Applied Probability Trust (1 October 2023)

ERROR BOUNDS FOR ONE-DIMENSIONAL CONSTRAINED LANGEVIN APPROXIMATIONS FOR NEARLY DENSITY DEPENDENT MARKOV CHAINS

FELIPE A. CAMPOS,* **
RUTH J. WILLIAMS,* *** *University of California, San Diego.*

Abstract

Continuous time Markov chains are frequently used to model the stochastic dynamics of (bio)chemical reaction networks. However, except in very special cases, they cannot be analysed exactly. Additionally, simulation can be computationally intensive. An approach to address these challenges is to consider a more tractable diffusion approximation. In [17], Leite and Williams proposed a reflected diffusion as an approximation for (bio)chemical reaction networks, which they called the Constrained Langevin Approximation (CLA) as it extends the usual Langevin approximation beyond the first time some chemical species becomes zero in number. Further explanation and examples of the CLA can be found in Anderson et al. [2].

In this paper, we extend the approximation of [17] to (nearly) density dependent Markov chains, as a first step to obtaining error estimates for the CLA, when the diffusion state space is one-dimensional and we provide a bound for the error in a strong approximation. We discuss some applications for chemical reaction networks and epidemic models, and illustrate these with examples. Our method of proof is designed to generalize to higher dimensions, provided there is a Lipschitz Skorokhod map defining the reflected diffusion process. The existence of such a Lipschitz map is an open problem in dimensions more than one.

Keywords: Density dependent Markov chains; diffusion approximation; constrained Langevin approximation; error bound; linear noise approximation; chemical reaction networks; stochastic differential equation with oblique reflection; systems biology.

2020 Mathematics Subject Classification: Primary 60J27; 60J60; 65C30
Secondary 60H10; 92C42; 92C45

* Postal address: Department of Mathematics, University of California - San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0112. USA
** Email address: fcamposv91@gmail.com
*** Email address: rjwilliams@ucsd.edu

Contents

1. Introduction

1.1. Background

Stochastic dynamics in chemistry and biology are often modeled by continuous-time Markov chains. While intuitive to formulate, these models often cannot be analysed exactly and can be computationally intensive to simulate. A way to tackle this problem is by considering a diffusion approximation, i.e., a continuous-path strong Markov process that approximates the Markov chain model. The diffusion approximations are usually more amenable to analysis, involving differential equations rather than difference equations (see Anderson et al. [2] and Linetsky [18]). Also, their simulation is typically less computationally intensive (see e.g., Leite & Williams [17]).

In the seminal work of Kurtz [16], a Langevin-type diffusion approximation was proposed under the assumption of *density dependence* for the Markov chain. Density dependent Markov chains were introduced earlier by Kurtz in [14] as a class of Markov processes which keep track of the number of individuals in a stochastic system, where a volume, area or total population parameter is considered. Kurtz [14] showed that under a suitable rescaling using this parameter, these Markov chains exhibit a functional Law of Large Numbers (LLN), where the deterministic limit is a solution of an ordinary differential equation. In [15], Kurtz proved a functional Central Limit Theorem (CLT) for these processes, rigorously establishing the Linear Noise Approximation (LNA) or van Kampen Approximation. Later, in [16], Kurtz developed the *Langevin diffusion approximation* for density dependent processes, which is valid until the first time that the concentration for a given species of individuals reaches zero. An extension of this approximation beyond such a stopping time, called a Constrained Langevin Approximation, is the object of attention in this paper.

A slight generalization of density dependence called *nearly density dependence* appears in Ethier & Kurtz ([9], chapter 11, equation 1.13) although the term was later introduced by Angius et al. [4], who proposed a hybrid diffusion approximation for nearly density dependent processes. Nearly density dependent processes include an important collection of Markov chain models called *chemical reaction networks*, which will be one of our main sources of examples.

Chemical reaction networks (CRNs) are models used to describe the stochastic dynamics of a chemical system (see Anderson & Kurtz [3] for an introduction to this subject). In these models, we have $d \geq 1$ different species subject to $n \geq 1$ different reactions. The system has a parameter $r \geq 1$, thought of as volume, area or total population. As time goes by, reactions are triggered randomly. A continuous-time Markov chain $X^r(t) = (X_1^r(t), \ldots, X_d^r(t))$ tracks the count for each species through time. Throughout this paper we will consider models of CRNs satisfying stochastic mass action kinetics and with rates scaled by r under the classical scaling (see Sections 2.1 and 2.2 of [2] for an exposition of these concepts). Using the results in Kurtz [16], it is possible to construct a Langevin diffusion approximation Y^r for the concentration process $\overline{X}^r := \frac{1}{r}X^r$ until the first time that Y^r reaches the boundary of the nonnegative d-dimensional orthant, $\mathbb{R}_+^d$.

In [17], Leite & Williams proposed an extension Z^r of this stopped Langevin approximation, which is defined for all time and which is such that when it hits the boundary of $\mathbb{R}_+^d$, it is instantaneously pushed back into $\mathbb{R}_+^d$. Leite & Williams called this the *Constrained Langevin Approximation (CLA)*, since it behaves like the Langevin

approximation in the interior of the orthant and it is constrained to remain in the orthant for all time. Leite & Williams proved, under some assumptions on the CRN, that the CLA is well defined and that it can be obtained as a weak limit of certain jump diffusion extensions of the Langevin approximation. However, no bounds were given for the error in their approximation. We note that although the CLA was proposed for chemical reaction networks, it can be formulated for nearly density dependent processes as well.

1.2. Overview of this paper

In this paper, we provide error bounds for the Constrained Langevin Approximation to nearly density dependent Markov chains when the diffusion state-space is either a bounded interval or the nonnegative halfline. While much of our method of proof generalizes to higher dimensions, one element, namely the Lipschitz continuity of the Skorokhod map (see Appendix B) is not generally known for smoothly varying oblique reflection vector fields on the boundary of $\mathbb{R}^d_+$ for $d > 1$.

We first give conditions under which our approximation exists, allowing for Hölder-continuous dispersion coefficients that could vanish at the boundary, extending the work of Leite & Williams [17] when $d = 1$. Then, we construct a coupling of the CLA and the Markov chain, where an explicit bound is given for the distance between the paths of these processes. This coupling is often referred to as a strong approximation.

The construction of this coupling provides a template for how we could estimate error bounds for diffusion approximations in higher dimensions. Despite the limitation to a one-dimensional setting, more complex models can sometimes be reduced to one dimension, for example, due to the presence of conservation laws or due to multiscaling.

