CERTAIN SYSTEMS ARISING IN STOCHASTIC GRADIENT DESCENT

Konstantinos Karatapanis

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Robin Pemantle, Professor of Mathematics

Acting Graduate Group Chairperson

David Harbater, Professor of Mathematics

Dissertation Committee:
Robin Pemantle, Professor of Mathematics
Philip Gressman, Professor of Mathematics
Henry Towsner, Associate Professor of Mathematics

Acknowledgments

I would like to thank my advisor Robin Pemantle for his continuous support, throughout my years at Penn. He was a teacher to me not only in math with his knowledge but also in life with him kind character.

Also I wish to thank my friends at Penn for supporting me. Among them I wish to especially thank Albert Chenchao, and Giorgos Gounaris for the stimulating conversations about my research.

Lastly, I owe a big thanks to my family for always being there for me, and a special thanks to my grandmother.

ABSTRACT

CERTAIN SYSTEMS ARISING IN STOCHASTIC GRADIENT DESCENT

Konstantinos Karatapanis

Robin Pemantle

Stochastic approximations is a rich branch of probability theory and has a wide range of application. Here we study stochastic approximations from the perspective of gradient descent. An important question is what is the asymptotic limit of a stochastic approximation. In that spirit we will provide a detailed description for the limiting behavior of certain one dimensional stochastic approximations.

Contents

Chapter 1

Introduction

1.1 Introduction

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

$$X_{n+1} = X_n + a_n \left(F(X_n) + \xi_{n+1} \right). \tag{1.1.1}$$

To elaborate on the parameters, let $\mathscr{F}_n$ be a filtration, then a_n, ξ_n are adapted, and ξ_n constitute martingale differences, i.e., $E(\xi_{n+1}|\mathscr{F}_n) = 0$. The sequence a_n can be deterministic or stochastic, and the sequence is assumed to be positive almost surely and is either converging to zero or staying constant, for a discussion on the different possibilities, see section 1.2. Also, when the dynamics of F are complicated, we usually require some additional assumptions on the the noise; one is a boundedness restraint in that we assume the existence of a constant M, such that $|\xi_n| \leq M$ a.s.; and secondly, we want ξ_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. However, in many instances more relaxed conditions on ξ_n are enough. This versatile system is well-studied, and it arises naturally in many different areas. In machine learning and statistics, (1.1.1) can be a powerful tool used for quick optimization and statistical inference (see [AAZB+17], [LLKC18], [CdlTTZ16]), among other uses. Furthermore, many urn models are represented by (1.1.1). These processes play 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].

Processes satisfying (1.1.1), when a_n goes to zero, are known as stochastic approximations, first introduced in [RM51]. These processes have been extensively studied since [KY03]. An important feature is that the step size a_n satisfies

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

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 ODE whose right-hand side corresponds to the expectation of the driving term $(F(X_t))$.

The previous heuristic can help us identify the support S of the limiting process $X_\infty := \lim_{n\to\infty} X_n$ in terms of the topological properties of the dynamical system $\frac{\mathrm{d}X_t}{\mathrm{d}t} = F(X_t)$ (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] chapter 5.8). However, there has not been a systematic approach 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, they 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 (1.1.1), where a_n is of the order of n^{-1}. Focused effort is being placed in understanding the support of the limiting process X_∞. In specific instances, the underlying problem boils down to understanding an SGD problem: characterizing the support of X_∞ in terms of the set of critical points of the corresponding potential V. For a comprehensive exposition on urn processes, see [Pem07].

1.2 Step sizes

In the literature of stochastic approximations, there is are ample choices for the behavior of the step sizes depending on the application at hand. Also, in different contexts, the sequence a_n as in recursion (1.1.1) can appear with different names. For example it is known as learning rate (machine learning), smoothing constant (forecasting) or gain (signal processing).

1. In [EDM04] they study the rate of convergence for polynomial step sizes, i.e., $a_n = n^{-\gamma}$ where $\gamma \in (1/2, 1)$ in the context of Q-learning for Markov decision processes, and they experimentally demonstrate that for γ, approximately $\frac{17}{20}$ the rate of convergence is optimal.

2. Kesten algorithm [Kes58] introduced a stochastic approximation process in hopes of accelerating the convergence of the Robbins-Monro algorithm [RM51]. The idea here is: when we are confident that the process is close to the value θ we wish to estimate, we decrease the step size in order to stabilize the convergence. And whenever we suspect we are far away from θ, we keep the step size large in order to allow faster exploration. In order to determine when it is warranted to decrease the step size, they followed the heuristic that when the process is close to θ then the sign of $X_n - X_{n-1}$ should fluctuate more intensely, as the process will keep overshooting. More formally, suppose that we have a sequence $a_n \geq 0$ such that $\sum_{n=1} a_n = \infty$ and $\sum_{n=1} a_n^2 < \infty$. Then

we define a stochastic approximation whose step size is $b_1 = a_1$, $b_2 = a_2$ and $b_n = a_{k_n}$

$$k_{n+1} = \begin{cases} k_n & \text{if } (X_n - X_{n-1})(X_{n-1} - X_{n-2}) > 0 \\ k_n + 1 & \text{if } (X_n - X_{n-1})(X_{n-1} - X_{n-2}) \leq 0 \end{cases}$$

3. In Gaivoronski [EW88], we see another variation of the Robbins-Monro algorithm where the step size is decreasing when $\frac{|X_n - X_{n-k}|}{\sum_{i=n-k}^{n-1} |X_{i+1} - X_i|} \leq \tilde{\gamma}$ for some threshold value $\tilde{\gamma}$, and otherwise it stays the same. The intuition on why this works is that the quantity $\frac{|X_n - X_{n-k}|}{\sum_{i=n-k}^{n-1} |X_{i+1} - X_i|} \leq \tilde{\gamma}$ is small when the iterates are fluctuating, and this is likely to happen when you are in the proximity of the minimum. Conversely, when the quantity $\frac{|X_n - X_{n-k}|}{\sum_{i=n-k}^{n-1} |X_{i+1} - X_i|}$ is big, the process is likely to be far away from the minimum.

4. In [SR96], they study the adaptive behavior of a lizard. In this model, the lizard wants to maximize the reward per unit of time in the long run. The lizard is at its home, unless it is hunting, and at some random time τ_n an ant appears whose weight is w_n. If the lizard is at home and an ant appears, it decides whether to go after the ant or not. If the lizard goes after the ant, it will catch it with some probability and it will take the lizard r_n to return to its home, regardless of whether he has successfully captured its prey or not. More formally, the lizard decides, based on past observations, to chase after the ant having in mind to maximize T_n/W_n the ratio of the total time it has

taken him to successfully have caught ants whose total weight amounts to W_n. This problem can be formulated as a stochastic approximation, whose step size is W_n^{-1}. The purpose of this research is to compare whether the strategy utilized by the lizard in this learning problem, will match the strategy that would arise via different optimization techniques; for example via the "return for effort" principle.

5. In [PJ92] and [Rup19] the step sizes are of polynomial order $a_n = n^{-\gamma}$. Here, they focus on studying an average of the iterates $S_n = \frac{\sum_{i=1}^{n} X_i}{a_n}$ properly scaled so that S_n converges in distribution to a normal. The optimal choice for γ is shown be less than 1, indicating that if you allow more fluctuations then by averaging you obtain a better estimate for the true value of the parameter. The point estimation for Robbins-Monro can be shown to be optimal for step size $a_n = \frac{1}{n}$ in terms of minimizing the square mean error at comparable rates to the averaging scheme. Even though we do not improve the performance in the long run, the averaging scheme is preferable since the larger step sizes increase robustness because in the early stages it is favorable to increase the rate of exploration.

1.3 Stochastic gradient descent

In machine learning, processes satisfying (1.1.1) 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

$$x_{n+1} - x_n = -\eta_n \nabla V(x_n), \tag{1.3.1}$$

where in many applications we take η_n to be a positive and constant. Notice that (1.3.1) is a specialization of (1.1.1), when $F = -\nabla V$, $\xi_{n+1} \equiv 0$ and $a_n = \eta_n$. The previous 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 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] chapter 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 because in other cases local minima provide equally good results as global minima [CHM$^+$15]. Furthermore, in certain applications saddle points lead to highly sub-optimal results ([JJKN15] and [SL16]), which highlights the importance of escaping saddle points.

An important variant of GD/SGD is the momentum GD/SGD, firstly introduced by Polyak in [Pol64]. This algorithm diminishes the effects of the gradient in directions where the iterates are fluctuating, and it increases the movement along stable directions by accumulating momentum. In the literature the momentum algorithm appears in two popular formats,

$$v_t = bv_{t-1} + \eta \nabla V(x_t)$$
$$x_t = x_{t-1} - v_{t-1}$$

and

$$v_t = bv_{t-1} + (1-\beta) \nabla V(x_t)$$
$$x_t = x_{t-1} - \eta v_{t-1}$$

The parameter $\beta \in (0,1)$, plays an important role in the performance. In practice, usually β is chosen around 0.9, but there is no fast and hard rule. SGD has difficulty navigating along ravines, and in such instances adding momentum can significantly improve performance.

A small variation of the previous algorithm, but which significantly improves performance is the accelerated gradient descent, or the look-ahead momentum gradient descent [NES83].

$$v_t = bv_{t-1} + \eta \nabla V(x_t - bv_{t-1})$$

$$x_t = x_{t-1} - v_{t-1}$$

Here, in order to make the transition from x_{t-1} to x_t, we incorporate into the gradient the quantity $-bv_{t-1}$ which is a good predictor on where the x_{t-1} will land, hence further improving the performance and giving a convergence rate of $\mathrm{O}(1/t^2)$ after t iterations.

One shortcoming of the momentum algorithms already described is that we have to guess the value of the parameter b as the these algorithms do not have a way to auto-tune. Some variations of SGD that try to ameliorate this are Adagard [DHS11], and Adadelta [DZ12]. For an overview for SGD algorithms with a neural network perspective, see [Rud16].

1.4 Stochastic approximations

1.4.1 Historic development

In 1951, R. Monro gave birth to stochastic approximation with his work [RM51]. Time proved that his ideas were very fruitful, and since then the theory has flourished. Suppose we perform an experiment at level $x_n \in [0, 1]$ giving rise to a random variable $\xi(\omega, x_n)$, whose distribution is unknown depending on x_n, measuring the response of the experiment. Here, $E(\xi(\omega, x_n))$ is again unknown, however we know it is increasing in its second coordinate. We wish to find the level $\tilde{x}$ such that $E(\xi(\omega, \tilde{x})) = a$, where a is a given constant. This should be done by establishing a recursive rule on how, given the past observations, we can determine the next level x_{n+1} for the experiment so that $x_n \to \theta$ in some sense. This can be accomplished by the recursion $x_{n+1} - x_n = \frac{1}{n}(a - \xi(\omega, x_n))$. To make the connection to (1.1.1), $a_n = \frac{1}{n}$, $F(x_n) = a - E(\xi(\omega, x_n))$ and $\xi_{n+1} = E(\xi(\omega, x_n)) - \xi(\omega, x_n)$. One of the main theorems is that when $E(\xi(\omega, x))$ is differentiable such that $\frac{\partial E(\xi(\omega,\tilde{x}))}{\partial x} > 0$ then $x_n \to \theta$ in probability. The step size factor $\frac{1}{n}$ is chosen appropriately, so that there is an implicit averaging that eliminates the effects of the noise, eventually. The previously described system hence behaves similarly to the ODE whose right-hand side corresponds to the expectation of the driving term in the sense that their limiting points coincide.

Later Kiefer [KW52], in a short paper, relaxed the conditions on $M(x) = E(\xi(\omega, x))$. In the following years Julius R. Blum first established that the convergence in the R. Monro model is almost surely, and then later on he proved the multidimensional analogue [RB54].

In an effort to improve the rate of convergence, in the multidimensional setting, and make it more applicable in the field of statistics in [PJ92] they consider an average of the iterates, i.e., $S_n = \sum_{i=1}^{n} x_i$. Then they show that if S_n is averaged properly then, it converges in distribution to a normal random variable. In that way it is possible to perform statistical testing like confidence intervals or hypothesis testing. Next, we provide a motivating example:

Example 1. *In paper [HLS80], they consider a random variable X_n taking values in $(0,1)$, which we interpret as counting the percentage of the red balls out of n balls in accordance to Polya's urn model. Recursively define X_{n+1} to be $\dfrac{nX_n+1}{n+1}$ with probability $f(X_n)$, and $X_{n+1} = \dfrac{nX_n}{n+1}$ with probability $1-f(X_n)$. The main result of [HLS80] is that X_n converges to a random variable X, whose range is a subset of $C = \{p|f(p) = p\}$; moreover for all points p such that $f'(p) < 1$ or ($f'(p) > 1$), we have $\mathbb{P}(X = p) > 0$ ($\mathbb{P}(X = p) = 0$).*

This process fits the general form (1.1.1). Indeed, we may rewrite X_n in the following form $X_{n+1} - X_n = A_n + Y_n$ where Y_n is the martingale

$$Y_n = \begin{cases} \frac{1-f(X_n)}{n+1}, & with\, probability\, f(X_n) \\ \\ \frac{-f(X_n)}{n+1}, & with\, probability\, 1-f(X_n) \end{cases}, \text{ and } A_n = \tfrac{f(X_n)-X_n}{n+1}.$$

Define $g_n = \begin{cases} 1-f(X_n), & with\, probability\, f(X_n) \\ -f(X_n), & with\, probability\, 1-f(X_n) \end{cases}.$

Then, the SDE becomes $X_{n+1} - X_n = \frac{f(X_n)-X_n}{n+1} + \dfrac{g_n}{n+1} = \frac{f(X_n)-X_n}{n+1} + \dfrac{\Theta(1)}{n+1}$, *when $f(X_n)$ is bounded away from $\{0,1\}$. We have already mentioned that X_n can only converge to points p, such that $f(p) = p$, $f'(p) < 1$. The idea is that the* *condition $f'(p) < 1$ implies that $f(X_n) - X_n$ is positive when $X_n \in (p - \delta, p)$ and negative when $X_n \in (p, p + \delta)$. Therefore, A_n pushes X_n towards p, when X_n lies in a neighborhood of p, and since $|X_{n+1} - X_n| = O(1/n)$, the process $(X_n)_{n \geq 0}$ may eventually get trapped in the neighborhood. Consequentially, as p is the sole point in the neighborhood that belongs in C, the convergence follows.*

1.4.2 Convergence in one dimension

After the result of [HLS80], in recent years, the corresponding proofs have been increasingly simplified using martingale techniques. Now, we offer a summary of some of the most fundamental results taking place in the one dimensional setting. Most of what follows is covered in [Pem07]. We define the following recursion,

$$X_{n+1} - X_n = a_n F(X_n) + a_n \xi_{n+1} + a_n R_n, \tag{1.4.1}$$

where ξ_i are martingale differences and R_n is a predictable process, that represents some approximation error or bias, depending on the context. The step sizes a_n are positive and satisfy $\sum_{n=1}^{\infty} a_n = \infty$ and $\sum_{n=1}^{\infty} a_n^2 < \infty$, the accumulated error is summable, i.e., $\tilde{R}_n = \sum_{n=1}^{\infty} a_n |R_n| < \infty$. We will borrow some constraints appearing in Urn processes and henceforth assume $X_n \in [0, 1]$ and $F \in [0, 1]$. The interval $(0, 1)$ and the previous constrains, altogether, are chosen for illustrative purposes. Next, we will see that the support of the limiting process is supported on the zero set of F,i.e., $X_\infty := \lim_{n \to \infty} X_n \in \{p | F(p) = 0\}$.

