup>17] JIN, C. et al. (2017). How to escape saddle points efficiently. Preprint. Available at <http://arxiv.org/abs/1703.00887>.
- [JKN15] JAIN, P., JIN, C., KAKADE, S. M. AND NETRAPALLI, P. (2015). Computing matrix squareroot via non convex local search. Preprint. Available at <http://arxiv.org/abs/1507.05854>.
- [KY03] KUSHNER, H. AND YIN, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York.
- [LLKC18] LI, T., LIU, L., KYRILLIDIS, A. AND CARAMANIS, C. (2018). Statistical inference using SGD. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, AAAI Press, Palo Alto, CA, pp. 3571–3578.

- [LSJR16] LEE, J. D., SIMCHOWITZ, M., JORDAN, M. I. AND RECHT, B. (2016). Gradient descent converges to minimizers. Preprint. Available at <http://arxiv.org/abs/1602.04915>.
- [Spr] Łojasiewicz inequality. *Encyclopedia of Mathematics*. Website, accessed 15 September 2019. Available at [https://www.encyclopediaofmath.org/index.php/Łojasiewicz\\_inequality](https://www.encyclopediaofmath.org/index.php/Łojasiewicz_inequality).
- [Pem90] PEMANTLE, R. (1990). Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Prob.* **18**, 698–712.
- [Pem91] PEMANTLE, R. (1991). When are touchpoints limits for generalized Pólya urns? *Proc. Amer. Math. Soc.* **113**, 235–243.
- [Pem07] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Prob. Surveys* **4**, 1–79.
- [PJ92] POLYAK, B. T. AND JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optimization* **30**, 838–855.
- [RM51] ROBBINS, H. AND MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22**, 400–407.
- [RRT17] RAGINSKY, M., RAKHLIN, A. AND TELGARSKY, M. (2017). Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. Preprint. Available at <http://arxiv.org/abs/1702.03849>.
- [RSS12] RAKHLIN, A., SHAMIR, O. AND SRIDHARAN, K. (2012). Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Conference on Machine Learning*, Omnipress, pp. 1571–1578.
- [Rup88] RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process. Tech. Rep., Cornell University Operations Research and Industrial Engineering.
- [RWW87] ROGERS, L. C. G. AND WILLIAMS, D. (1987). *Diffusions, Markov Processes and Martingales, Vol. 2: Itô Calculus*. John Wiley, New York.
- [SL87] SURI, R. AND LEUNG, Y. T. (1987). Single run optimization of a SIMAN model for closed loop flexible assembly systems. In *Proceedings of the 19th Conference on Winter Simulation*, Association for Computing Machinery, New York, pp. 738–748.
- [SL16] SUN, R. AND LUO, Z. Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inf. Theory* **62**, 6535–6579.
- [Son12] SON, P. T. (2012). An explicit bound for the Łojasiewicz exponent of real polynomials. *Kodai Math. J.* **35**, 311–319.
- [SQW17] SUN, J., QU, Q. AND WRIGHT, J. (2017). Complete dictionary recovery over the sphere I: overview and the geometric picture. *IEEE Trans. Inf. Theory* **63**, 853–884.