The paper is organized as follows. We begin in Section 1.3 by establishing the notation that will be used throughout the paper. In Section 2, we give the precise definitions of a nearly density dependent family and its associated Constrained Langevin Approximation (CLA). In Section 3, we describe the main results in this work for the two cases: when the diffusion state space is a bounded interval and when it is the nonnegative halfline. We provide examples of our results applied to chemical reaction networks and epidemic models in Section 4. In Section 5, we give the notion of a Stochastic Differential Equation with Reflection (SDER) and use this in establishing existence and uniqueness for the CLA. In Section 6 we provide proofs of our main results and corollaries. Two of the main ingredients of the proofs are a variant of the Komlós, Major & Tusnády (KMT) approximation [11, 12] (see Appendix A) and the known existence of a Lipschitz map defining the solution of the Skorokhod Problem on a bounded or unbounded real interval (see Appendix B).

1.3. Preliminaries and notation

For any integer $d \geq 1$, let $\mathbb{R}^d$ denote the d-dimensional Euclidean space. We usually write $\mathbb{R}$ for $\mathbb{R}^1$. For $x \in \mathbb{R}^d$, let $|x| = (\sum_{i=1}^d x_i^2)^{1/2}$ be the usual Euclidean norm. We denote by $\mathbb{R}^d_+$ the set of vectors $x \in \mathbb{R}^d$ such that $x_i \geq 0$ for $i = 1, \ldots, d$. The subset of $\mathbb{R}^d_+$ consisting of vectors with integer entries will be denoted by $\mathbb{Z}^d_+$.

We denote by $\mathcal{C}$ the set of continuous functions $x : [0, \infty) \longrightarrow \mathbb{R}$, equipped with the topology of uniform convergence on compact sets. This topology makes $\mathcal{C}$ a Polish space (metrizable, complete and separable). We endow $\mathcal{C}$ with its Borel σ-algebra $\mathcal{M}$. We denote by $\mathcal{D}$ the set of functions $x : [0, \infty) \longrightarrow \mathbb{R}$ that are right-continuous on $[0, \infty)$ with finite left-hand limits on $(0, \infty)$. The space $\mathcal{D}$ will be equipped with

Skorokhod's J_1-topology, which makes $\mathcal{D}$ a Polish space. We also denote by $\mathcal{M}$ the Borel σ-algebra associated with this topology. For $T > 0$, we denote by $\mathcal{D}[0,T]$ the set of functions $x : [0,T] \longrightarrow \mathbb{R}$ that are right-continuous on $[0,T)$ with finite left-hand limits on $(0,T]$. The space $\mathcal{D}[0,T]$ will be equipped with Skorokhod's J_1-topology, which also makes $\mathcal{D}[0,T]$ a Polish space. The space $\mathcal{D}[0,T]$ is endowed with its Borel σ-algebra $\mathcal{M}_T$. We denote by $\mathcal{C}[0,T]$ the subset of continuous functions in $\mathcal{D}[0,T]$. For any $T > 0$ and $x \in \mathcal{D}$, let $\|x\|_T = \sup_{0\le t\le T} |x(t)|$. For an integer $k \ge 1$ and an interval $I \subseteq \mathbb{R}$ (possibly unbounded), we denote by $\mathcal{C}^k(I)$ the set of functions $f : I \longrightarrow \mathbb{R}$ for which the first k derivatives exist on I and define continuous functions there.

A d-dimensional process $\{B(t),\ 0 \le t < \infty\}$ will be called a d-dimensional Brownian motion if it has continuous paths, independent increments and for every $0 \le s < t$ the increment $B(t)-B(s)$ is distributed as a multivariate Gaussian vector with mean vector 0 and covariance matrix $(t-s)I_d$, where I_d is the $d \times d$ identity matrix. In particular, the component processes of a d-dimensional Brownian motion are independent. We will call the d-dimensional Brownian motion *standard*, if in addition $B(0) = 0$ a.s.

Given a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, denote by $\mathcal{N}$ the collection of $\mathbb{P}$-null sets in $\mathcal{F}$. A filtration $\{\mathcal{F}_t : 0 \le t < \infty\}$ is an increasing family of sub-σ-algebras of $\mathcal{F}$. A quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, called a filtered probability space, is said to satisfy the usual conditions if $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space and $\{\mathcal{F}_t\}$ is a filtration such that $\mathcal{F}_0$ contains $\mathcal{N}$ and $\mathcal{F}_t = \cap_{t<u}\mathcal{F}_u$ for every $t \ge 0$. A d-dimensional process $\{X(t),\ 0 \le t < \infty\}$ will be called an $\{\mathcal{F}_t\}$-martingale if X is adapted to $\{\mathcal{F}_t\}$ and each component is a martingale with respect to $\{\mathcal{F}_t\}$. All of the stochastic processes considered in this paper have sample paths that are right-continuous with finite left-hand limits, and some will have continuous sample paths.

Let $(\mathcal{S}, d)$ be a complete and separable metric space, endowed with its Borel σ-algebra $\mathcal{H}$. For P and $\tilde{P}$, probability measures on $(\mathcal{S}, \mathcal{H})$, we denote by $\Pi(P, \tilde{P})$ the set of all probability measures on the product space $(\mathcal{S}\times\mathcal{S}, \mathcal{H}\otimes\mathcal{H})$ whose first and second marginals are P and $\tilde{P}$, respectively. Furthermore, for $p \in [1,\infty)$, $\mathcal{W}_p$ will denote the Wasserstein distance, defined by:

$$\mathcal{W}_p(P,\tilde{P}) = \inf_{\pi\in\Pi(P,\tilde{P})} \left(\int_{\mathcal{S}\times\mathcal{S}} d(x,y)^p d\pi(x,y)\right)^{1/p}.$$

2. Key stochastic processes

2.1. Nearly density dependent families

Let $\mathcal{I} \subseteq \mathbb{R}_+$ be a closed interval. In this article, we consider two cases: $\mathcal{I}$ is either $[0,a]$, where $a > 0$, or $[0,\infty)$. Consider a positive parameter $r > 0$, which can be thought, for example, as representing volume, area or total population. Define,

$$\mathcal{I}^r := r\mathcal{I} \cap \mathbb{Z} = \{k \in \mathbb{Z} \mid k = rx, \ \text{ for some } x \in \mathcal{I}\},$$
$$\overline{\mathcal{I}}^r := \mathcal{I} \cap \left\{\frac{k}{r} \mid k \in \mathbb{Z}\right\}.$$

Notice that $\overline{\mathcal{I}}^r = \frac{1}{r}\mathcal{I}^r$. Let $\mathcal{R} = \{r_n\}_{n\ge1}$ be a positive increasing sequence such that $r_n \to \infty$ as $n \to \infty$.