Theorem 1.1. *Suppose that X_n solve* (1.4.1)*. Also, $E(\xi_{n+1}^2|\mathcal{F}_n) \leq M$, for some positive constant M. If $F > \delta$ or $F < -\delta$ on $[a,b] \subset (0,1)$, then for any closed interval $I \subset [a,b]$ we have $\mathbb{P}(X_\infty \in I) = 0$.*

To see why this is true, assume that X_n eventually stays inside I, then because $\sum_{n=1}^{\infty} a_n = \infty$ the iterates X_n, would have to travel along a path of infinite length, which contradicts the fact that there is a finite amount of noise in the system.

A direct corollary of the previous result is that $X_\infty \in \{p|F(p) = 0\}$, when F is continuous, as the sets $\{F > \frac{1}{n}\}$ can be written as a countable union of open sets.

Corollary 1.2. *Suppose that X_n solve* (1.4.1)*. Also, $E(\xi_{n+1}^2|\mathcal{F}_n) \leq M$, for some positive constant M. If F is a continuous function then $X_\infty \subset \{p|F(p) = 0\}$.*

The next two results provide a more detailed description for the support of X_∞ as a subset of $\{p|F(p) = 0\}$.

Theorem 1.3. *Suppose that X_n solve* (1.4.1)*. Let $p \in (0,1)$, and assume that $F < 0$ on $(p-\epsilon, p)$, $F > 0$ on $(p, p+\epsilon]$ and $F(p) = 0$. If X_n visits $(p-\epsilon, p+\epsilon)$ infinitely often then $\mathbb{P}(X_\infty = 0) > 0$.*

Lastly, we provide a non-convergence theorem,

Theorem 1.4. *Suppose that X_n solve* (1.4.1)*. Let $p \in (0,1)$, and assume that $F = sign(x-p)$ on $(p-\epsilon, p+\epsilon)$ then $\mathbb{P}(X_\infty = p) = 0$.*

We interpret the previous results in the context of SGD. In that setting corollary 1.2 says that X_∞ is supported on the stationary points of the corresponding potential function V such that $V' = F$. Theorem 1.3 says that a local minimum for V, when visited infinitely often, is a point where SGD may convergent to. And finally, Theorem 1.4 argues that a local maximum is always avoided.

1.4.3 Convergence in multiple dimensions

Here, we will discuss the ODE-method for a more detailed exposition, see [KY03], which is used to establish convergence results for stochastic approximations. This method links the asymptotic behavior of the discrete process to the autonomous continuous dynamical system. It can be shown that there is a sequence of continuous approximation of the tail of the discrete process, that converges to the continuous dynamical system. Subsequently, under certain conditions imposed on the discrete process, our knowledge about the asymptotic behavior of the deterministic continuous model transfers to the corresponding discrete stochastic approximation. All this will be made more precise after we have laid out the necessary apparatus. The main results, in this section, will be a convergence and a non-convergence result Theorem 1.7, and [Pem90] respectively. We will provide a sketch for the Theorem 1.7 and the rest of the results will be quoted.

The main analytic concept for this method is equicontinuity. We give the definition

Definition 1.5. *Let $f_n : \mathbb{R} \to \mathbb{R}$ be a family of measurable functions. Then $\{f_n\}$ is called equicontinuous in the extended sense if* $\sup |f_n(0)|$ *is bounded and for all*

$\delta > 0$

$$\limsup_{n\to\infty} \sup_{0\le|s-t|<\delta} |f_n(t) - f_n(s)| = 0$$

So, we have the following version of Arzela-Ascoli theorem.

Theorem 1.6. *Let $f_n : \mathbb{R} \to \mathbb{R}$ be a family of measurable functions. Then, there is a subsequence f_{k_n} such that f_{k_n} converges uniformly on a continuous function.*

Next, we will see an application of the following theorem to a deterministic model, which can serve as a template for establishing our main convergence result.

Example 2. *Suppose that $x'(t) = F(x(t))$ for a continuously differentiable function F. Define $x^n(t) = \sum_{m=0}^{\infty} 1_{\left[\frac{m}{n}\le t<\frac{m+1}{n}\right]} x\left(\frac{m}{n}\right)$. Then we can see that $\{x^n\}$ is an equicontinuous family of functions, that satisfy $x^n(t) = \int_0^t F(x^n(u))du + \rho^n(t)$. It is easy to see that $\rho^n(t) \to 0$ uniformly. Therefore, the limiting object for $x^n(t)$ satisfies $x'(t) = F(x(t))$.*

We define the associated continuous process for the stochastic approximations solving equation (1.1.1). Define $t_n = \sum_{i=1}^{n-1} a_i$ for $n \ge 1$. Set

$$X^1(t) = \begin{cases} X_1 & \text{for } t \le 0 \\ X_n & \text{when } t_n \le t < t_{n+1} \end{cases},$$

and for $n > 1$ define $X^n(t) = X^1(t_n + t)$. Now that we have set up the necessary terminology, we can state a general convergence theorem for processes that satisfy (1.1.1).

Theorem 1.7. *Suppose that X_n solve* (1.1.1)*. Furthermore, assume that $|X_n|$ is almost surely bounded. Then there is a null set N such that for all $\omega \notin N$, the sequence X^n is equicontinuous and the limits x'_t of convergent subsequences of X^n are trajectories of $x'_t = F(x_t)$ in some bounded invariant set and X_n converges to this invariant set. Furthermore, if F admits a potential functions $-\nabla V$, then X_n converges to a stationary component of V.*

The sketch of the proof consists of two main steps, the first step is to establish that the shifted interpolated process converges to the deterministic one. After establishing that one can argue that if the interpolated process did not behave as expected it would contradict the fact that its limit satisfies the stationary ODE.

For the first part one can argue that the tail of the martingale sequence is going to zero, using the Markov inequality. For the second part we characterize the limit points as follows. Because the of boundedness assumption and the fact that the iterates are bounded there must be a point inside the locally stable neighborhood S where the iterates are converging to. The deterministic system remains inside S, therefore the iterates eventually must stay in S, as well. Since $X^n(0) = X^1(t_n)$ is convergent, the iterates must converge to the limit set of S. When F arises from a potential function the limit sets are the stationary points. Furthermore the iterates must converge to a unique component since otherwise, as the difference of successive iterates is going to zero, there would exist a point, outside of the stationary component, where the iterates converged to.

In the previous theorem an important assumption was that the iterates were bounded. When the iterates are not constrained in a bounded domain the analysis is more complicated. In this case the trajectories can possibly be unbounded, and stability techniques, a prominent one is establishing the existence of a Liapunov function, (see [KY03] chapter 5.4) is one of the main ways forward. These methods can be used to establish that the iterates return to a bounded trajectory infinitely often, in which case asymptotic analysis techniques similar to Theorem 1.7 can be utilized.

The main ingredient for the proof of Theorem 1.7 was that the family of the interpolated process was equicontinuous. In, principle, one can demand minimal conditions for the step sizes and the stochastic approximation algorithm so that this property is met. An effort to do that can be seen in [KY03] chapter 5.4, producing results under weaker conditions. Furthermore, in a certain classes of problems the conditions specified there can be seen not only being necessary, but also sufficient.

We have established that the limit of a stochastic approximation is supported on distinct stationary components. An important next step for the development of the theory is the investigate which stationary points are limiting. Generally it is hard to establish whether a saddle point is in the support of X_∞, as the underlying dynamics can lead to very complicated saddle point structures. However, a very important result [Pem91], shows that when the linearization of $F(\cdot)$ as in (1.1.1), at a point p, has a positive eigenvalue value then p will never be in the support of X_∞.

1.4.4 Motivating example

Here we will revisit Example 1, and the follow up paper of Pemantle [Pem91] which provides a more detailed description for the support of X_∞. The analysis developed in section 1.4.1 establishes that the support of the limiting process is exactly the set of fixed points of f (critical points of the corresponding potential). More precisely, the iterates X_n will avoid local maxima with probability 1, and X_n will converge to a local minimum with some positive probability. However, at that point in history, it was unknown whether X_n can converge to saddle points. Later Pemantle, with his work [Pem91], settled this; giving explicit conditions, and surprisingly depending on the local behavior of f, the process may or may not converge there. Next, we will define a quantity which we will need in the next paragraph. Let $Z_{n,m} = \sum_{i=n}^{m-1} Y_i$, so $E(Z_{n,m}^2) \leq \sum_{i\geq n} \frac{1}{(i+1)^2} \sim \frac{1}{n}$. The last equation, after taking $m \to \infty$, is called the remaining variance for the process X_n, and it measures how much X_n can potentially deviate from the "mean flow" by the influence of future noise.

We will give the high level intuition, in qualitative terms, utilizing objects already described, namely the mean flow and the remaining variance. It is clear that the occurrence of convergence or non convergence to a point p, depends on the behavior of the process X_n when lying in the stable trajectory. Now, for simplicity, we assume the stable trajectory lies in a left neighborhood of p namely $(p-\delta, p)$, and recalling that p is a saddle point $(p, p+\delta)$ realizes the unstable trajectory. Consequently, assume X_n is moving towards p. The notion of the expected rate of convergence $\mathfrak{o}_1(n) := (X_n - p)$ can be explicitly computed via solving the mean flow differential equation. To continue further, as promised, we need to introduce $\mathfrak{o}_2(n) = \sqrt{E(Z^2_{n,\infty})}$ the order of the square root of the remaining variance. When $\mathfrak{o}_1(n) = o(\mathfrak{o}_2(n))$, in every instance where X_n behaves as expected, with h.p. X_n will be pushed, by the remaining noise, to the unstable trajectory i.e. $X_{n+k} \in (p, p+\delta)$ for some $k > 0$. Whenever this happens X_{n+k} may fail to return to $(p-\delta, p)$ with some positive fixed probability. Finally, by Borel-Cantelli the process will not converge to p with probability 1. Similarly, we can argue that when $\mathfrak{o}_2(n) = o(\mathfrak{o}_1(n))$, X_n will converge to p with some positive probability. To elaborate, the probability that X_n will escape the stable trajectory is decaying rapidly whence by Borel-Cantelli, in the event that X_n behaves as expected, the process will fail to visit the unstable trajectory, thereby establishing the convergence of X_n to p.

1.5 Non-convex optimization

Non-convex optimization problems are, generally, NP-hard (for a discussion in the context of escaping saddle points see [AG16]). The difficulty with high order saddle points can be seen from the fact that it is NP-hard to confirm when a polynomial of degree 4 is non-negative. We have the following theorem

Theorem 1.8. *It is NP-hard to determine whether a homogeneous polynomial of degree* 4 *is non-negative.*

For a reference see [Nes00] and [HL13]. This limitation does not mean that there could not exist algorithms that distinguish higher order degenerate saddle points efficiently when noise is added appropriately into the system. Which in turn could provide fast convergence to local minima with high probability.

So far, there has been a lot of effort finding fast converging SGD type of algorithms when assuming some non-degeneracy conditions on the Hessian matrix. Although there are results when the Hessian is degenerate, all the results require using knowledge of the second order terms (Hessian and higher derivatives) which are computationally very expensive because they need to calculate the inverse of the Hessian matrix etc.. So, in practice, they mostly use results that require only first order information, or at least an oracle calculating the Hessian along one direction. The latter has been shown that it can implemented efficiently by running a subroutine of a cubic regularization problem [AAZB+17].

One popular condition that guarantees that the SGD will avoid saddle points is the strict saddle property. Using the terminology of equation (1.1.1), a point p is a strict saddle when the linearization of F at p contains a positive eigenvalue see [Pem90] [JGN+17] and [Lev2006]. The paper [Pem90] establishes that if a stochastic approximation satisfies (1.1.1) then it will avoid, asymptotically, a strict saddle point with probability 1. A result of similar flavor is [LSJR16], where under the same condition they show that if you randomly initialize GD, then with probability 1 you avoid strict saddle points. Both of the problems use a stable manifold theorem. The former result using this decomposition finds a good approximation of the trajectories in the proximity of the saddle point, and by a change of coordinates it manages to see how the process gets jiggled in the unstable direction. The latter, via the stable manifold theorem argues that the set of stable trajectories is of measure zero, hence if you randomly initialize and then do GD you will avoid the saddle points.

The previous results established some confidence that the first order information is enough to evade saddle points asymptotically. Next, using the strict saddle property, in [JGN$^+$17] they managed to evade saddles point with high probability in $\mathrm{O}\left(\mathrm{polylog}(d)\frac{1}{\epsilon^2}\right)$ iterations. Here, using again the strict saddle property they managed to find a novel description of the geometry surrounding a saddle point. The idea is that when the Hessian has a negative eigenvalue, then only in a small band parallel to the corresponding eigenvector the process can get stuck. Since outside of this small band the eigenvector has a dominant effect which forces the process to decrease rapidly. Using these ideas the algorithm they came up with, repeats unperturbed gradient descent as long as the gradient is big. When the gradient is smaller than a certain threshold value they perturb the process, having in mind, that if the process lands outside the small band the gradient is dominant again. Finally, since, the area of the band is small, with high probability the gradient will become large so that performing gradient descent is efficient again.

1.6 Further discussion

Here, we will be occupied understanding the support of X_∞ in an one dimensional setting. More specifically, we will work with processes that solve

$$X_{n+1} - X_n = \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},\ \gamma \in (1/2, 1]. \tag{1.6.1}$$

To put in the SGD 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 $f(p) = 0$, then point p would be a saddle point.

Problem 1. *Let $(X_n)_{\geq 1}$ solve (1.6.1). Suppose that p is a saddle point. Find the threshold value:=$\tilde{\gamma}$ for γ, should it exist, such that:*

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

Part 1 of Problem 1 guarantees that the SGD avoids saddle points, and hence converging to local minima. Choosing different γ in the first regime i.e. $\gamma \in (1/2, \tilde{\gamma})$, enables us to optimize SGD's performance by choosing γ appropriately. 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 [EDM04] they 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 γ approximately $\frac{17}{20}$ the rate of convergence is optimal.

Here, we are trying to establish that if we understand the underlining dynamical system sufficiently, then by adding enough noise, the process will wander until it is captured by a downhill path, and, eventually, will escape the unstable neighborhood. Furthermore, an extension of the results, could, potentially, lead to SGD type algorithms (in higher dimensions) that converge fast to local minima, even in the proximity of degenerate saddle points, with high probability.

1.7 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 term. However, it is natural to consider, without the need to contemplate, a process defined by

$$\mathrm{d}L_t = \frac{f(L_t)}{t^\gamma}\mathrm{d}t + \frac{1}{t^\gamma}\mathrm{d}B_t,\ \gamma \in (1/2, 1]. \tag{1.7.1}$$

We assume that $f(0) = 0$, and 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 if it 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 1.9. *Suppose that $\mathcal{N}$ is a neighborhood of zero. Let $(L_t)_{t\geq 1}$ be a solution of (1.7.1), where $f(x)$ is Lipschitz. We distinguish two cases depending on f and the parameters of the system.*

1. $k|x| \leq f(x)$, $k \geq \frac{1}{2}$ *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 $\mathbb{P}(L_t \to 0) = 0$.

In part 1, we have only considered $\gamma = 1$ since that is the only critical case, namely for $\gamma < 1$ the effects of the noise would be overwhelming and for all k, we would obtain $\mathbb{P}(L_t \to 0) = 0$.

We now state the main convergence theorem:

Theorem 1.10. *Suppose that $\mathcal{N}$ is a neighborhood of zero. Let $(L_t)_{t\geq 1}$ be a solution of (1.7.1). 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 $\mathbb{P}(L_t \to 0) > 0$.

This is accomplished by first establishing the previous results for monomials i.e. $f(x) = |x|^k$ or $f(x) = k|x|$, which is done in sections 3.1, and 3.2. We prove the stated theorems in section 3.3, by utilizing the comparison results found in section 2.

In section 3.1, we deal with the linear case, i.e. when $f(x) = k|x|$. There, the SDE can be explicitly solved, which simplifies matters to a great extent. Firstly, in subsection 3.1.2, 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 3.1.3 we show that process will converge to 0 with some positive probability.

In section 3.2, 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, which is accomplished by studying the process $L_t/h(t)$. In subsection 3.2.2, the main theorem is that when $\frac{1}{2} + \frac{1}{2k} \geq \gamma$, then $L_t \to \infty$ a.s.. In subsection 3.2.3, we show that when $\frac{1}{2} + \frac{1}{2k} < \gamma$ holds, the process may converge to 0 with positive probability.

Qualitatively, the previous constrains 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 γ decreases the remaining variance increases, hence we should expect that the process visits the unstable trajectory with greater ease, due to higher fluctuations.

1.8 Results for the discrete model

In this section we will state the corresponding results for the discrete model. Furthermore at the end we will provide some examples, and experimental results corroborating the theoretical ones. The asymptotic behavior of the discrete processes is the expected one, depending on the parameters of the problem. Here, we study processes satisfying

$$X_{n+1} - X_n \geq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},\ \gamma \in (1/2, 1), \tag{1.8.1}$$

and

$$X_{n+1} - X_n \leq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},\ \gamma \in (1/2, 1), \tag{1.8.2}$$

or

$$X_{n+1} - X_n \geq \frac{f(X_n)}{n} + \frac{Y_{n+1}}{n}, \tag{1.8.3}$$

and

$$X_{n+1} - X_n \leq \frac{f(X_n)}{n} + \frac{Y_{n+1}}{n}, \tag{1.8.4}$$

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

Theorem 1.11. *Suppose that $\mathcal{N}$ is a neighborhood of zero. We separate two cases*

1. *Suppose $(X_n)_{n\geq 1}$ solve* (1.8.1), $|x|^k \leq f(x)$, $\frac{1}{2} + \frac{1}{2k} > \gamma$ *and* $k > 1$ *for all*

$x \in \mathcal{N}$

2. *Suppose* $(X_n)_{n\geq 1}$ *solve* (1.8.3), $k|x| \leq f(x),\ k > 1/2$ *for all* $x \in \mathcal{N}$

then in both cases $\mathbb{P}(X_n \to 0) = 0$.

For the convergence result the non-degeneracy condition $E(Y_{n+1}^2|\mathscr{F}_n) \geq l$ is replaced with the assumption stated in part 1 of Theorem 1.12.

Theorem 1.12. *Let* $\mathcal{N} = (-3\epsilon, 3\epsilon)$ *be a neighborhood of zero. Suppose* $(X_n)_{n\geq 1}$ *solve* (1.8.1). *We separate two cases*

1. *There exist* $-\epsilon_2 > -3\epsilon$, $-\epsilon_1 < -\epsilon$, *such that for all* $M > 0$, *there exists* $n > M$ *such that* $\mathbb{P}(X_n \in (-\epsilon_2, -\epsilon_1)) > 0$.

2. *Suppose* $(X_n)_{n\geq 1}$ *solve* (1.8.2) $0 < f(x) \leq |x|^k$, $\frac{1}{2} + \frac{1}{2k} < \gamma$ *and* $k > 1$ *for all* $x \in \mathcal{N}$.

3. *Suppose* $(X_n)_{n\geq 1}$ *solve* (1.8.4) $0 < k|x| \leq f(x),\ k > 1/2$ *for all* $x \in \mathcal{N}$.

Then when the 1 holds in both cases 2 and 3 we have $\mathbb{P}(X_n \to 0) > 0$

The assumption imposed on X_n, part 1 of Theorem 1.12, 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 process, the previous result need not hold. For instance, the drift could dominate the noise, and, consequentially, the process may never reach a neighborhood of the origin with probability 1. There are processes that naturally satisfy this property; such an example is the urn process defined seen in [Pem91], which is discussed in section 1.4.4.

Example 3. *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 U_n are I.I.D 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 $X_n \to 0$ holds with positive probability. In figure 1.1 we can see a typical example where convergence of the iterates occurs.*

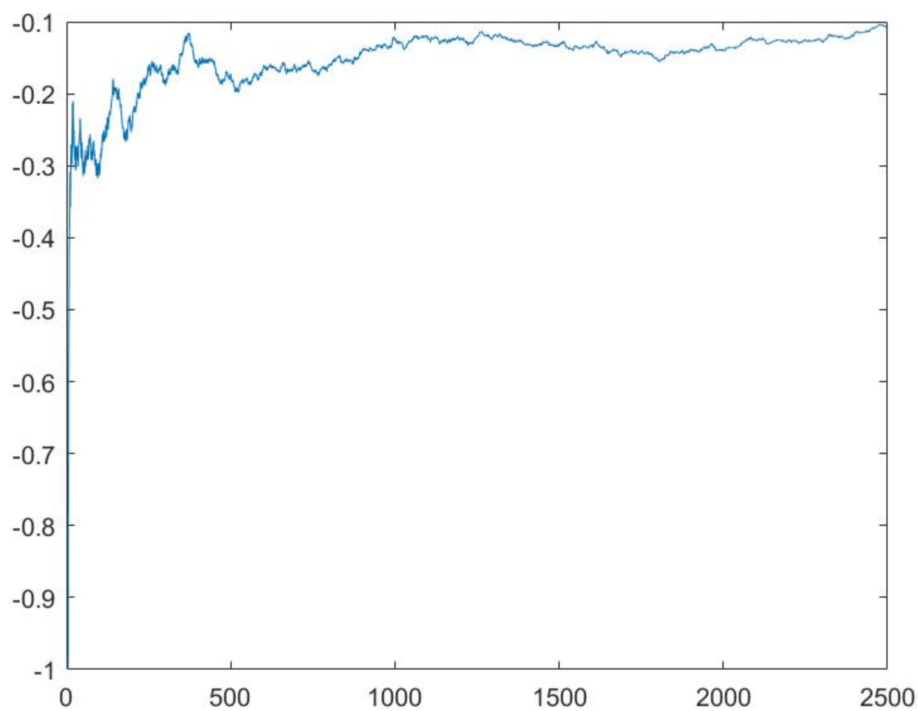

Figure 1.1: $(X_n)_{n\geq 10}$ and $X_{10} = -1$

Example 4. *Suppose that X_n satisfies $X_{n+1} - X_n = \frac{\max(k|X_n|,1)}{n} + \frac{U_n}{n}$ where U_n are I.I.D uniformly distributed on $(-2,2)$, and $k = \frac{1}{3}$. As before we have that the* *assumption 1 is satisfied. And since $k < \frac{1}{2}$ we expect $X_n \to 0$ to hold with positive probability. In figure 1.2 we can see a typical example where convergence occurs.*

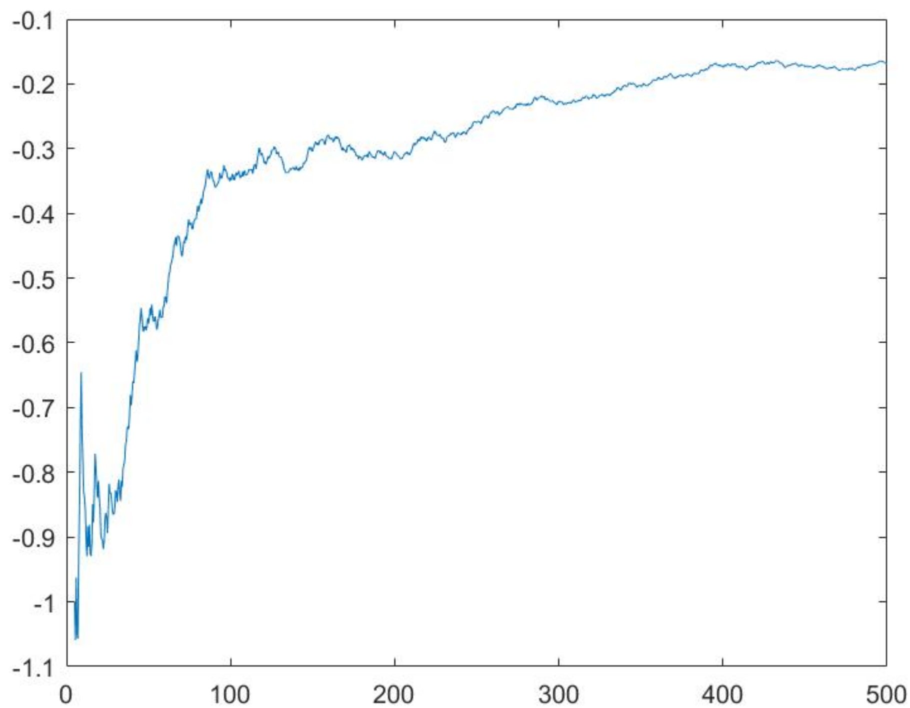

Figure 1.2: $(X_n)_{n\geq 10}$ and $X_{10} = -1$

Next we will provide some corroborating experimental results. Suppose X_n satisfies $X_{n+1} - X_n = \frac{f(X_n)}{n^\gamma} + \frac{U_n}{n^\gamma}$ where U_n are I.I.D uniformly distributed on $(-.5, .5)$. In the next table we have run simulations in order to investigate whether X_n converges to 0 or not for various values of γ. The simulations were run for initial $n = 100$ and $X_{100} = -1$. The criteria for nonconvergence is whether the process at some point exceeds $X_n > 1$, at which point the process explodes. The next table is for the specialization $f(x) = |x|^3$. According to Theorem 1.12 the threshold value for γ is $\tilde{\gamma} = 2/3$.

	γ	iterations	nonconv#/100
1	55/100	10^4	79%
2	55/100	10^5	100%
3	60/100	10^4	61%
4	60/100	10^5	86%
5	70/100	10^4	4%
6	70/100	10^5	17%
7	70/100	10^6	32%
8	70/100	10^7	38%

The next table is for the specialization $f(x) = |x|^2$. According to Theorem 1.12 the threshold value for γ is $\tilde{\gamma} = 3/4$.

	γ	iterations	nonconv#/100
1	60/100	10^4	91%
2	60/100	10^5	100%
3	70/100	10^4	33%
4	70/100	10^5	68%
5	70/100	10^6	94%
6	78/100	10^4	0%
7	78/100	10^5	1%
8	78/100	10^6	8%

Next we will provide a similar analysis for the linear case, i.e., when $f(x) = k|x|$ and $\gamma = 1$. Here, we have a threshold value for k, namely $\tilde{k} = 1/2$.

	k	iterations	nonconv#/100
1	1	10^4	21%
2	1	10^5	64%
3	1	10^6	79%
4	2/5	10^4	0%
5	2/5	10^5	3%
6	2/5	10^6	6%

1.9 Further conjectures

Suppose that $(X_n)_{n\geq 1}$ satisfies (1.1.1), where $F = -\nabla V$ and $a_n = \frac{1}{n^\gamma}$. Here, $V : \mathbb{R}^d \to \mathbb{R}$ is an analytic function, such that $V(0) = \nabla V(0) = 0$. Hereby, we assume that 0 is a saddle point that it is, also, an isolated critical point.

For this section we will focus on problem 1 part (1). More specifically, the goal is to find $\gamma \in (1/2, 1]$ such that $\mathbb{P}(X_n \to 0) = 0$. We will start by discussing about a potential strategy for a generic analytic function F which arises from some potential function V. Then we will provide specific examples.

One of the main tools we will need, for the initial discussion, is a Lojasiewicz type inequality, for a reference see [Spr], Theorem 2 and [Son12], Lemma 3.2, page 315.

Definition 1.13. *Suppose that* $V : \mathbb{R}^n \to \mathbb{R}$*. Then denote the zero set of* V *by* $Z_V = \{x \in \mathbb{R}^n : V(x) = 0\}$.

Theorem 1.14. *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:*

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

The line of attack consists of three distinct steps.

- We start by studying the process $(V(X_n))_{n\geq 1}$. Using Theorem 1.14 we get an upper bound on the $|V(X_n)|$.

- 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 concluding that $\mathbb{P}(X_n \to 0) = 0$.

For the first step using Theorem 1.14, part (a) we see that $Y_n := V(X_n)$ satisfies a recursion similar to the ones in section 1.8, namely

$$Y_{n+1} - Y_n = -\frac{|Y_n|^k}{n^\gamma} + \frac{\text{Noise}_{n+1}}{n^\gamma} + \mathrm{O}\left(\frac{1}{n^{2\gamma}}\right), \tag{1.9.1}$$

where the noise satisfies $E\left(\text{Noise}_{n+1}^2 | \mathscr{F}_n\right) \geq |Y_n|^k$. **Proof of equation** (1.9.1): We expand V centered at X_n and we obtain

$$\begin{aligned} V(X_{n+1}) - V(X_n) &= \nabla V(X_n) \cdot (X_{n+1} - X_n) + M||X_{n+1} - X_n||^2 \\ &\leq -\frac{||\nabla V(X_n)||^2}{n^\gamma} + \frac{\nabla V(X_n) \cdot B_{n+1}}{n^\gamma} + \frac{1}{n^{2\gamma}} \\ &< -\frac{|V(X_n)|^k}{n^\gamma} + \frac{\nabla V(X_n) \cdot B_{n+1}}{n^\gamma} + \frac{1}{n^{2\gamma}} \end{aligned}$$

Now, using that the noise is quasi-isotropic we see that the new noise satisfies

$$\begin{aligned} E((\nabla V(X_n) \cdot B_{n+1})^2 | \mathcal{F}_n) &= E((\frac{\nabla V(X_n)}{||\nabla V(X_n)||} \cdot B_{n+1})^2 ||\nabla V(X_n)||^2 | \mathcal{F}_n) \\ &= E((\frac{\nabla f(v_n)}{||\nabla V(X_n)||} \cdot B_{n+1})^2 | \mathcal{F}_n) ||\nabla V(X_n) - V(0)||^2 \\ &\geq 1 \cdot |V(X_n)|^k. \end{aligned}$$

The recursion defined in equation (1.9.1), can provide an upper bound $|V(X_n)|$, however this could be far from optimal as we have the bias term $\frac{1}{n^{2\gamma}}$ that we need take into account.