Definition 2.1. (NDDF) A family $\{X^r\}_{r\in\mathcal{R}}$ of continuous-time Markov chains with state spaces $\{\mathcal{I}^r\}_{r\in\mathcal{R}}$, will be called a **nearly density dependent family (NDDF)**

if there exist a function $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ which does not depend on r, and a family of functions $\{\epsilon^r : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow \mathbb{R}\}_{r \in \mathcal{R}}$ such that (i)-(iv) below hold. For convenience, we define

$$\lambda^r(x, \ell) := \beta(x, \ell) + \epsilon^r(x, \ell), \qquad x \in \mathcal{I},\ \ell \in \mathbb{Z} \setminus \{0\},\ r \in \mathcal{R}. \tag{2.1}$$

(i) For each $\ell \in \mathbb{Z} \setminus \{0\}$, $\beta(\cdot, \ell)$ is continuous.

(ii) For each compact subset $\mathcal{K} \subseteq \mathcal{I}$ and $\ell \in \mathbb{Z} \setminus \{0\}$,

$$\sup_{x \in \mathcal{K}} |\epsilon^r(x, \ell)| \leq \frac{M_{\mathcal{K},\ell}}{r} \qquad \text{for every } r \in \mathcal{R}, \tag{2.2}$$

where $M_{\mathcal{K},\ell} > 0$ is a constant not depending on r.

(iii) For each $r \in \mathcal{R}$, $k \in \mathcal{I}^r$ and $\ell \in \mathbb{Z} \setminus \{0\}$, if $k + \ell \notin \mathcal{I}^r$, then $\lambda^r(\frac{k}{r}, \ell) = 0$.

(iv) For each $r \in \mathcal{R}$, X^r has an infinitesimal generator $Q^r = (q^r_{k,j})_{k,j \in \mathcal{I}^r}$ whose off-diagonal entries are given by:

$$q^r_{k,k+\ell} = r\lambda^r\left(\frac{k}{r}, \ell\right), \qquad k \in \mathcal{I}^r,\ \ell \in \mathbb{Z} \setminus \{0\},\ k + \ell \in \mathcal{I}^r. \tag{2.3}$$

As a special case, if the family $\{\epsilon^r : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow \mathbb{R}\}_{r \in \mathcal{R}}$ is such that $\epsilon^r(\cdot, \cdot) = 0$, for every $r \in \mathcal{R}$, then $\{X^r\}_{r \in \mathcal{R}}$ will be called a **density dependent family (DDF)**.

Definition 2.1 is a blend of definitions in Kurtz [14], Ethier & Kurtz ([9], Chapter 11, (1.13)) and Angius et al. ([4], Definition 2). Their definitions are valid in a finite-dimensional space. Here, we focus on one-dimension, but this definition can be generalized to higher dimensions. Ethier & Kurtz [9] introduced the notion of nearly density dependence, although the term was coined afterwards by Angius et al. [4], motivated by chemical reaction networks. Our notion of density dependent family (a special case of Definition 2.1) is consistent with the one introduced by Kurtz [14]. It is implicit in Definition 2.1 that the Markov chains do not explode in finite time. We will assume this throughout the paper.

In Section 4 we provide examples of NDDFs, including CRNs with only one species (Examples 4.2 and 4.3) and CRNs reduced by mass-conservation conditions (Examples 4.1 and 4.5). Some of these examples will be density dependent families, such as the SIS and SI models for epidemics (Examples 4.4 and 4.6 respectively).

Throughout this paper, for any NDDF $\{X^r\}_{r \in \mathcal{R}}$, we make the following additional assumption.

Assumption 2.1. *There exists a finite set $\boldsymbol{L} = \{\ell_1, \ldots, \ell_n\} \subseteq \mathbb{Z} \setminus \{0\}$ such that for each $r \in \mathcal{R}$ and for each $j = 1, \ldots, n$, $\lambda^r(\cdot, \ell_j)$ is not identically zero and*

$$\lambda^r(\cdot, \ell) = 0, \quad \textit{for each } \ell \notin \boldsymbol{L}. \tag{2.4}$$

We regard $\mathbf{L}$ as the set of possible jumps. Assumption 2.1 asserts that the whole family $\{X^r\}_{r \in \mathcal{R}}$ shares a common finite set of possible jumps. Under this condition, because of (2.2), $\beta(\cdot, \ell) = 0$ for a given $\ell \notin \mathbf{L}$ and, for a given compact set $\mathcal{K}$, the

constants $M_{\mathcal{K},\ell}$ in (2.2) can be chosen independent of ℓ. We suppress the ℓ and denote this constant by $M_{\mathcal{K}}$.

For NDDFs coming from chemical reaction networks, Assumption 2.1 is usually satisfied. In this context, each $\ell \in \mathbf{L}$ comes from a reaction, and usually only finitely many reactions are possible.

By the reasoning of Theorem 6.4.1 of Ethier & Kurtz [9], we can construct a realization of a NDDF $\{X^r\}_{r\in\mathcal{R}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with independent unit rate Poisson processes $N_1, \ldots, N_n$ such that the following holds:

$$X^r(t) = X^r(0) + \sum_{j=1}^{n} \ell_j N_j \left(\int_0^t r\lambda^r \left(\frac{X^r(s)}{r}, \ell_j \right) ds \right) \tag{2.5}$$

for every $r \in \mathcal{R}$ and for every $t \geq 0$, where $X^r(0) \in \mathcal{I}^r$. We normalize X^r by considering the concentration process $\overline{X}^r := \frac{1}{r}X^r$, which takes values in $\overline{\mathcal{I}}^r \subseteq \mathcal{I}$.

We next introduce the Constrained Langevin Approximation of $\overline{X}^r$. This is a slight generalization to NDDFs of the CLA introduced by Leite & Williams [17] for chemical reaction networks.

2.2. Constrained Langevin Approximation

Consider a NDDF $\{X^r\}_{r\in\mathcal{R}}$ with its corresponding function β. Consider the continuous coefficients $\mu, \sigma : \mathcal{I} \longrightarrow \mathbb{R}$ defined for $x \in \mathcal{I}$ by

$$\mu(x) := \sum_{j=1}^{n} \ell_j \beta(x, \ell_j), \qquad \sigma(x) := \left(\sum_{j=1}^{n} \ell_j^2 \beta(x, \ell_j) \right)^{1/2}. \tag{2.6}$$

Consider the function $\gamma : \mathcal{I} \longrightarrow \mathbb{R}$ defined by:

$$\gamma = \begin{cases} \mathbb{1}_{\{0\}} - \mathbb{1}_{\{a\}} & \text{if } \mathcal{I} = [0, a] \text{ for } a > 0, \\ \mathbb{1}_{\{0\}} & \text{if } \mathcal{I} = [0, \infty). \end{cases} \tag{2.7}$$

It can be verified, under the conditions of Definition 2.1, that $\mu(x)\gamma(x) \geq 0$ for each $x \in \partial\mathcal{I}$ where $\partial\mathcal{I} = \{0, a\}$ if $\mathcal{I} = [0, a]$ and $\partial\mathcal{I} = \{0\}$ if $\mathcal{I} = [0, \infty)$. Consider a family $\{\nu^r\}_{r\in\mathcal{R}}$ of Borel probability measures on $\mathcal{I}$.

Definition 2.2. (CLA) For each $r \in \mathcal{R}$, a **Constrained Langevin Approximation (CLA)** Z^r with initial distribution ν^r is a process defined together with processes W^r, L^r on $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r)$ such that

(i) $(\Omega^r, \mathcal{F}^r, \mathbb{P}^r)$ is a probability space and $\{\mathcal{F}_t^r\}$ is a filtration on $\mathcal{F}^r$ such that the filtered probability space $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r)$ satisfies the usual conditions,

(ii) $Z^r = \{Z^r(t),\ 0 \leq t < \infty\}$ is a continuous, $\{\mathcal{F}_t^r\}$-adapted process such that $\mathbb{P}^r$-a.s. $Z^r(t) \in \mathcal{I}$ for all $t \geq 0$ and $Z^r(0)$ has distribution ν^r,

(iii) $W^r = \{W^r(t),\ 0 \leq t < \infty\}$ is a one-dimensional standard Brownian motion and an $\{\mathcal{F}_t^r\}$-martingale under $\mathbb{P}^r$,

(iv) $L^r = \{L^r(t),\ 0 \le t < \infty\}$ is a continuous, $\{\mathcal{F}_t^r\}$-adapted, one-dimensional process, that is non-decreasing $\mathbb{P}^r$-a.s., such that $\mathbb{P}^r[L^r(0) = 0] = 1$, and $\mathbb{P}^r$-a.s:

$$L^r(t) = \int_0^t \mathbb{1}_{\{Z^r(s)\in\partial\mathcal{I}\}} dL^r(s), \qquad \text{for every } t \ge 0, \tag{2.8}$$

(v) $\mathbb{P}^r$-a.s. for every $t \ge 0$:

$$Z^r(t) = Z^r(0) + \int_0^t \mu(Z^r(s))ds + \frac{1}{\sqrt{r}}\int_0^t \sigma(Z^r(s))dW^r(s) + \frac{1}{\sqrt{r}}\int_0^t \gamma(Z^r(s))dL^r(s). \tag{2.9}$$

Remark 2.1. Using the language of Section 5, $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r, Z^r, W^r, L^r)$, the tuple of elements from Definition 2.2, is a *(weak) solution* of the Stochastic Differential Equation with Reflection (SDER) (2.9). Conversely, a (weak) solution to equation (2.9) produces a CLA Z^r. Here we allow $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r)$ to depend on r in the definition. In Theorems 3.2 and 3.4 we will prove that this filtered probability space can be chosen to be the same for all r.

The CLA Z^r behaves like the Langevin Approximation to $\overline{X}^r$ in the interior of $\mathcal{I}$. By (2.8), the continuous process L^r can only increase when Z^r is at the boundary of $\mathcal{I}$ and this is used to keep the process Z^r in the interval $\mathcal{I}$.

3. Main results

We present our main results for the two cases, where $\mathcal{I}$ is a bounded interval or the nonnegative halfline. In either case, functions β and $\{\epsilon^r\}_{r\in\mathcal{R}}$ are associated with a NDDF and μ and σ are defined by (2.6).

3.1. Bounded interval case

We make the following assumption in this section.

Assumption 3.1. *The interval $\mathcal{I}$ is of the form $\mathcal{I} = [0, a]$ where $a > 0$. Furthermore, there exists a constant $A > 0$, such that*

$$|\beta(x, \ell_j) - \beta(y, \ell_j)| \le A|x - y| \tag{3.1}$$

for every $x, y \in \mathcal{I}$ and $1 \le j \le n$.

Then for each $r \in \mathcal{R}$, $\mathcal{I}^r$ is a finite set and this implies that the Markov chain X^r automatically does not explode in finite time. Additionally, by (2.2), we can pick a constant $M_{\mathcal{I}} > 0$, not depending on r or ℓ, such that:

$$\sup_{x\in\mathcal{I}} |\epsilon^r(x, \ell_j)| \le \frac{M_{\mathcal{I}}}{r} \tag{3.2}$$

for every $r \in \mathcal{R}$ and $1 \le j \le n$. Under Assumption 3.1, we can prove that the CLA is well defined.

Theorem 3.1. *Suppose Assumption 3.1 holds. Then, for every $r \in \mathcal{R}$ and every Borel probability measure ν^r on $\mathcal{I}$, there exists a CLA Z^r with initial distribution ν^r and associated tuple $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r, Z^r, W^r, L^r)$. Furthermore, if $\tilde{Z}^r$ is another CLA with initial distribution ν^r and tuple $(\tilde{\Omega}^r, \tilde{\mathcal{F}}^r, \{\tilde{\mathcal{F}}_t^r\}, \tilde{\mathbb{P}}^r, \tilde{Z}^r, \tilde{W}^r, \tilde{L}^r)$, then (Z^r, L^r) and $(\tilde{Z}^r, \tilde{L}^r)$ have the same law.*

This result is a consequence of Theorem 5.2 and Proposition 5.1. In Theorem 5.2 and Corollary 5.1 we show that strong existence and pathwise uniqueness in fact hold for equation (2.9). We focused on existence and uniqueness in law here, because those are the notions needed for the next theorem. Note that by (3.1), the coefficient σ is Hölder continuous and not necessarily Lipschitz continuous.

Now, we give our main result.