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 (1.1.1) are comparable, then X_n may wander in the realm where $V(X_n) < 0$.

Definition 1.15. *Suppose that* $V : \mathbb{R}^n \to \mathbb{R}$ *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 $\mathcal{O}_{X_n} \cap Z_V$ can be very steep so that under the slightest perturbation the iterate X_n may return to the realm $V(X_n) > 0$.

Example 5. *Suppose that* $V(x,y) = x^4 + x^2y^2 - y^4$, *that is* V *is a homogeneous polynomial of degree* 4. *Then since* $V(r\vec{x}) = r^4 V(\vec{x})$, *we see that* $\mathcal{O}_{\vec{y}} \cap Z_V$ *is a cone for any* $\vec{y}$ *such that* $V(\vec{y}) < 0$. *Define* $W_n = (X_n, Y_n)$ *given by*

$$W_{n+1} - W_n = \frac{\nabla V(W_n)}{n^\gamma} + \frac{B_{n+1}}{n^\gamma}$$

where for simplicity we assume that $B_{n+1} = (U_{1,n+1}, U_{2,n+2})$ *where* $U_{i,n}$ *for* $i = 1, 2$ *and* $n \in \mathbb{N}$ *is a collection of independent uniforms on* $(-1, 1)$. *Now, we may write* *the recursion for* X_n *and* Y_n*, namely*

$$X_{n+1} - X_n = \frac{-4X_n^3 - 2X_nY_n^2}{n^\gamma} + \frac{N_{1,n+1}}{n^\gamma} \tag{1.9.2}$$

and

$$Y_{n+1} - Y_n = \frac{-4Y_n^3 + 2X_n^2Y_n}{n^\gamma} + \frac{N_{2,n+1}}{n^\gamma} \tag{1.9.3}$$

Using symmetry we may assume that $X_n > 0$*. Under this assumption we obtain the following bound*

$$X_{n+1} - X_n \leq \frac{-4X_n^3}{n^\gamma} + \frac{N_{1,n+1}}{n^\gamma}.$$

Therefore, by Theorem 4.1 it is always possible to choose γ *such that* X_n *crosses* 0*. From here, we can see that the process* W_n *will eventually land with probability* 1 *in the realm* $V(W_n) < 0$*.*

Example 6. *Suppose that* $V(x, y) = x^6 + x^2y^2 - y^6$*. Here the conical surface is similar to the surface* $y = \sqrt{|x|}$*. However, even though the region* $\mathcal{O}_y$ *is very steep, see figure 1.3, simulations suggest that for certain values of* γ *the process defined by*

$$X_{n+1} - X_n = \frac{\nabla V(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},$$

will not get stuck along a stable trajectory, i.e., $\mathbb{P}(X_{n+1} \to 0) = 0$*.*

In the simulations the parameter γ *was set* $\gamma = .6$.

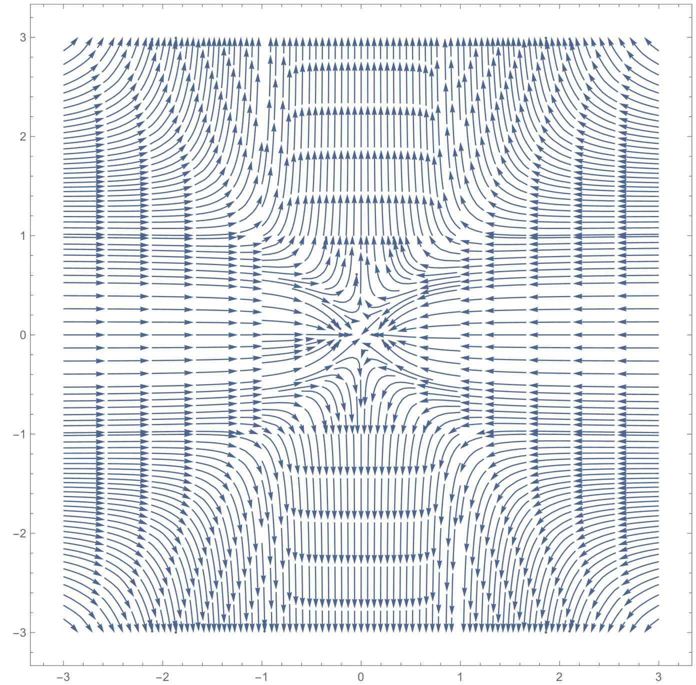

Figure 1.3: Vector flow

Chapter 2

Preliminary results

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

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(t)\mathrm{d}B_t. \tag{2.0.1}$$

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

Proof: We will argue by contradiction. Assume that $\limsup_{t\to\infty} X_t < 0$, and pick $\delta > 0$ such that $\limsup_{t\to\infty} X_t < -\delta$ 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 $\int_1^t f(X_s)\mathrm{d}s \to \infty$. However, since the process $G_t = \int_1^t g(s)\mathrm{d}B_s$ has finite quadratic variation, i.e., $\sup_t \langle G_t \rangle = \int_0^\infty g^2(t)\mathrm{d}t < \infty$, G_t stays a.s. finite. The last two observations imply that $X_t \to \infty$, which is a contradiction. ■

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

Proof: We will again argue by contradiction. Assume that $\liminf_{t\to\infty} X_t < 0$ on a set of positive probability. Take an enumeration of the pair 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 $\limsup_{t\to\infty} X_t \geq 0$, we have $\bigcup_{n\geq 0} A_n = \{\liminf_{t\to\infty} X_t < 0\}$. 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

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

Hence we conclude that $\int_{t_1}^{t_2} g(s)\mathrm{d}B_s \leq -q_n + p_n$. By the definition of A_n, on event A_n we can find a sequence of times (t_{2k}, t_{2k+1}) such that $t_{2k} < t_{2k+1}$ and $\int_{t_{2k}}^{t_{2k+1}} g(s)\mathrm{d}B_s \leq -q_n + p_n$. Now, if we define $G_{u,t} = \int_u^t g(s)\mathrm{d}B_s$, 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., $\mathbb{P}(\liminf_{t\to\infty} X_t < 0) = 0$. ■

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

Proposition 2.3. *Let $(C_t)_{t\geq 0}$ and $(D_t)_{t\geq 0}$ stochastic processes in the same Wiener space, that satisfy $\mathrm{d}C_t = f_1(C_t)\mathrm{d}t + g(t)\mathrm{d}B_t$, $\mathrm{d}D_t = f_2(D_t)\mathrm{d}t + g(t)\mathrm{d}B_t$ 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 $C_t > D_t \forall t \geq s_0$ a.s..*

Proof: Define $\tau = \inf\{\tau > s_0 | C_\tau = D_\tau\}$, and set $D_\tau = C_\tau = c$, for $\tau < \infty$. Now, from continuity of f_1, and f_2 we can find δ such that $f_1(x) > f_2(x), \forall x \in (c - \delta, c]$. However, for all s we have $C_\tau - D_\tau - (C_s - D_s) = -(C_s - D_s) = \int_s^\tau f_1(C_u) - f_2(D_u) \mathrm{d}u$. Thus, for s such that $C_y, D_y \in (c - \delta, c) \forall y \in (s, \tau)$ we have

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

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

$$X_{n+1} - X_n \geq \frac{f(X_n)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma}, \ \gamma \in (1/2, 1) \tag{2.0.2}$$

where f satisfies $\forall \epsilon > 0, \exists c > 0, f(x) \geq c, \ x \in (-\infty, -\epsilon)$, and the Y_n are defined similarly, as in (1.8.1).

Lemma 2.4. $\limsup X_n \geq 0$ *a.s..*

Proof: The proof is nearly identical as in the continuous case (Lemma 2.1). ■

Lemma 2.5. $\liminf_{t\to\infty} X_t \geq 0$ *a.s.*.

Proof: The proof is identical as in 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 2.6. *Let $\mathscr{F}_n$, $n \geq 0$ be a filtration with $\mathscr{F}_0 = \{0, \Omega\}$, and A_n, $n \geq 1$ a sequence of events with $A_n \in \mathscr{F}_n$. Then*

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

Chapter 3

Continuous model

3.1 Continuous model, simplest case

3.1.1 Introduction

Let L_t be defined by (1.7.1), 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,

$$\begin{aligned} X_{t+\mathrm{d}t} - X_t &= L_{e^t + e^t \mathrm{d}t} - L_{e^t} \\ &= k|L_{e^t}|\mathrm{d}t + e^{-t}(B_{t+e^t\mathrm{d}t} - B_{e^t}) \\ &= k|X_t|\mathrm{d}t + e^{-\frac{t}{2}}\mathrm{d}B_t. \end{aligned}$$

Which will be the model we will study. We begin with some definitions.

$$\mathrm{d}X_t = k|X_t|\mathrm{d}t + e^{-\frac{t}{2}}\mathrm{d}B_t. \tag{3.1.1}$$

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

$$\mathrm{d}K_t = kK_t\mathrm{d}t + e^{-\frac{t}{2}}\mathrm{d}B_t. \tag{3.1.2}$$

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 $(\Omega, \mathscr{F}, \mathbb{P})$. The solution for SDE (3.1.2), is given by $K_t = e^{kt}(e^{-t_0 k}K_{t_0} + \int_{t_0}^t e^{-s(k+\frac{1}{2})}\mathrm{d}B_s)$. Indeed, substituting in (3.1.2), and using Itô's formula we get

$$\begin{aligned}\mathrm{d}K_t &= a'(t)(k_0 + \int_{t_0}^t b(s)\mathrm{d}B_s) + a(t)b(t)\mathrm{d}B_t \\ &= \frac{a'(t)}{a(t)}K_t + a(t)b(t)\mathrm{d}B_t.\end{aligned}$$

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

Proposition 3.1. *Let $(X_t)_{t\geq t_0}$, $(K_t)_{t\geq t_0}$ in the Wiener probability space $(\Omega, \mathscr{F}, \mathbb{P})$ be the solutions of* (3.1.1),(3.1.2) *respectively. We start them at time t_0, $X_{t_0} \geq K_{t_0} \geq 0$. Then $X_t \geq K_t$, $\forall t \geq t_0$.*

It is a direct application of Proposition 2.3. ■

3.1.2 Analysis of X_t when $k \geq 1/2$.

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

Theorem 3.2. *Let $(X_t)_{t\geq 1}$ the solution of* (3.1.1) *for $k \geq \frac{1}{2}$, then $X_t \to \infty$ a.s..*

We will prove the theorem at the end of the subsection. 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 3.3. *Let $(X_t)_{t\geq 1}$ the solution of* (3.1.1) *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 (3.1.1), run from time s with initial condition $X_s < 0$ coincides with the solution of the SDE $\mathrm{d}X_t = -kX_t\mathrm{d}t + e^{-\frac{t}{2}}\mathrm{d}B_t$ before X_t hits 0. Formally, we define $\tau_0 = \inf\{t | t \geq s,\ X_t = 0\}$. Using the same method when solving SDE (3.1.2), we obtain $X_t = e^{-kt}(e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})}\mathrm{d}B_u)$, on $\{t < \tau_0\}$. First we deal with the case $k \neq \frac{1}{2}$. Set $G_t = \int_s^t e^{u(k-\frac{1}{2})}\mathrm{d}B_u$, 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.

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

When $k = \frac{1}{2}$, the solution simplifies to $X_t = e^{-kt}(e^{ks} X_s + B_t)$ in distribution, where $\{B_t\}_{t \ge s}$ with initial condition $B_s = 0$. We repeat the previous calculation,

$$
\begin{aligned}
\mathbb{P}(\tau = \infty) &= \mathbb{P}\left(\sup_{s<u<\infty} X_u \le 0\right) \\
&= \mathbb{P}\left(\sup_{s<u<\infty} B_u \le -e^{ks} X_s\right) \\
&= 0
\end{aligned}
$$

as for a Brownian motion we have $\limsup_{t\to\infty} B_t = \infty$ almost surely. ■

We will now prove two important lemmas, that are true for solutions of (3.1.1) for any $k > 0$.

Lemma 3.4. *Let* $(X_t)_{t\geq 1}$ *the solution of* (3.1.1). *Then on the event* $\{X_t \geq 0 \ i.o.\}$, *there is a positive constant* $c < 1$ *such that* $\{X_t \geq ce^{-t/2} \, i.o.\}$ *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 $X_t \geq \int_{t_i}^t k|X_u|\mathrm{d}u + \int_{t_i}^t e^{-\frac{u}{2}}\mathrm{d}B_u \geq \int_{t_i}^t e^{-\frac{u}{2}}\mathrm{d}B_u$.

Set $G_t = \int_{t_i}^t e^{-\frac{u}{2}}\mathrm{d}B_u$. The quadratic variation of G_t, is $\langle G_t \rangle = e^{-t_1} - e^{-t}$. Fix $0 < c < 1$. Now, observe that we can always choose t big enough such that $\langle G_t \rangle \geq ce^{-t_1}$ for any t_1.

Then,

$$
\begin{aligned}
\mathbb{P}(\sup_{t_i<u<t} X_t > e^{-t_1/2}) &\geq \mathbb{P}(\sup_{t_i<u<t} G_t > e^{-t_1/2}) \\
&= 2\mathbb{P}(G_t > e^{-t_1/2}) \\
&\geq 2\mathbb{P}(N(0, ce^{-t_1}) > e^{-t_1/2}) \\
&= 2\mathbb{P}(N(0, c) > 1) > \gamma > 0.
\end{aligned}
$$

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 $\mathscr{F}_n = \mathscr{F}_{\tau_n}$, for $n \geq 1$ and $\mathscr{F}_0 = \{0, \Omega\}$. Now, let $A_n = \{\exists t,\, \tau_{n-1} < t < \tau_n\,$, s.t. $X_t \geq ce^{-t/2}\}$. So, by definition $A_n \in \mathscr{F}_n$. We find a lower bound for

$\mathbb{P}(A_n|\mathscr{F}_{n-1})$.