Theorem 3.2. *Suppose Assumption 3.1 holds. Then, there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ satisfying the usual conditions, such that for each $r \geq 8$ in $\mathcal{R}$ and $x_0^r \in \overline{\mathcal{I}}^r$, there exist processes Z^r, W^r, L^r, X^r and a family of nonnegative random variables $\{\Theta_T^r\}_{T\geq 1}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:*

(i) X^r is a continuous-time Markov chain with infinitesimal generator Q^r such that $\mathbb{P}$-a.s. $\overline{X}^r(0) = x_0^r$,

(ii) Z^r is a CLA with associated tuple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W^r, L^r)$ such that $\mathbb{P}$-a.s. $Z^r(0) = x_0^r$,

and

(iii) for every $T \geq 1$,

$$\sup_{0\leq t\leq T} |\overline{X}^r(t) - Z^r(t)| \leq \Theta_T^r \frac{\log r}{r} \qquad \mathbb{P} - a.s. \tag{3.3}$$

and

$$\mathbb{P}[\Theta_T^r > C_T + x] \leq \frac{K_T}{r^2} \exp\left(-\lambda_T x \log r\right) \qquad \text{for all } x \geq 0, \tag{3.4}$$

where λ_T, C_T and K_T are positive constants depending on $T, \mathcal{I}, \boldsymbol{L}, M_{\mathcal{I}}$ and β.

Remark 3.1. Using the language of Section 5 we note the following. Although strong existence holds for equation (2.9) (under the given assumptions), the solution in Theorem 3.2 is a weak solution. In particular, the filtration $\{\mathcal{F}_t\}$ there is generated by a multidimensional Brownian motion W that occurs in the precursor equation (5.6). See Section 6.1 for further discussion.

Loosely speaking, this result proves that the pathwise error of the Constrained Langevin Approximation on a bounded interval is of order $O(\frac{\log r}{r})$ on compact time intervals. The proof of Theorem 3.2 can be found in Section 6.2. This constructs an appropriate coupling of the processes Z^r and $\overline{X}^r$.

We can derive from Theorem 3.2 the following result concerning the Wasserstein distance $\mathcal{W}_p$ between the laws of $\overline{X}^r$ and Z^r. For $r \in \mathcal{R}$ and $T > 0$, denote by P_T^r the probability measure on $(\mathcal{D}[0,T], \mathcal{M}_T)$ that is the law of $\{\overline{X}^r(t),\ 0 \leq t \leq T\}$ under $\overline{X}^r(0) = x_0^r$, $\mathbb{P}$-a.s. Also, denote by $\tilde{P}_T^r$ the probability measure on $(\mathcal{D}[0,T], \mathcal{M}_T)$ that is the law of $\{Z^r(t),\ 0 \leq t \leq T\}$ under the initial condition $Z^r(0) = x_0^r$, $\mathbb{P}^r$-a.s.

Corollary 3.1. *Suppose Assumption 3.1 holds. Then, for every $T, p \geq 1$, there exists a constant $C = C(T, p, \mathcal{I}, \boldsymbol{L}, M_{\mathcal{I}}, \beta) > 0$ depending only on $T, p, \mathcal{I}, \boldsymbol{L}, M_{\mathcal{I}}$ and β such that*

$$\mathcal{W}_p(P_T^r, \tilde{P}_T^r) \leq C \frac{\log r}{r} \tag{3.5}$$

for every $r \geq 8$ in $\mathcal{R}$.

The proof of this result is given in Section 6.3.

3.2. Halfline case

We make the following assumption in this section.

Assumption 3.2. *The interval $\mathcal{I}$ is of the form $\mathcal{I} = [0, \infty)$. Furthermore, the following hold:*

(i) for every compact set $\mathcal{K} \subseteq \mathcal{I}$, there exists a constant $A_{\mathcal{K}} > 0$, such that

$$|\beta(x, \ell_j) - \beta(y, \ell_j)| \leq A_{\mathcal{K}}|x - y| \tag{3.6}$$

for every $x, y \in \mathcal{K}$ and $1 \leq j \leq n$,

(ii) there exists a constant $C > 0$ such that for each $x \geq 0$,

$$\begin{aligned} x\mu(x) &\leq C(1 + x^2), \\ \sigma^2(x) &\leq C(1 + x^2), \end{aligned} \tag{3.7}$$

(iii) no member of the NDDF associated with β, $\{\epsilon^r\}_{r\in\mathcal{R}}$ explodes in finite time.

Under (3.6), both μ and σ^2 are locally Lipschitz continuous on $[0, \infty)$. Also, (3.7) implies that $\beta(x, \ell_j) \leq C\ell_j^{-2}(1 + x^2)$ for $x \geq 0$ and $1 \leq j \leq n$. In the context of CRNs, this allows for the possibility of some binary reactions (see Example 4.3). Observe that conditions (3.6) and (3.7) alone may not prevent the explosion of the members of the NDDF. Consequently, we include *(iii)* in our assumptions.

As in the previous case, we can prove that the CLA Z^r is properly defined.

Theorem 3.3. *Suppose Assumption 3.2 holds. Then, for every $r \in \mathcal{R}$ and every Borel probability measure ν^r on $\mathcal{I}$, there exists a CLA Z^r with initial distribution ν^r and associated tuple $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r, Z^r, W^r, L^r)$. Furthermore, if $\tilde{Z}^r$ is another CLA with initial distribution ν^r and tuple $(\tilde{\Omega}^r, \tilde{\mathcal{F}}^r, \{\tilde{\mathcal{F}}_t^r\}, \tilde{\mathbb{P}}^r, \tilde{Z}^r, \tilde{W}^r, \tilde{L}^r)$, then (Z^r, L^r) and $(\tilde{Z}^r, \tilde{L}^r)$ have the same law.*

Similar to Theorem 3.1, this result is a consequence of Corollary 5.1, Theorem 5.2 and Proposition 5.1. Strong existence and pathwise uniqueness hold for equation (2.9). The following is our main result for the halfline case.

Theorem 3.4. *Suppose Assumption 3.2 holds. Then, there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ satisfying the usual conditions, such that for each $r \geq 8$ in $\mathcal{R}$ and $x_0^r \in \overline{\mathcal{I}}^r$, there exist processes Z^r, W^r, L^r, X^r defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:*

(i) X^r is a continuous-time Markov chain with infinitesimal generator Q^r such that $\mathbb{P}$-a.s. $\overline{X}^r(0) = x_0^r$,

(ii) Z^r is a CLA with associated tuple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W^r, L^r)$ such that $\mathbb{P}$-a.s. $Z^r(0) = x_0^r$,

and

(iii) for every compact set $\mathcal{K} \subseteq \mathcal{I}$ such that $x_0^r \in \mathcal{K}$, there is a family of nonnegative random variables $\{\Theta_{T,\mathcal{K}}^r\}_{T\geq 1}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that for every $T \geq 1$

$$\sup_{0\leq t\leq T\wedge\tau_{\mathcal{K}}^r} |\overline{X}^r(t) - Z^r(t)| \leq \Theta_{T,\mathcal{K}}^r \frac{\log r}{r} \qquad \mathbb{P} - a.s. \tag{3.8}$$

and

$$\mathbb{P}[\Theta^r_{T,\mathcal{K}} > C_{T,\mathcal{K}} + x] \leq \frac{K_{T,\mathcal{K}}}{r^2} \exp\left(-\lambda_{T,\mathcal{K}} x \log r\right) \qquad \text{for all } x \geq 0, \tag{3.9}$$

where $\tau^r_{\mathcal{K}} = \inf\{t \geq 0 \mid \overline{X}^r(t) \notin \mathcal{K} \text{ or } Z^r(t) \notin \mathcal{K}\}$ *and* $\lambda_{T,\mathcal{K}}, C_{T,\mathcal{K}}$ *and* $K_{T,\mathcal{K}}$ *are positive constants depending on* $T, \mathcal{K}, \boldsymbol{L}, M_{\mathcal{K}}$ *and* β.