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

On $\{X_t \geq 0 \text{ i.o.}\}$ the sum $\sum_{n\geq 1}\mathbb{P}(A_n|\mathscr{F}_{n-1})$ has infinite non zero terms bigger than γ, hence $\sum_{n\geq 1}\mathbb{P}(A_n|\mathscr{F}_{n-1}) = \infty$ a.s.. Finally, by Lemma 2.6 (Borel-Cantelli) we conclude. ■

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

Lemma 3.5. *Let* $(X_t)_{t\geq 1}$ *the solution of* (3.1.1)*. Then on the event* $\{X_t \geq 0$ *i.o.*$\}$ *we have that* $\{\liminf_{t\to\infty} X_t > 0\}$ *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 (3.1.1), before hitting 0, is given by

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

Denote $G_t = \int_s^t e^{-s(k+\frac{1}{2})}\mathrm{d}B_s$. We calculate its quadratic variation

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

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

$$\begin{aligned}
\mathbb{P}(\inf_{s\le u<\infty} X_u > \frac{c}{2}e^{-\frac{s}{2}}) &= \mathbb{P}(\inf_{s\le u<\infty} e^{ku}(ce^{-s(k+\frac{1}{2})} + G_u) > \frac{c}{2}e^{-\frac{s}{2}}) \\
&\ge \mathbb{P}(\inf_{s\le u<\infty} e^{ks}(ce^{-s(k+\frac{1}{2})} + G_u) > \frac{c}{2}e^{-\frac{s}{2}}) \\
&= \mathbb{P}(\inf_{s\le u<\infty} ce^{-s(k+\frac{1}{2})} + G_u > \frac{c}{2}e^{-s(k+\frac{1}{2})}) \\
&= \mathbb{P}(\inf_{s\le u<\infty} G_u > -\frac{c}{2}e^{-s(k+\frac{1}{2})}) \qquad (3.1.3) \\
&= 1 - \mathbb{P}(\sup_{s\le u<\infty} G_u > \frac{c}{2}e^{-s(k+\frac{1}{2})}) \\
&= 1 - 2\lim_{t\to\infty} \mathbb{P}(G_t > -\frac{c}{2}e^{-s(k+\frac{1}{2})}), \text{ by the reflection principle} \\
&= 1 - 2\mathbb{P}(N(0, \frac{e^{-s(2k+1)}}{2k+1}) > \frac{c}{2}e^{-s(k+\frac{1}{2})}) \\
&= 1 - 2\mathbb{P}(N(0, \frac{1}{k+1}) > \frac{c}{2}) > \delta > 0.
\end{aligned}$$

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

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

$$\mathbb{P}(A|\mathscr{F}_n) \leq 1 - \mathbb{P}(\inf_{\tau_n \leq u < \infty} X_u > \frac{c}{2} e^{-\frac{\tau_n}{2}} |\mathscr{F}_n)$$
$$< 1 - \delta.$$

■

Now, we can prove Theorem 3.2.

Proof of Theorem 3.2: From Proposition 3.3 we know that $\{X_t \geq 0 \text{ i.o.}\}$ has probability 1. Therefore, from Lemma 3.5 we deduce $\liminf_{t\to\infty} X_t > 0$ almost surely. Consequently, $\int_0^\infty |X_u| \mathrm{d}u \to \infty$ a.s.. At the same time $\limsup_{t\to\infty} \int_0^t e^{-\frac{u}{2}} \mathrm{d}B_u < \infty$ a.s., hence $X_t \to \infty$ a.s.. ■

3.1.3 Analysis of X_t when $k < 1/2$.

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

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 3.6. *Let $(X_t)_{t\geq 1}$ solve* (3.1.1) *with $k < \frac{1}{2}$, and define $A = \{X_t \to 0\}$, $B = \{X_t \to \infty\}$. Then the following hold:*

1. *Let A, B as before. Then $\mathbb{P}(A \cup B) = 1$.*

2. Both A and B are non trivial, i.e., $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.

3. On $\{X_t \geq 0 \text{ i.o.}\}$ we get $X_t \to \infty$.

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

Proposition 3.7. *Let $(X_t)_{t\geq 1}$ solve (3.1.1) with $k < \frac{1}{2}$. Assume that at time s, $X_s < 0$. Then $(X_t)_{t\geq 1}$ will hit 0 with probability α where $0 < \alpha < 1$.*

Proof: Define the stopping time $\tau_1 = \inf\{t \geq s | X_t = 0\}$. As in Proposition 3.3, the solution for X_t started at time s up till time τ_1, is given by $X_t = e^{-kt}(e^{ks}X_s + \int_s^t e^{u(k-\frac{1}{2})} \mathrm{d}B_u)$.

$$
\begin{aligned}
\mathbb{P}(\tau = \infty) &= \mathbb{P}(\sup_{s<u<\infty} X_u \leq 0) \\
&= 1 - \lim_{t\to\infty} 2\mathbb{P}(N(0, \frac{e^{2t(k-\frac{1}{2})} - e^{2t(k-\frac{1}{2})}}{2k-1}) > -e^{ks}X_s), \text{ as in Proposition 3.3} \\
&= 1 - 2\mathbb{P}(N(0, -e^{2s(k-\frac{1}{2})}/(2k-1)) > -e^{ks}X_s) = 1 - \alpha.
\end{aligned}
$$

Therefore $0 < \alpha < 1$. ■

Proof of Theorem 3.6:

1. Define the events $N = \{\exists s, \text{ s.t.} X_t < 0 \forall t \geq s\}$, 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 *1*, we will show that $N \subset \{X_t \to 0\}$ up to a null set and $P = \{X_t \to \infty\}$.

From Lemma 2.2 we know that $\liminf_{t\to\infty} X_t \geq 0$ a.s., therefore on $N \subset \{X_t \to 0\}$ up to a null set.

To show that $P = \{X_t \to \infty\}$, note that Lemma 3.5 shows that on $\{X_t \geq 0\,\text{i.o.}\}$, $\liminf_{t\to\infty} X_t > 0$ almost surely. Consequently, on $\{X_t \geq 0 \text{ i.o.}\}$, we have $X_t \to \infty$, as $\int_0^\infty |X_u|\mathrm{d}u \to \infty$ and $\limsup_{t\to\infty} \int_0^t e^{-\frac{u}{2}}\mathrm{d}B_u < \infty$ a.s.. Therefore, $P = \{X_t \to \infty\}$. Which concludes part *1.*

2. The fact that $\mathbb{P}(A) > 0$, follows immediately from Proposition 3.7. 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

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

3. Follows immediately from the proof of *1.*

■

Lastly, we prove a proposition that will be used in section 3.3.

Proposition 3.8. *Suppose* $(X_t)_{t\geq1}$, $(Y_t)_{t\geq1}$ *solve* (3.1.1)*, 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 both $X_t, Y_t \in (-3\epsilon, 0)$, $\forall t > s$.

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

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

So,

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

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

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

Similarly we obtain $-\int_s^t (k_1-k)e^{(k_1-k)u}G_u\mathrm{d}u \ge \epsilon\left(e^{(k_1-k)t} - e^{(k_1-k)s}\right)$. Thus on A, we obtain the following inequalities

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

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 following upper bound,

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

and lower bound

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

Doing similarly for Y_t, we conclude. ■

3.2 Analysis of $\mathrm{d}L_t = \frac{|L_t|^k}{t^\gamma}\mathrm{d}t + \frac{1}{t^\gamma}\mathrm{d}B_t$.

3.2.1 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 $\mathrm{d}B_{\theta(t)} = \sqrt{\theta'(t)}\mathrm{d}B_t$. Therefore

$$\begin{aligned}\mathrm{d}X_t &= \frac{|X_t|^k}{\theta(t)^\gamma}\theta'(t)\mathrm{d}t + \frac{1}{\theta(t)^\gamma}\sqrt{\theta'(t)}\mathrm{d}B_t \\ &= c_1|X_t|^k\mathrm{d}t + c_2 t^{-\frac{\gamma}{1-\gamma}}\sqrt{\theta'(t)}\mathrm{d}B_t \\ &= c_1|X_t|^k\mathrm{d}t + c_2 t^{-\frac{\gamma}{2(1-\gamma)}}\mathrm{d}B_t\end{aligned}$$

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

$$\mathrm{d}X_t = c|X_t|^k\mathrm{d}t + t^{-\frac{\gamma}{2(1-\gamma)}}\mathrm{d}B_t,\ k>1 \text{ and } \gamma\in(1/2,1),\ c\in(0,\infty). \tag{3.2.1}$$

By a time scaling, we may assume that X_t solves

$$\mathrm{d}X_t = |X_t|^k\mathrm{d}t + t^{-\frac{\gamma}{2(1-\gamma)}}\mathrm{d}B_t,\ k>1 \text{ and } \gamma\in(1/2,1). \tag{3.2.2}$$

Notice, that the noise is scaled differently. However, it will be evident that only the order of the noise is relevant. The SDE (3.2.2) will be the primary focus of the next subsection and the results will apply to solutions of (3.2.1), 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 3.9. *Suppose that* $(X_t)_{t\geq1}$ *solve* (3.2.1), *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*

$$Z_t - Z_s = \int_s^t c\frac{X_u}{h(u)}\left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right)\mathrm{d}u + \int_s^t -\frac{1}{h(u)}u^{-\frac{\gamma}{2(1-\gamma)}}\mathrm{d}B_u. \quad (3.2.3)$$

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

$$Z_t - Z_s = \int_s^t c|h(u)|^{k-1}Z_u\left(C - (-Z_u)^{k-1}\right)\mathrm{d}u + \int_s^t -\frac{1}{h(u)}u^{-\frac{\gamma}{2(1-\gamma)}}\mathrm{d}B_u. \quad (3.2.4)$$

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

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

Thus,

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

The SDE (3.2.4) is an immediate consequence of the last line of the previous calculation. ■

In the next proposition, we describe some properties of the noise for the process Z_t, and give a very important inequality for subsection 3.2.2, which relates the order of the deterministic system converging to zero and order of the remaining noise for X_t, i.e., the order of $\langle \int_s^\infty u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u \rangle$.

Proposition 3.10. *Set* $G'_{s,t} = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u$*, the noise term of* (3.2.4) *and* (3.2.3).

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, γ)*, 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, $\langle G'_{s,t} \rangle = \int_s^t \frac{1}{h(u)^2} u^{-\frac{\gamma}{(1-\gamma)}} \mathrm{d}u$. Notice that by the definition of $h(t)$, we have $h(t)^{-1} = \Theta(t^{\frac{1}{k-1}})$, therefore $-\frac{1}{h(u)^2} u^{-\frac{\gamma}{(1-\gamma)}} = \Theta(u^{\frac{2}{k-1} - \frac{\gamma}{1-\gamma}})$. 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,

$$\begin{aligned}\frac{1}{2}+\frac{1}{2k}\geq\gamma \iff \frac{k-1}{2k}\leq 1-\gamma \iff \frac{2k}{k-1}\geq\frac{1}{1-\gamma}\\ \iff \frac{2}{k-1}+2\geq\frac{1}{1-\gamma} \iff \frac{2}{k-1}+\frac{1}{1-\gamma}\geq -2\end{aligned} \tag{3.2.5}$$

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 $\frac{1}{2}+\frac{1}{2k}<\gamma$ then $\frac{2}{k-1}-\frac{\gamma}{1-\gamma}<-1$, therefore when $\frac{1}{2}+\frac{1}{2k}<\gamma$, $\langle G'_{s,t}\rangle<\infty$. Finally, rearranging the first inequality of (3.2.5) we obtain $\frac{1}{k-1}\geq\frac{2\gamma-1}{2(1-\gamma)}$. ■

The solution of the SDE (3.2.2), 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 it would imply that SDE (3.2.2) 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 when $|x|>M$. This SDE can be seen to admit strong solutions for infinite time. The reason is that this process X_t is a.s. bounded from below, as 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 as shown in (3.2.2).

3.2.2 Analysis of X_t when $1/2+1/2k\geq\gamma$, $k>1$ and $\gamma\in(1/2,1)$

The main result of this subsection is the following theorem.

Theorem 3.11. *Let $(X_t)_{t\geq 1}$, that solves* (3.2.2)*. When $1/2+1/2k \geq \gamma, X_t \to \infty$ a.s.*

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

Proposition 3.12. *Let $(X_t)_{t\geq 1}$ solve* (3.2.1) *for $c=1$. Then, for some $\beta<0$, the event $\{X_t \geq \beta t^{\frac{1-2\gamma}{2(1-\gamma)}}\, i.o.\}$ has probability* 1*.*

Proof: Set $G'_t = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u$, which corresponds to the noise term of (3.2.4). 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 $\int_s^t \dfrac{X_u}{h(u)} \left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right) \mathrm{d}u > 0$ for all $t > 0$. However, since $G'_{w,t} := \int_w^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u$ for any fixed w has infinite quadratic variation, we may find $G'_{s,t} > -Z_s$. Now, from (3.2.3) we get

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

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

Finally, in Proposition 3.10 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 3.13. *Let $(X_t)_{t\geq 1}$ solve (3.2.1) for $c = 1$. Then $\liminf_{t\to\infty} X_t > 0$ almost surely.*

Proof: Set $G_{s,t} = \int_s^t u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u$ 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 by using the reflection principle $W(u) > u > 0$ such that

$$\mathbb{P}\left(\sup_{u<t<W(u)} G_{u,t} > \gamma u^{\frac{1-2\gamma}{2(1-\gamma)}}\right) > \delta, \tag{3.2.6}$$

for δ independent of u. Take $\gamma > -\beta$, where β 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

$$\mathbb{P}\left(\sup_{s<t<W(s)} X_t - X_s > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}}\right) \geq \mathbb{P}\left(\sup_{s<t<W(s)} G_{s,t} > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}}\right) > \delta. \tag{3.2.7}$$

When $X_s \geq \beta s^{\frac{1-2\gamma}{2(1-\gamma)}}$ observe that on the event $\{\sup_{s<t<W(s)} X_t - X_s > \gamma s^{\frac{1-2\gamma}{2(1-\gamma)}}\}$ there is τ_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 $\mathbb{P}(\sup_{\tau_n<t<\tau_{n+1}} G_{\tau_n,t} > \gamma {\tau_n}^{\frac{1-2\gamma}{2(1-\gamma)}} | \mathcal{F}_{\tau_n}) > \delta$. So, by Borel-Cantelli (Lemma 2.6) on the events $\{\sup_{\tau_n<t<\tau_{n+1}} X_t - X_{\tau_n} > \gamma {\tau_n}^{\frac{1-2\gamma}{2(1-\gamma)}}\}$ we may conclude that $\{X_t \geq 0 \text{ i.o.}\}$ has probability 1. Define τ_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 $\langle G_{s,\infty}\rangle = \Theta(s^{\frac{1-2\gamma}{(1-\gamma)}})$

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

for some δ' independent of s.

Define τ_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 $A = \{\liminf_{\to\infty} X_t \leq 0\}$ has probability zero, it suffices to argue that there is a δ such that $\mathbb{P}(A|\mathscr{F}_n) < 1-\delta$, a.s. for all $n \geq 1$. This is immediate from the previous calculation. Indeed,

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

■

Proof of Theorem 3.11: Since X_t is a solution of (3.2.2), we have $X_t - X_1 = \int_1^t |X_u|^k \mathrm{d}u + G_{1,t}$. From Corollary 3.13 we know that $\liminf_{t\to\infty} X_t > 0$ a.s., therefore $\int_1^t |X_u|^k \mathrm{d}u \to \infty$ almost surely. However, since $\langle G_{1,\infty}\rangle < \infty$, we have that $\limsup_{t\to\infty} |G_{1,t}| < \infty$ a.s.. Therefore, $X_t \to \infty$ a.s..