The proof of this result can be found in Section 6.4. In this case, the coefficients $\beta(\cdot, \ell_j)$ could be unbounded. This is the main reason for truncation by $\tau^r_{\mathcal{K}}$ in Theorem 3.4. As in the bounded interval case, an analogue of Remark 3.1 holds here.

4. Examples

In this section we give several examples of Density Dependent Families (DDF), Nearly Density Dependent Families (NDDF) and the associated Constrained Langevin Approximations (CLA). These examples mostly come from stochastic chemical reaction network models. For a general description of these models, see Anderson & Kurtz [3].

Example 4.1. This example was considered in Anderson et al. [2]. Consider the following chemical reactions:

$$S_1 \xrightarrow{\alpha_1} S_2, \qquad S_2 \xrightarrow{\alpha_2} S_1$$

where $\alpha_1, \alpha_2 > 0$. For a volume parameter $r > 0$, the process

$$\tilde{X}^r = \{(\tilde{X}^r_1(t), \tilde{X}^r_2(t)),\ 0 \leq t < \infty\}, \tag{4.1}$$

which tracks the amount of S_1 and S_2, respectively, over time, is a continuous-time Markov chain taking values in $\mathbb{Z}^2_+$ with infinitesimal generator given by:

$$\tilde{q}^r_{\tilde{k},\tilde{k}+v} = \begin{cases} \alpha_1 k_1 & \text{if } v = (-1, 1), \\ \alpha_2 k_2 & \text{if } v = (1, -1), \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{k} = (k_1, k_2) \in \mathbb{Z}^2_+$, $v \in \mathbb{Z}^2 \setminus \{0\}$ and $\tilde{k} + v \in \mathbb{Z}^2_+$. This model exhibits the following conservation of mass:

$$\tilde{X}^r_1(t) + \tilde{X}^r_2(t) = \tilde{X}^r_1(0) + \tilde{X}^r_2(0) \quad \text{for every } t \geq 0.$$

Suppose $a^r := \tilde{X}^r_1(0) + \tilde{X}^r_2(0) > 0$ is deterministic and define $X^r := \tilde{X}^r_1$. Then X^r is a continuous-time Markov chain taking values in $\{0, 1, \ldots, a^r\}$ with infinitesimal generator given by:

$$q^r_{k,k+\ell} = \begin{cases} \alpha_1 k & \text{if } \ell = -1, \\ \alpha_2 (a^r - k) & \text{if } \ell = 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k \in \{0, 1, \ldots, a^r\}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \{0, 1, \ldots, a^r\}$. We can re-write these "birth-death" rates as:

$$q^r_{k,k-1} = r\alpha_1 \left(\frac{k}{r}\right), \qquad q^r_{k,k+1} = r\alpha_2 \left(\overline{a}^r - \frac{k}{r}\right), \qquad \text{for suitable } k,$$

where $\overline{a}^r := r^{-1}a^r$.

In order to construct a DDF, consider a constant $a > 0$, and a positive increasing sequence $\mathcal{R} = \{r_n\}_{n\geq 1}$ that tends to infinity, such that $a^r := ar$ is an integer for every $r \in \mathcal{R}$. This can be achieved by choosing $\mathcal{R} = \frac{1}{a}\mathbb{Z}_{>0}$ or $\mathcal{R} = \mathbb{Z}_{>0}$ (the latter provided a is an integer). For $r \in \mathcal{R}$, consider X^r to be the continuous-time Markov chain described previously with total mass $a^r = ar$, so that $\bar{a}^r = a$ for all $r \in \mathcal{R}$. Under these conditions, $\{X^r\}_{r\in\mathcal{R}}$ is a DDF for $\mathcal{I} = [0, a]$ with function $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by

$$\beta(x, \ell) = \begin{cases} \alpha_1 x & \text{if } \ell = -1, \\ \alpha_2 (a - x) & \text{if } \ell = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Additionally, the coefficients $\{\lambda^r\}_{r\in\mathcal{R}}$ defined in (2.1), are given by $\lambda^r(x, \ell) = \beta(x, \ell)$ for $r \in \mathcal{R}$. Consequently, Assumptions 2.1 and 3.1 hold with $\mathbf{L} = \{-1, 1\}$. The coefficients for the Constrained Langevin Approximation (CLA) Z^r are given by:

$$\mu(x) = \alpha_2 a - (\alpha_1 + \alpha_2)x, \qquad \sigma(x) = \sqrt{\alpha_2 a + (\alpha_1 - \alpha_2)x} \qquad \text{for } x \in \mathcal{I}.$$

By Theorem 3.1, for a prescribed initial distribution, existence and uniqueness in law hold for Z^r. Furthermore, Theorem 5.2 shows that strong existence and uniqueness hold for equation (2.9). Finally, by Theorem 3.2, given an initial condition $x_0^r \in \overline{\mathcal{I}}^r$, we can realize a version of Z^r and $\overline{X}^r$ on the same probability space satisfying the error estimates (3.3) and (3.4). From Remark 3.1, this version of Z^r is a (weak) solution of (2.9).

In Figure 1 we show a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r. We simulated the Markov chain X^r using the Doob-Gillespie algorithm. For the CLA Z^r we used an Euler-Maruyama scheme where the discretized process is projected back to the interval in the event of escaping it. More precisely, let $0 = t_0 < t_1 < \ldots < t_N = T$ be a uniform partition of $[0, T]$ with step size $h := t_{n+1} - t_n$. For $a > 0$, $\mathcal{I} = [0, a]$ and an initial condition $x_0^r \in \overline{\mathcal{I}}^r$, define $Z_h^r(0) := x_0^r$ and for $0 \leq n \leq N - 1$:

$$Z_h^r(t_{n+1}) := \pi_{\mathcal{I}} \left[Z_h^r(t_n) + \mu(Z_h^r(t_n))h + \sigma(Z_h^r(t_n))(W(t_{n+1}) - W(t_n)) \right] \tag{4.2}$$

where W is a standard Brownian motion and $\pi_{\mathcal{I}} : \mathbb{R} \longrightarrow \mathcal{I}$ is given by, $\pi_{\mathcal{I}}(x) = 0$ if $x < 0$, x if $x \in \mathcal{I}$ and a if $x > a$. For Figure 1 we used an Euler step size of $h = 0.04$, which is comparable to the mean time until the next reaction when the Markov chain $\overline{X}^r$ is at the initial level $x_0^r = 0.1$. We chose this step size and the time length $T = 10$ to illustrate transient behavior.

Since μ and σ are Lipschitz continuous and bounded, it can be shown that the approximation Z_h^r to Z^r is of *strong order* $h^{1/2-\varepsilon}$ for every $\varepsilon > 0$, i.e., for every $\varepsilon, T > 0$ there exists a constant $C_{\varepsilon,T} > 0$ such that:

$$\left(\mathbb{E} \left[\sup_{0\leq t\leq T} |Z^r(t) - Z_h^r(t)|^2 \right] \right)^{1/2} \leq C_{\varepsilon,T} h^{1/2-\varepsilon}, \tag{4.3}$$

for sufficiently small $h > 0$. See Slominski [20] for a proof.

Although our justification of the CLA approximation Z^r for the continuous-time Markov chain $\overline{X}^r$ is over a finite time horizon, it is interesting to compare long run

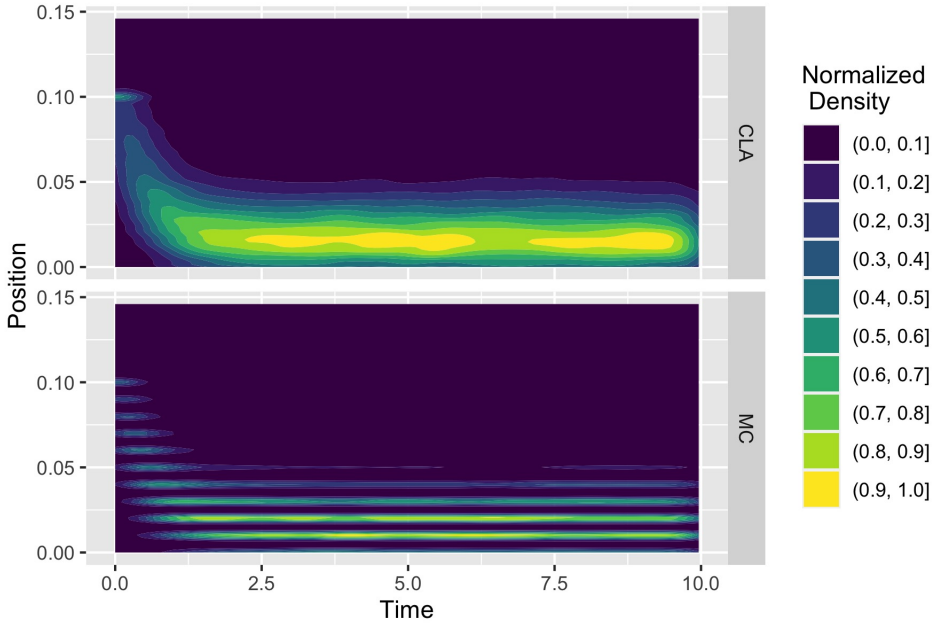

FIGURE 1: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 10$, for Example 4.1. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 2$, $\alpha_2 = 0.01$, $r = 100$, $\bar{a}^r = a = 4$ and $x_0^r = 0.1$. The figures were generated with the ggplot2 R package.

behavior, in particular, stationary distributions. For this simple model, a comparison of exact formulas for the stationary distributions of the continuous-time Markov chain $\overline{X}^r$, the CLA Z^r and the linear noise approximation (LNA) was given in Example 2 of Anderson et al. [2]. We refer the interested reader to that work, especially Figure 2 there, where it is seen that the continuous-time Markov chain and CLA results have good agreement, and that the CLA is considerably more accurate than the LNA. Beyond such specific examples, obtaining general error estimates between stationary distributions for the continuous-time Markov chain and the CLA is an interesting topic for future investigation.

Example 4.2. Consider the following production-degradation chemical reactions:

$$\phi \xrightarrow{\alpha_1} S_1, \qquad S_1 \xrightarrow{\alpha_2} \phi$$

where $\alpha_1, \alpha_2 > 0$. This example, also considered in Anderson et al. [2], corresponds to an $M/M/\infty$ queue. For a volume parameter $r > 0$, the process $X^r = \{X^r(t),\ 0 \leq t < \infty\}$, which tracks the amount of S_1 over time, is a continuous-time Markov chain taking values in $\mathbb{Z}_+$ with infinitesimal generator given by:

$$q^r_{k,k+\ell} = \begin{cases} \alpha_1 r & \text{if } \ell = 1, \\ \alpha_2 k & \text{if } \ell = -1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k \in \mathbb{Z}_+$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \mathbb{Z}_+$. With $\mathcal{R} = \{r_n\}_{n\geq 1}$ being any positive increasing sequence converging to ∞, the family $\{X^r\}_{r\in\mathcal{R}}$ is a DDF with $\mathcal{I} = [0, \infty)$ and $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by:

$$\beta(x, \ell) = \begin{cases} \alpha_1 & \text{if } \ell = 1, \\ \alpha_2 x & \text{if } \ell = -1, \\ 0 & \text{otherwise.} \end{cases}$$

The coefficients $\{\lambda^r\}_{r\in\mathcal{R}}$ defined in (2.1), are given by $\lambda^r(x, \ell) = \beta(x, \ell)$. Assumption 2.1 holds with $\mathbf{L} = \{-1, 1\}$. Regarding Assumption 3.2, condition (3.6) is satisfied, and the coefficients

$$\mu(x) = \alpha_1 - \alpha_2 x, \qquad \sigma(x) = \sqrt{\alpha_1 + \alpha_2 x}, \qquad x \in \mathcal{I}, \tag{4.4}$$

satisfy condition (3.7). Finally, it is well known that X^r does not explode in finite time for each $r > 0$. Consequently, Assumption 3.2 is satisfied. By Theorems 3.3 and 5.2 existence and uniqueness in law, as well as strong existence and uniqueness hold for equation (2.9). The process Z^r is a nonnegative diffusion that is instantaneously reflected upon touching 0. Also, Theorem 3.4 shows that given an initial condition $x_0^r \in \overline{\mathcal{I}}^r$, we can realize Z^r and $\overline{X}^r$ on the same probability space so that (3.8) and (3.9) hold.