■

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

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

Theorem 3.14. *The process $(X_t)_{t\geq 1}$ the solution of (3.2.1), converges to zero with positive probability, when $X_1 < 0$.*

We prove a technical lemma first.

Lemma 3.15. *Let $(Z_t)_{t\geq s}$ solve (3.2.3), and set $G'_t = \int_s^t -\frac{1}{h(u)} u^{-\frac{\gamma}{2(1-\gamma)}} \mathrm{d}B_u$. Suppose that $Z_s > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$. Define $A = \{G'_t \in (-\epsilon, \epsilon) \forall t \in (s, s+\delta)$, and $G'_t \in (-2\epsilon, -\frac{9}{10}\epsilon) \forall t \in (s+\delta, \infty)\}$. Then:*

1. *$\mathbb{P}(A) > 0$, $\forall \epsilon, \delta > 0$.*

2. *For all $\epsilon > 0$ small enough, there is $\delta > 0$ such that $Z_t < -\dfrac{5\epsilon}{3}$, $\forall t \in (s, s+\delta)$ on A.*

3. *Define $\tau_C = \inf\{t > s | Z_t = -2\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\}$, then $\tau_C > s + \delta$ almost surely, where δ is the same as part 2.*

Proof:

1. This is immediate, since in Proposition 3.10 part *2* we have shown that $\langle G'_\infty \rangle < \infty$.

2. The first restriction on ϵ is such that $Z_s < -3\epsilon$. Next, we begin by defining f_1 and f_2 on $(s, s+\delta)$ satisfying

$$f'(x) = c|h(x)|^{k-1} f(x)(C - (-f(x)^{k-1})) \tag{3.2.8}$$

where c and C are the same as the parameters of SDE (3.2.3), 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 $(\left(-\frac{C}{k}\right)^{\frac{1}{k-1}}, 0)$. This function will be important later. We should also note, that f is decreasing in intervals where $f(x) \in (-\left(\frac{C}{k}\right)^{\frac{1}{k-1}}, 0)$, since there $f'(x) < 0$.

We can pick the $\delta > 0$, such that $f_2(t) > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$, $\forall t \in (s, s+\delta)$. We will show that $Z_t > f_2(t)$ on $(s, s+\delta)$ by contradiction. Using the SDE (3.2.4) for Z_t, we get that

$$Z_t - Z_s = \int_s^t c|h(u)|^{k-1} Z_u \left(C - (-Z_u)^{k-1}\right) \mathrm{d}u + g(t), \tag{3.2.9}$$

where $g(t)$ is a continuous function such that $\sup_{t \in (s, s+\delta)} |g(t)| \leq \epsilon$. Assume that f_2, Z become equal at some point, and choose t to be the first time. Using the integral form of (3.2.8), and subtracting it from (3.2.4), we get

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

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 $\tau_C > s+\delta$

almost surely.

■

Before proving the theorem we will need the following proposition.

Proposition 3.16. *Let* $(X_t)_{t\geq s}$ *solve* (3.2.1). *Assume that at time* s, $X_s < 0$, *and* $Z_s > -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}$. *Then the process with positive probability never returns to* 0.

Proof: The condition $1/2+1/2k < \gamma$ as it has already been shown in Proposition 3.10 part *2*, implies that $\langle G'_\infty\rangle < \infty$. On the event A as defined in Lemma 3.15, using (3.2.3), we get the following upper and lower bounds for all $t \geq s+\delta$

$$-\frac{X_t}{h(t)} \leq -\frac{X_s}{h(s)} + \int_s^t c\frac{X_u}{h(u)}\left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right)\mathrm{d}u - \frac{9}{10}\epsilon \tag{3.2.10}$$

$$-\frac{X_t}{h(t)} \geq -\frac{X_s}{h(s)} + \int_s^t c\frac{X_u}{h(u)}\left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right)\mathrm{d}u - 2\epsilon \tag{3.2.11}$$

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 ϵ, such that $\frac{3\epsilon}{2} < C^{\frac{1}{k-1}}$. Now, define $\tau_l = \sup\{t \leq \tau_0 | -\dfrac{X_t}{h(t)} = -\frac{3\epsilon}{2}\}$ and notice that Lemma 3.15, implies that $\tau_{l\epsilon} > s+\delta$, since $Z_t < -\dfrac{5\epsilon}{3}$, on $(s, s+\delta)$. Also, on $\{\tau_0 < \infty\} \cap A$ we have $\tau_l < \infty$. Then from (3.2.11) we see that

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

Therefore,

$$-\frac{X_s}{h(s)} + \int_s^{\tau_l} c\frac{X_u}{h(u)}\left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right)\mathrm{d}u - \frac{9}{10}\epsilon \le -\frac{2\epsilon}{5}. \tag{3.2.12}$$

Now, notice that $X_t > \frac{3}{2}\epsilon h(t)$, $\forall t \in (\tau_l, \tau_0)$, 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

$$\int_{\tau_l}^{\tau_0} c\frac{X_u}{h(u)}\left(C\frac{|h(u)|^k}{h(u)} - \frac{|X_u|^k}{X_u}\right)\mathrm{d}u < 0. \tag{3.2.13}$$

Combining (3.2.12), and (3.2.13), we get that

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

a contradiction. ■

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

Proof of Theorem 3.14: 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 3.16 implies that X_t converges to zero with positive probability. Indeed, remember from Lemma 2.2 that $\liminf_{t\to\infty} X_t \ge 0$ a.s.. Therefore, since on the event A as defined in Lemma

3.15 we have $\limsup_{t\to\infty} X_t \leq 0$, we deduce $\lim_{t\to\infty} X_t = 0$. To finish the proof, it suffices to show that when $\{\sigma < \infty\}$ has zero probability then $X_t \to 0$ with positive probability. This is easy to see since $\mathbb{P}(\{\sigma < \infty\}) = 0$ implies that Z_t, never hits zero, therefore $\limsup_{t\to\infty} X_t \leq 0$ on $\{\sigma < \infty\}$. ■

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

Proposition 3.17. *Let* $(X_t)_{t\geq s}$ *solve* (3.2.1)*. Take the event* A*, such that Lemma 3.15 holds, where* $\epsilon < \left(\frac{C}{k}\right)^{\frac{1}{k-1}}$*, where* $C(c) = \frac{1}{c(k-1)}$ *as in the parameter* C *of SDE* (3.2.3)*. 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 $\tau_C = \inf\{t > s | Z_t = -2\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\}$, and define $\sigma = \sup\{\tau_C > t > s | Z_t = -\left(\frac{C}{k}\right)^{\frac{1}{k-1}}\}$. We will show that $\tau_C = \infty$ a.s.. We assume otherwise, and reach a contradiction. From Lemma 3.15 part *3*, we know that $\tau_C > s + \delta$. Therefore,

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

the desired contradiction. ■

3.3 Analysis of $\mathrm{d}L_t = \frac{f(L_t)}{t^\gamma}\mathrm{d}t + \frac{1}{t^\gamma}\mathrm{d}B_t$.

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

$$\mathrm{d}L_t = \frac{f(L_t)}{t^\gamma}\mathrm{d}t + \frac{1}{t^\gamma}\mathrm{d}B_t, \gamma \in (\frac{1}{2}, 1] \tag{3.3.1}$$

By our assumptions on f, the SDE (3.3.1) admits strong solutions. Also, we define a more general SDE, namely

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(t)\mathrm{d}B_t \tag{3.3.2}$$

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

Proposition 3.18. *Let $(X_t)_{t\geq 1}$ be a solution of* (3.3.2)*. 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 $\xi(t) = \int_0^t g^2(t)\mathrm{d}t$, and define $\tilde{X}_t = X_{\xi^{-1}(t)}$. Then,

$$\mathrm{d}\tilde{X}_t = \frac{f(\tilde{X}_t)}{g^2(\xi^{-1}(t))}\mathrm{d}t + \mathrm{d}B_t \tag{3.3.3}$$

This gives a well defined SDE whose solution is defined on $[0, T']$ for $T' \in \mathbb{R},\ T' < T$. 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 proof of Theorems 1.9 and 1.10. For the proofs, we use that the theorems hold if and only if they hold for their corresponding reparameterizations.

Proof of Theorem 1.9 part 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 $\mathscr{F}_\tau$ that satisfies (3.2.2) with $U_\tau = -2\epsilon$. From Proposition 2.3, we see that $U_t < X_t, \forall t \in (\tau, \tau')$. Now, we can see that $\mathbb{P}(\tau' = \infty) = 0$. Indeed, $\mathbb{P}(\tau' = \infty) \leq \mathbb{P}(U_t \leq 3\epsilon \forall t \geq \tau) \leq 1 - \mathbb{P}(U_t \to \infty) = 0$. ■

Proof Theorem 1.10 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 3.18, $\tau < \infty$ 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 $\mathscr{F}_\tau$, that satisfy (3.2.1) with drift constant 1 and c respectively. From Proposition 2.3, 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'_τ, such that $X_\tau > Y'_\tau$, and $Z_t^{Y'} = -\frac{Y'_t}{h(t)} > \max\left(-\left(\frac{C(1)}{k}\right)^{\frac{1}{k-1}}, -\left(\frac{C(c)}{k}\right)^{\frac{1}{k-1}}\right)$. Now, we should show that $\{\tau' = \infty\} \cap \{Y_t \to 0\} \cap \{Y'_t \to 0\}$ is non trivial. Take ϵ_1 and ϵ_c, both less than ϵ, as in the statement of Lemma 3.15 for Y_t and Y'_t respectively, and pick $\epsilon' = \min(\epsilon_1, \epsilon_c)$. For ϵ', using Lemma 3.15, we know we can find δ_1 and δ_c such that on $A_1 = \{G_t \in (-\epsilon', \epsilon') \text{ for all } t \in (s, s + \delta_1), \text{and } G_t \in (-2\epsilon', -\frac{9}{10}\epsilon') \text{ for all } t \in (s + \delta_1, \infty)\}$ we have $Y_s \to 0$, and on $A_c = \{G_t \in (-\epsilon', \epsilon') \text{ for all } t \in (s, s + \delta_c), \text{and } G_t \in (-2\epsilon', -\frac{9}{10}\epsilon') \text{ for all } t \in (s + \delta_c, \infty)\}$ we have $Y'_s \to 0$. From here, since $A_1 \cap A_c$ is non trivial, we only need to argue that $\{\tau' = \infty\} \supset A \cap A'$. From the remark of Lemma 3.15 we see that Y_t and Y'_t always stay below 0 on $A_1 \cap A_c$. Also, from Proposition 3.17, we see that $Z_t^{Y'} > -2\left(\frac{C(c)}{k}\right)^{\frac{1}{k-1}}$. Equivalently, and using that $h(t) \geq -\frac{3}{2}\left(\frac{k}{C(c)}\right)^{\frac{1}{k-1}} \epsilon$,

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

■

We now prove the second part of Theorem 1.10.

Proof Theorem 1.10 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 the exit time from $\mathcal{N}$, $\tau_e = \inf\{t | X_t \notin (-3\epsilon, 0)\}$. From Proposition 3.18, 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 (3.1.1) 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 from Proposition 2.3, we get $Y_t < X_t < Y'_t$, for all $t \in (\tau, \tau_e)$. Now, using Proposition 3.8, there is an event A such that $Y_t, Y'_t \in (-3\epsilon, 0)$, for all $t \geq \tau$. Consequentially $X_t \in (-3\epsilon, 0)$, for all $t \geq \tau$ since $\tau_e = \infty$ on A. Finally, using Lemma 2.2 we conclude that $Y_t \to 0$ on A, hence also $X_t \to 0$ on A. ■

Chapter 4

The discrete model

4.1 Analysis of X_n when $\frac{1}{2} + \frac{1}{2k} > \gamma$, $k > 1$ and $\gamma \in (1/2, 1)$

Throughout this section we assume $\frac{1}{2} + \frac{1}{2k} > \gamma$. We will prove Theorem 1.11 at the end of the subsection with the help of Theorem 4.1. The former theorem is a local version of the latter. For Theorem 4.1 we will need X_n that satisfy

$$X_{n+1} - X_n \geq \frac{|X_n|^k}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},\ k > 1 \text{ and } \gamma \in (1/2, 1), \tag{4.1.1}$$

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

Theorem 4.1. *Let $(X_n)_{n\geq 1}$ solve* (4.1.1). *Then $X_t \to \infty$ a.s..*

Now, we develop the necessary tools to prove this theorem.

Proposition 4.2. *Let* $(X_n)_{n\geq 1}$ *solve* (4.1.1)*. The process* $(X_n)_{n\geq 1}$ *gets close to the origin infinitely often. More specifically, for* $\beta < 0$ *the event* $\{X_n \geq \beta n^{\frac{1-2\gamma}{2}} \, i.o.\}$ *has probability* 1.

Proof: Now, from the restrictions on k we obtain

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

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,

$$\begin{aligned}
Z_{n+1} - Z_n &\geq -\frac{X_{n+1}}{h(n+1)} + \frac{X_n}{h(n)} && (4.1.2)\\
&\geq -X_n\left(\frac{1}{h(n+1)} - \frac{1}{h(n)}\right) - \frac{|X_n|^k}{n^\gamma h(n+1)} - \frac{Y_{n+1}}{n^\gamma h(n+1)}\\
&\geq X_n \frac{1-\gamma}{k-1}\xi_n^{-\frac{1-\gamma}{1-k}-1} - \frac{|X_n|^k}{n^\gamma h(n+1)} - \frac{Y_{n+1}}{n^\gamma h(n+1)}, \text{ where } \xi_n \in (n, n+1)\\
&\geq \frac{X_n}{h(n+1)n^\gamma}\left(\frac{1-\gamma}{k-1}\xi_n^{-\frac{1-\gamma}{1-k}-1} h(n+1)n^\gamma - \frac{|X_n|^k}{X_n}\right) - \frac{Y_{n+1}}{n^\gamma h(n+1)}\\
&\geq \frac{X_n}{h(n+1)n^\gamma}\left(-a_n\frac{1-\gamma}{k-1}|h(n)|^{k-1} - \frac{|X_n|^k}{X_n}\right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} && (4.1.3)\\
&\geq \frac{X_n}{h(n+1)n^\gamma}\left(-\frac{2(1-\gamma)}{k-1}|h(n)|^{k-1} - \frac{|X_n|^k}{X_n}\right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} && (4.1.4)
\end{aligned}$$

where

$$a_n = \frac{\xi_n^{\frac{-(1-\gamma)}{1-k}-1} h(n+1)n^\gamma}{-|h(n)|^{k-1}}. \tag{4.1.5}$$

To justify the inequality (4.1.4) for large enough n, notice that $a_n \to 1$.

Define $G'_{s,n} = \sum_{i=s}^{n-1} \frac{Yi+1}{ih(i+1)}$. As seen in 4.1

Lemma 4.3. $\limsup_{n\to\infty} G'_{1,n} = \infty$ *a.s.*.

Proof: We use the following theorem, for a reference see [Fis92] page 676 Theorem 1,

Theorem 4.4. *Let X_n be a martingale difference such that $E(X_i^2|\mathscr{F}_{i-1}) < \infty$. Set $s_n^2 = \sum_{i=1}^n E(X_i^2|\mathscr{F}_{i-1})$, and define $\phi(x) = (2\log_2(x^2 \vee e^2))^{\frac{1}{2}}$. We assume that $s_n \to \infty$ a.s. and that $|X_i| \leq \frac{K_i s_i}{\phi(s_i)}$ a.s. where K_i is $\mathscr{F}_{i-1}$ measurable with $\limsup_{i\to\infty} K_i < K$ for some constant K. Then there is a positive constant $\epsilon(K)$ such that $\limsup_{n\to\infty} \sum_{i=1}^n \frac{X_i}{s_n\phi(s_n)} \geq \epsilon(K)$ a.s..*

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

From Lemma 4.3, it is immediate that for any random time s (not necessarily a stopping time) $\limsup_{n\to\infty} G'_{s,n} = \infty$, a.s..

Now, we return to prove proposition 4.2. 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,

$$Z_n \geq Z_{n_0} + \sum_{i=n_0}^{n} \frac{X_n}{h(n+1)n^\gamma} \left(-\frac{2(1-\gamma)}{k-1} |h(n)|^{k-1} - \frac{|X_n|^k}{X_n} \right) - \frac{Y_{n+1}}{n^\gamma h(n+1)} + {G'}_{n_0,n}$$

$$> Z_{n_0} + {G'}_{n_0,n},$$

which gives $\limsup_{n\to\infty} Z_n = \infty$ which is a contradiction since this would imply $X_n \geq 0$.

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 $G_{n,u} = \sum_{i=n}^{u-1} \frac{Yi+1}{i^\gamma}$, which is an important quantity for the next lemma and the remaining of the section.

Lemma 4.5. *For any n large enough, we can find $a_1 > 0, \delta > 0$ such that $\mathbb{P}(\sup_{u\geq n} G_{n,u} \geq a_1 n^{\frac{1-2\gamma}{2}} | \mathscr{F}_n) > \delta$ and $\mathbb{P}(G_{n,\infty} \geq a_1 n^{\frac{1-2\gamma}{2}} | \mathscr{F}_n) > \delta$.*

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

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

Therefore,

$$\begin{aligned}
E((G_{\tau\wedge m})^2|\mathscr{F}_n) &\geq E((G_{\tau\wedge n})^2|\mathscr{F}_n) + c\mathbb{P}(\tau = \infty|\mathscr{F}_n)(n^{1-2\gamma} - (m-1)^{1-2\gamma}) \\
&= c\mathbb{P}(\tau = \infty|\mathscr{F}_n)(n^{1-2\gamma} - (m-1)^{1-2\gamma}). \qquad (4.1.6)
\end{aligned}$$

Notice that since 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 (4.1.6), we obtain

$$\frac{4a_2^2 n^{1-2\gamma}}{\epsilon c' n^{1-2\gamma}} = \frac{4a_2^2}{\epsilon c'} \geq \mathbb{P}(\tau = \infty|\mathscr{F}_n).$$

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

Now, we take any martingale M_n starting at 0, such that it exits an interval $(-2a, 2a)$, with at least probability p, and $|M_{n+1} - M_n| < a$, a.s.. Then, we stop the martingale upon exiting the interval $(-2a, 2a)$; namely, define τ_- to be the first time M_n goes below $-2a$ and τ_+ 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

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

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

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

For the second part of the lemma, we use the following inequality: let M_n be a martingale such that $M_0 = 0$ and $E(M_n^2) < \infty$. Then $\mathbb{P}(\max_{n \geq u \geq 0} M_u \geq \lambda) \leq \frac{E(M_n^2)}{E(M_n^2)+\lambda^2}$ (for a reference see [Dur13], page 213, exercise 5.4.5). Let τ be the first time $G_{n,u}$, surpasses $a_2 n^{1-2\gamma}$. Condition on $[\tau < \infty]$, and notice that $G_{n,\infty} \geq \frac{a_2}{2} n^{1-2\gamma}$ 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 gives

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

Therefore, $\mathbb{P}(G_{n,\infty} \geq \frac{a_2}{2} n^{\frac{1-2\gamma}{2}} | \mathscr{F}_n) \geq \mathbb{P}(\tau < \infty) \frac{\frac{a_2^2}{4}}{c + \frac{a_2^2}{4}}$, which concludes the lemma.
■

So, for any stopping time σ, we get the following version of the previous lemma:

Lemma 4.6. *For any n, we can find $a_1 > 0, \delta_1 > 0, \delta_2 > 0$ such that $\mathbb{P}(\sup_{u \geq \sigma} G_{\sigma,u} \geq a_1 \sigma^{\frac{1-2\gamma}{2}} | \mathscr{F}_\sigma) > \delta_1$ and $\mathbb{P}(G_{\sigma,\infty} \geq a_1 \sigma^{\frac{1-2\gamma}{2}} | \mathscr{F}_\sigma) > \delta_2$.*

■

Before proving Theorem 4.1 we will need the following corollary.

Corollary 4.7. *Let $(X_n)_{n \geq 1}$ solve* (4.1.1)*. The event $\{X_n \geq 0 \, i.o.\}$ 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 τ_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 the statement of Lemma 4.6 holds. From Proposition 4.2, we can do so, with all τ_n a.s. finite. Hence, $\mathbb{P}(\sup_{\infty \geq u \geq \tau_n} X_m - X_{\tau_n} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} | \mathscr{F}_{\tau_n}) \geq \mathbb{P}(\sup_{\infty \geq u \geq \tau_n} G_{\tau_n,u} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} | \mathscr{F}_{\tau_n}) > \delta_1 > 0$. Therefore, by Borel-Cantelli on the events $\{X_{\tau_n} \geq \beta \tau_n^{\frac{1-2\gamma}{2}}\}$, we get $\{X_{\tau_n} \geq 0 \text{ i.o.}\}$. Therefore $\{X_n \geq 0 \text{ i.o.}\}$ holds a.s. ■

Proof of Theorem 4.1: Define τ_n, as in the proof of the previous corollary, such that $X_{\tau_n} \geq 0$. Since $\mathbb{P}(G_{\tau_n,\infty} \geq a_1 \tau_n^{\frac{1-2\gamma}{2}} | \mathscr{F}_{\tau_n}) > \delta_2$, 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 a.s. there are constants $c(\omega) > 0$ $m(\omega)$ such that $\{X_n > c \text{ for all } n \geq m\} = \{\liminf_{\to\infty} X_n > 0\}$. Indeed, if we define $\tau_0 = 0$ and $\tau_{n+1} = \inf\{m > \tau_n + 1 | X_m \geq \frac{a_1}{2} m^{\frac{1-2\gamma}{2}}\}$ we see that $\tau_n < \infty$ a.s. and $\tau_n \to \infty$. This gives a corresponding filtration, namely $\mathscr{F}_n = \sigma(\tau_n)$.

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

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

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

Finally, we can prove Theorem 1.11. In the next proof X_n, X_n' solve (1.8.1) and (4.1.1), respectively.

Proof of Theorem 1.11: We define, $\tau = \inf\{n | X_n \in (-\epsilon, \epsilon)\}$, and $\tau' = \inf\{\tau' > \tau | X_n \notin (-\epsilon', \epsilon')\}$. When ϵ 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 (4.3.1). Since $X'_n \to \infty$, a.s., we have that $\tau' < \infty$ a.s.. Thus, on $\{\lim_{n\to\infty} X_n = 0\}$, Borel-Cantelli implies $\{X_n = 0 \text{ i.o.}\}$. Therefore, $\mathbb{P}(\lim_{n\to\infty} X_n = 0) = 0$. ■

4.2 Analysis of X_n when $\frac{1}{2} + \frac{1}{2k} < \gamma$, $k > 1$ and $\gamma \in (1/2, 1)$

Before proving the main theorem Theorem 1.12, as described in subsection 1.8 we will study a process $(X_n)_{n\geq 1}$ that satisfies

$$X_{n+1} - X_n \leq \frac{f(x)}{n^\gamma} + \frac{Y_{n+1}}{n^\gamma},\ \gamma \in (1/2, 1),\ k \in (1, \infty), \tag{4.2.1}$$

where, $f(x) \leq |x|^k$ when $x \in (-\epsilon, \epsilon)$, and $f(x) = |x|^k$ when $x \in \mathbb{R} \setminus (-\epsilon, \epsilon)$. Also, as before, $|Y_n| < M$ a.s. and $E(Y_{n+1} | \mathscr{F}_n) = 0$. Let $x_0 < 0$, such that $f(x) > M$, $\forall x \leq x_0$. We will use x_0 in the next lemma.

Lemma 4.8. *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 $X_n \geq -C$. Then

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

$\blacksquare$

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 the previous subsection 4.1, first appearing in (4.1.3) and defined in (4.1.5).

Claim: *Any n_0 large enough satisfies the following properties*

1. $a_n > 1/2$, $n \geq n_0$.
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. $\mathbb{P}({G'}_{n_0,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq n_0 | \mathscr{F}_{n_0}) > 0$.

Proof:

1. This is is trivial.

2. Since $|Y_n| < M$, and $X_n > C$ a.s., then 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 $\dfrac{h(n)}{h(n+1)} \to 1$.

   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 2.

3. Using Doob's inequality, and the fact that $m^\gamma h(m+1) \sim m^{\frac{1-\gamma}{k-1}-\gamma} \leq m^{\frac{1-\gamma}{k-1}-\gamma} \leq m^{\frac{-1-\delta}{2}}$ for some $\delta > 0$, we have:

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

$\blacksquare$

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

Lemma 4.9. *Let τ be a stopping time such that $\tau \geq n_0$, where n_0 is the same as in the previous claim. Then, $\mathbb{P}(G'_{\tau,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq \tau | \mathscr{F}_\tau) > 0$*

$\blacksquare$

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 4.10. *Let $(X_n)_{n\geq 1}$ that satisfies (4.2.1). When $\tau < \infty$, with positive probability, then $\mathbb{P}(X_n \to 0) > 0$. 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 $Z_n = -\dfrac{X_n}{h(n)}$ and obtain, as done in (4.1.3), expect, 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 $D_n = \dfrac{X_n}{h(n+1)n^\gamma}\left(-\dfrac{1-\gamma}{2(k-1)}|h(n)|^{k-1} - \dfrac{|X_n|^k}{X_n}\right)$. Then we have

$$Z_m - Z_\tau \leq \sum_{i=\tau}^{m-1} D_i + G'_{\tau,m}. \tag{4.2.2}$$

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 ϵ, 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 $\{\sigma < \infty\} \cap A$ also does. 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, $X_n < 0$ for all $n > \tau$, therefore $\limsup_{n\to\infty} X_n \leq 0$ on A. However, by Lemma 2.5 we have $\limsup_{n\to\infty} X_n \geq 0$ a.s.. Therefore, on A, $X_n \to 0$. ■

Remark: 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 1.12: We define $\tau = \inf\{n \geq n_0 | X_n \in (-\epsilon_2, -\epsilon_1)\}$, where n_0 is the same as in Lemma 4.9, and $\tau_e = \inf\{n | X_n \notin (-3\epsilon, 3\epsilon)\}$. Let $(X'_n : n \geq \tau)$ be a process that satisfies (4.1.1). Then we couple (X_n) with (X'_n) on $\{\tau < \infty\}$ such 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 4.10 are met. The only thing we need to check is that $Z'_\tau = -\frac{X'_\tau}{h(\tau)} < -2\epsilon$. However, since $h(t) \to 0$ this is always possible by choosing n_0 large enough. Furthermore, by Proposition 4.10, we see that there is an event of positive probability such that $X'_n \to 0$, where τ_e is infinite conditioned on this event. Therefore, X_n converges to 0 with positive probability. ∎

4.3 Analysis of X_t when $k > 1/2$ and $\gamma = 1$

Throughout this section we assume $k > 1/2$. We will prove the second part of theorem 1.11. This section will be almost identical to section 4.1. As before first we will prove a local version of the latter theorem. For Theorem 4.11 we will need X_n that satisfy

$$X_{n+1} - X_n \geq k\frac{|X_n|}{n} + \frac{Y_{n+1}}{n}, \ k > 1/2, \tag{4.3.1}$$

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

Theorem 4.11. *Let $(X_n)_{n\geq 1}$ solve* (4.3.1). *Then $X_n \to \infty$ a.s..*

We have the following proposition corresponding to Proposition 4.2.

Proposition 4.12. *Let $(X_n)_{n\geq 1}$ solve* (4.1.1). *The process $(X_n)_{n\geq 1}$ gets close to the origin infinitely often. More specifically, for $\beta < 0$ the event $\{X_n \geq \beta n^{-\frac{1}{2}} i.o.\}$ has probability* 1.

Proof: Set $h(t) = -t^{-k}$, and define $Z_n = -\frac{X_n}{h_n}$. Also we pick $1/2 < k_1 < k$. From here, on the event $\{X_m < 0 \,\text{for all}\, m \geq n\}$, we get the following recursion,

$$
\begin{aligned}
Z_{n+1} - Z_n &\geq -\frac{X_{n+1}}{h(n+1)} + \frac{X_n}{h(n)} && (4.3.2)\\
&\geq -X_n\left(\frac{1}{h(n+1)} - \frac{1}{h(n)}\right) - k\frac{|X_n|}{nh(n+1)} - \frac{Y_{n+1}}{nh(n+1)}\\
&\geq X_n k\xi_n^{k-1} - k\frac{|X_n|}{nh(n+1)} - \frac{Y_{n+1}}{nh(n+1)}, \text{ where } \xi_n \in (n, n+1)\\
&\geq \frac{X_n}{h(n+1)n}\left(k\xi_n^{k-1}h(n+1)n - k\frac{|X_n|}{X_n}\right) - \frac{Y_{n+1}}{nh(n+1)}\\
&\geq \frac{X_n}{h(n+1)n}\left(-a_n k - k\frac{|X_n|}{X_n}\right) - \frac{Y_{n+1}}{nh(n+1)} && (4.3.3)\\
&\geq \frac{X_n}{h(n+1)n}\left(-k_1 - k\frac{|X_n|}{X_n}\right) - \frac{Y_{n+1}}{nh(n+1)} && (4.3.4)\\
&\geq \frac{X_n}{h(n+1)n}(k - k_1) - \frac{Y_{n+1}}{nh(n+1)} && (4.3.5)
\end{aligned}
$$

where

$$a_n = -\xi_n^{k-1}h(n+1)n. \tag{4.3.6}$$

To justify the inequality (4.3.4) for large enough n, notice that $a_n \to 1$.

Define $G'_{s,n} = \sum_{i=s}^{n-1} \frac{Y i+1}{ih(i+1)}$ as done in section 4.1 To this end, we have the following lemma,

Lemma 4.13. $\limsup_{n\to\infty} G'_{1,n} = \infty$ *a.s.*.

Proof: Same proof as Lemma 4.3. ■

From Lemma 4.13, it is immediate that for any random time s (not necessarily a stopping time) $\limsup_{n\to\infty} G'_{s,n} = \infty$, a.s..

Now, we return to prove proposition 4.12. Assume that there is n_0 such that $X_n < 0$, for all $n \geq n_0$. Then,

$$
\begin{aligned}
Z_n &\geq Z_{n_0} + \sum_{i=n_0}^{n} \frac{X_i}{h(i+1)i}(k-k_1) + G'_{n_0,n} \\
&> Z_{n_0} + G'_{n_0,n},
\end{aligned}
$$

which gives $\limsup_{n\to\infty} Z_n = \infty$ which is a contradiction since this would imply $X_n \geq 0$.

■

We define $G_{n,u} = \sum_{i=n}^{u-1} \frac{Y i+1}{i^\gamma}$. We are almost done since, the noise estimates done in section 4.1 was a big bulk of the work and they also hold when $\gamma = 1$.

Before proving Theorem 4.1 we will need the following corollary.

Corollary 4.14. *Let* $(X_n)_{n\geq 1}$ *solve* (4.3.1). *The event* $\{X_n \geq 0 \, i.o.\}$ *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 τ_n, going to infinity a.s., such that

$X_{\tau_n} \geq \beta \tau_n^{-\frac{1}{2}}$ for $|\beta| < a_1$, where a_1 is such that the statement of Lemma 4.6 holds. From Proposition 4.2, we can do so, with all τ_n a.s. finite. Hence, $\mathbb{P}(\sup_{\infty \geq u \geq \tau_n} X_m - X_{\tau_n} \geq a_1 \tau_n^{-\frac{1}{2}} | \mathscr{F}_{\tau_n}) \geq \mathbb{P}(\sup_{\infty \geq u \geq \tau_n} G_{\tau_n,u} \geq a_1 \tau_n^{-\frac{1}{2}} | \mathscr{F}_{\tau_n}) > \delta_1 > 0$. Therefore, by Borel-Cantelli on the events $\{X_{\tau_n} \geq \beta \tau_n^{-\frac{1}{2}}\}$, we get $\{X_{\tau_n} \geq 0 \text{ i.o.}\}$. Therefore $\{X_n \geq 0 \text{ i.o.}\}$ holds a.s. ■

Proof of Theorem 4.11: Define τ_n, as in the proof of the previous corollary, such that $X_{\tau_n} \geq 0$. Since $\mathbb{P}(G_{\tau_n,\infty} \geq a_1 \tau_n^{-\frac{1}{2}} | \mathscr{F}_{\tau_n}) > \delta_2$, an application of Borel-Cantelli shows that $\{X_n \geq \frac{a_1}{2} n^{-\frac{1}{2}} \text{ i.o.}\}$ holds a.s.. We claim a.s. there are constants $c(\omega) > 0$ $m(\omega)$ such that $\{X_n > c \text{ for all } n \geq m\} = \{\liminf_{\to\infty} X_n > 0\}$. Indeed, if we define $\tau_0 = 0$ and $\tau_{n+1} = \inf\{m > \tau_n + 1 | X_m \geq \frac{a_1}{2} m^{-\frac{1}{2}}\}$ we see that $\tau_n < \infty$ a.s. and $\tau_n \to \infty$. This gives a corresponding filtration, namely $\mathscr{F}_n = \sigma(\tau_n)$.

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

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

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

Finally, we can prove Theorem 1.11. In the next proof X_n, X'_n solve (1.8.1) and (4.3.1), respectively.

Proof of Theorem 1.11: We define, $\tau = \inf\{n | X_n \in (-\epsilon, \epsilon)\}$, and $\tau' = \inf\{\tau' > \tau | X_n \notin (-\epsilon', \epsilon')\}$. When ϵ 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 (4.1.1). Since $X'_n \to \infty$, a.s., we have that $\tau' < \infty$ a.s.. Thus, on $\{\lim_{n\to\infty} X_n = 0\}$, Borel-Cantelli implies $\{X_n = 0 \text{ i.o.}\}$. Therefore, $\mathbb{P}(\lim_{n\to\infty} X_n = 0) = 0$. ■

4.4 Analysis of X_n when $k < 1/2$ and $\gamma = 1$

Before proving the second part of the main theorem Theorem 1.12, as described in subsection 1.8 we will study a process $(X_n)_{n\geq 1}$ that satisfies

$$X_{n+1} - X_n \leq \frac{f(x)}{n} + \frac{Y_{n+1}}{n}, k < 1/2, \tag{4.4.1}$$

where, $f(x) \leq k|x|$ when $x \in (-\epsilon, \epsilon)$, and $f(x) = k|x|$ when $x \in \mathbb{R} \setminus (-\epsilon, \epsilon)$. Also, as before, $|Y_n| < M$ a.s. and $E(Y_{n+1}|\mathscr{F}_n) = 0$. Let $x_0 < 0$, such that $f(x) > M$, $\forall x \leq x_0$. We will use x_0 in the next lemma.

Lemma 4.15. *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

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

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

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

■

Let a_n, be defined as in the previous subsection 4.3, first appearing in (4.3.3) and defined in (4.3.6).

Claim: *Any n_0 large enough satisfies the following properties*

1. $a_n > 1/2$, $n \geq n_0$.

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. $\mathbb{P}({G'}_{n_0,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq n_0 | \mathscr{F}_{n_0}) > 0$.

Proof:

1. This is is trivial.

2. Since $|Y_n| < M$, and $X_n > C$ a.s., then whenever $X_n < 0$, we have $|X_{n+1} - X_n| = O(n^{-1})$ and $n^{-1} = o(h(n))$. Furthermore, notice that $\dfrac{h(n)}{h(n+1)} \to 1$.

   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 2.

3. Using Doob's inequality, and the fact that $mh(m+1) \sim m^{1-k} \leq m^{\frac{-1-\delta}{2}}$ for some $\delta > 0$, we have:

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

■

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

Lemma 4.16. *Let τ be a stopping time such that $\tau \geq n_0$, where n_0 is the same as in the previous claim. Then,* $\mathbb{P}(G'_{\tau,n} \in (\frac{-\epsilon}{2}, \frac{\epsilon}{2}) \forall n \geq \tau | \mathscr{F}_\tau) > 0$

■

Let $\epsilon > 0$, and define a stopping time $\tau = \inf\{n \geq n_0 | Z_n < -2\epsilon\}$.

Proposition 4.17. *Let $(X_n)_{n\geq 1}$ that satisfies* (4.2.1). *When $\tau < \infty$, with positive probability, then $\mathbb{P}(X_n \to 0) > 0$. More specifically, the process $(X_n : n \geq \tau)$ converges to zero with positive probability.*

Proof: Pick $k_2 > k$. On the event $\{X_m < 0 \, \text{for all} \, m \geq n\}$ we use the expression for $Z_n = -\dfrac{X_n}{h(n)}$ and obtain, as done in (4.3.3), expect, now the inequalities are reversed,

$$Z_{n+1} - Z_n \leq \frac{X_n}{h(n+1)n}\left(-a_n k - k\frac{|X_n|}{X_n}\right) - \frac{Y_{n+1}}{nh(n+1)} \tag{4.4.2}$$

$$\leq \frac{X_n}{h(n+1)n}\left(-k_2 - k\frac{|X_n|}{X_n}\right) - \frac{Y_{n+1}}{nh(n+1)} \tag{4.4.3}$$

$$\leq \frac{X_n}{h(n+1)n}(k - k_2) - \frac{Y_{n+1}}{nh(n+1)} \tag{4.4.4}$$

Set $D_n = \frac{X_n}{h(n+1)n}(k - k_2)$. Then we have

$$Z_m - Z_\tau \leq \sum_{i=\tau}^{m-1} D_i + G'_{\tau,m}. \tag{4.4.5}$$

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 $D_n < 0$ whenever $Z_n \geq -2\epsilon$. If $\{\tau_0 < \infty\} \cap A$ has positive probability, then $\{\sigma < \infty\} \cap A$ also does. 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, $X_n < 0$ for all $n > \tau$, therefore $\limsup_{n\to\infty} X_n \leq 0$ on A. However, by Lemma 2.5 we have $\limsup_{n\to\infty} X_n \geq 0$ a.s.. Therefore, on A, $X_n \to 0$. ■

Remark: 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 1.12: We define $\tau = \inf\{n \geq n_0 | X_n \in (-\epsilon_2, -\epsilon_1)\}$, where n_0 is the same as in Lemma 4.9, and $\tau_e = \inf\{n | X_n \notin (-3\epsilon, 3\epsilon)\}$. Let $(X'_n : n \geq \tau)$ be a process that satisfies (4.1.1). Then we couple (X_n) with (X'_n) on $\{\tau < \infty\}$ such 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

4.17 are met. The only thing we need to check is that $Z'_\tau = -\frac{X'_\tau}{h(\tau)} < -2\epsilon$. However, since $h(t) \to 0$ this is always possible by choosing n_0 large enough. Furthermore, by Proposition 4.17, we see that there is an event of positive probability such that $X'_n \to 0$, where τ_e is infinite conditioned on this event. Therefore, X_n converges to 0 with positive probability. ■

Bibliography

[AAZB+17] Naman Agarwal, Zeyuan Allen-Zhu, Brian Bullins, Elad Hazan, and Tengyu Ma. Finding approximate local minima faster than gradient descent. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 1195–1199, New York, NY, USA, 2017. ACM.

[AG16] Animashree Anandkumar and Rong Ge. Efficient approaches for escaping higher order saddle points in non-convex optimization. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 81–102, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[BR95] Robert W. Brennan and Paul Rogers. Stochastic optimization applied to a manufacturing system operation problem. In *Proceedings of the 27th Conference on Winter Simulation*, WSC '95, pages 857–864, Washington, DC, USA, 1995. IEEE Computer Society.

[CdlTTZ16] Xi Chen, jason d. lee, Xin T. Tong, and Yichen Zhang. Statistical inference for model parameters in stochastic gradient descent, 10 2016.

[CHM+15] Anna Choromanska, Mikael Henaff, Michael Mathieu, Grard Ben Arous, and Yann LeCun. The loss surfaces of multilayer networks. *Journal of Machine Learning Research*, 38:192–204, 2015.

[DHS11] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, July 2011.

[DKLH18] Hadi Daneshmand, Jonas Kohler, Aurelien Lucchi, and Thomas Hofmann. Escaping saddles with stochastic gradients. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1155–1164, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[Dur13] R. Durrett. *Probability: Theory and Examples.* Duxbury Press, 2013.

[DZ12] Matthew D. Zeiler. Adadelta: An adaptive learning rate method. 1212, 12 2012.

[EDM04] Eyal Even-Dar and Yishay Mansour. Learning rates for q-learning. *J. Mach. Learn. Res.*, 5:1–25, December 2004.

[EW88] Y.M. Ermoliev and R.J.-B. Wets. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Heidelberg, 1988.

[Fis92] Evan Fisher. On the law of the iterated logarithm for martingales. *Ann. Probab.*, 20(2):675–680, 1992.

[GHJY15] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points — online stochastic gradient for tensor decomposition. In Peter Grnwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 797–842, Paris, France, 03–06 Jul 2015. PMLR.

[GM91] Saul B. Gelfand and Sanjoy K. Mitter. Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$. *SIAM Journal on Control and Optimization*, 29(5):999–1018, 1991.

[HL13] Christopher J. Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *J. ACM*, 60(6):45:1–45:39, November 2013.

[HLS80] Bruce M. Hill, David Lane, and William Sudderth. A strong law for some generalized urn processes. *Ann. Probab.*, 8(2):214–226, 1980.

[JGN+17] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan. How to Escape Saddle Points Efficiently. *ArXiv e-prints*, March 2017.

[JJKN15] Prateek Jain, Chi Jin, Sham M. Kakade, and Praneeth Netrapalli. Computing matrix squareroot via non convex local search. *CoRR*, abs/1507.05854, 2015.

[Kes58] Harry Kesten. Accelerated stochastic approximation. *Ann. Math. Statist.*, 29(1):41–59, 03 1958.

[KW52] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[KY03] H. Kushner and G.G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Stochastic Modelling and Applied Probability. Springer New York, 2003.

[LLKC18] Tianyang Li, Liu Liu, Anastasios Kyrillidis, and Constantine Caramanis. Statistical inference using sgd, 2018.

[LSJR16] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht. Gradient Descent Converges to Minimizers. *ArXiv e-prints*, February 2016.

[NES83] Y. E. NESTEROV. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269:543–547, 1983.

[Nes00] Yurii Nesterov. *Squared Functional Systems and Optimization Problems*, pages 405–440. Springer US, Boston, MA, 2000.

[Pem90] Robin Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Probab.*, 18(2):698–712, 1990.

[Pem91] Robin Pemantle. When are touchpoints limits for generalized Pólya urns? *Proc. Amer. Math. Soc.*, 113(1):235–243, 1991.

[Pem07] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surveys*, 4:1–79, 2007.

[PJ92] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992.

[Pol64] Boris Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr Computational Mathematics and Mathematical Physics*, 4:1–17, 12 1964.

[RB54] Julius R. Blum. Approximation methods which converge with probability one. *Ann. Math. Statist*, 25, 06 1954.

[RM51] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[RRT17] M. Raginsky, A. Rakhlin, and M. Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. *ArXiv e-prints*, February 2017.

[RSS12] Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pages 1571–1578, USA, 2012. Omnipress.

[Rud16] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[Rup88] D. Ruppert. Efficient estimations from a slowly convergent robbins-monro process. *Technical report, Cornell University Operations Research and Industrial Engineering, Cornell University Operations Research and Industrial Engineering*, 1988.

[Rup19] David Ruppert. Efficient estimators from a slowly converging robbins-monro process. 03 2019.

[RWW87] L. C. G. Rogers, D. Williams, and John Wiley. *Diffusions, Markov processes and Martingales, vol 2: Ito calculus.* 1987.

[SL87] Rajan Suri and Ying Tat Leung. Single run optimization of a siman model for closed loop flexible assembly systems. In *Proceedings of the*

19th Conference on Winter Simulation, WSC '87, pages 738–748, New York, NY, USA, 1987. ACM.

[SL16] R. Sun and Z. Q. Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, Nov 2016.

[Son12] Pham Tin Son. An explicit bound for the ojasiewicz exponent of real polynomials. *Kodai Math. J.*, 35(2):311–319, 2012.

[Spr] Springer Verlag GmbH, European Mathematical Society. Encyclopedia of Mathematics. Website. URL:https://www.encyclopediaofmath.org/index.php/Lojasiewicz_inequality. Accessed on 2017-10-11.

[SQW17] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere i: Overview and the geometric picture. *IEEE Transactions on Information Theory*, 63(2):853–884, Feb 2017.

[SR96] Sharoni Shafir and Jonathan Roughgarden. Adaptive individuals in evolving populations. chapter The Effect of Memory Length on Individual Fitness in a Lizard, pages 173–186. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1996.