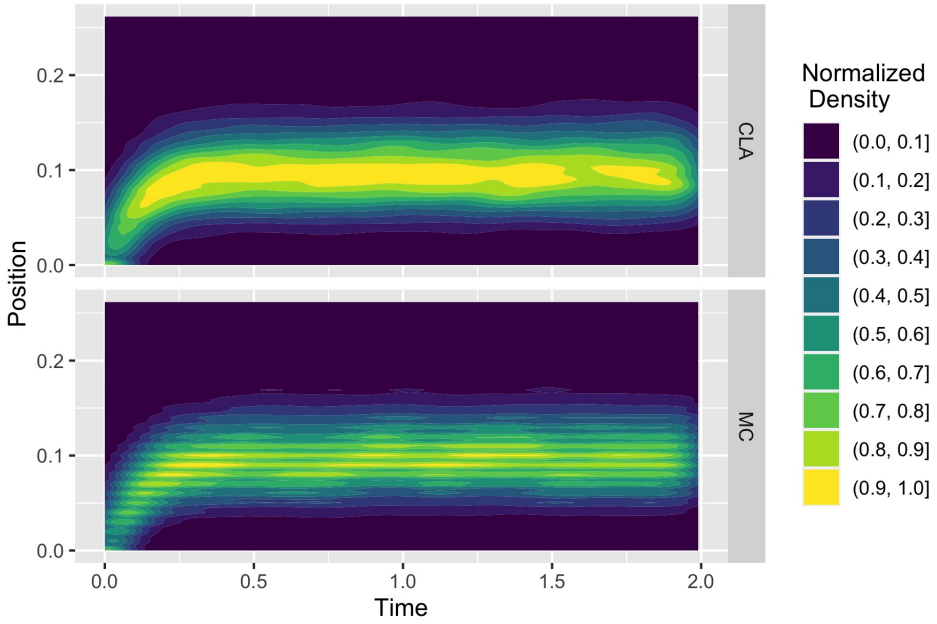

FIGURE 2: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 2$, for Example 4.2. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 1$, $\alpha_2 = 10$, $r = 100$, and $x_0^r = 0$. The figures were generated with the ggplot2 R package.

In Figure 2 we show a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r for Example 4.2. As for Example 4.1, we simulated the Markov chain X^r using the Doob-Gillespie algorithm and the CLA Z^r using an Euler-Maruyama scheme where the discretized process is projected back to the interval in the event of escaping it. For Figure 2 we used an Euler step size of $h = 0.01$, which is comparable to the mean time until the next reaction when the Markov chain $\overline{X}^r$ is at the initial level $x_0^r = 0$ and is twice the mean time until the next reaction when the Markov chain $\overline{X}^r$ is at the level 0.1, where the production and degradation rates are balanced.

Also, for this example, a comparison of exact formulas for the stationary distributions of the continuous-time Markov chain $\overline{X}^r$, the CLA Z^r and the LNA was given in Example 1 of Anderson et al. [2]. We refer the interested reader to that work, especially Figure 1 there, where it is seen that the Markov chain and CLA results have good agreement, and that the CLA is considerably more accurate than the LNA.

Example 4.3. The following example shows how, in the halfline case, our assumptions allow for reactions with more than one unit of a given species in the input. Consider the chemical reactions:

$$\phi \xrightarrow{\alpha_1} S_1, \qquad 2S_1 \xrightarrow{\alpha_2} S_1$$

where $\alpha_1, \alpha_2 > 0$. For a volume parameter $r > 0$, with standard mass action kinetics, the process $X^r = \{X^r(t),\ 0 \le t < \infty\}$, which tracks the amount of S_1 over time, is a continuous-time Markov chain taking values in $\mathbb{Z}_+$ with infinitesimal generator given by:

$$q^r_{k,k+\ell} = \begin{cases} \alpha_1 r & \text{if } \ell = 1, \\ \frac{\alpha_2}{r} k(k-1) & \text{if } \ell = -1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k \in \mathbb{Z}_+$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \mathbb{Z}_+$. With $\mathcal{R} = \{r_n\}_{n\ge1}$ being any positive increasing sequence converging to ∞, the family $\{X^r\}_{r\in\mathcal{R}}$ is a NDDF in the case $\mathcal{I} = [0, \infty)$ with $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by:

$$\beta(x, \ell) = \begin{cases} \alpha_1 & \text{if } \ell = 1, \\ \alpha_2 x^2 & \text{if } \ell = -1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\epsilon^r(x, \ell) = \begin{cases} 0 & \text{if } \ell = 1, \\ -\frac{\alpha_2 x}{r} & \text{if } \ell = -1, \\ 0 & \text{otherwise.} \end{cases}$$

The function β satisfies the local Lipschitz condition (3.6). The collection $\{\lambda^r\}_{r\in\mathcal{R}}$ is given by $\lambda^r(x, \ell) = \beta(x, \ell) + \epsilon^r(x, \ell)$. Consequently, Assumption 2.1 holds with $\mathbf{L} = \{-1, 1\}$. The coefficients in (2.6) are given by:

$$\mu(x) = \alpha_1 - \alpha_2 x^2, \qquad \sigma(x) = \sqrt{\alpha_1 + \alpha_2 x^2}, \qquad x \in \mathcal{I}, \tag{4.5}$$

and they satisfy condition (3.7). Since the Markov chain is dominated by a Poisson process, it does not explode in finite time. Consequently, Assumption 3.2 is satisfied.

For $r \in \mathcal{R}$, the Constrained Langevin Approximation (CLA) Z^r satisfies existence and uniqueness in law as well as strong existence and uniqueness. A coupling construction of $\overline{X}^r$ and Z^r is provided by Theorem 3.4.

In Figure 3 we show a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r for Example 4.3. As for prior examples, we simulated the Markov chain X^r using the Doob-Gillespie algorithm and the CLA Z^r using an Euler-Maruyama scheme where the discretized process is projected back to the interval in the event of escaping it. For Figure 3 we used an Euler step size of $h = 0.01$, which is comparable to the mean time until the next reaction for the Markov chain $\overline{X}^r$ at the initial level $x_0^r = 0$ and is twice the mean time until the next reaction when the Markov chain $\overline{X}^r$ is at the level 0.125, where the production and degradation rates are balanced.

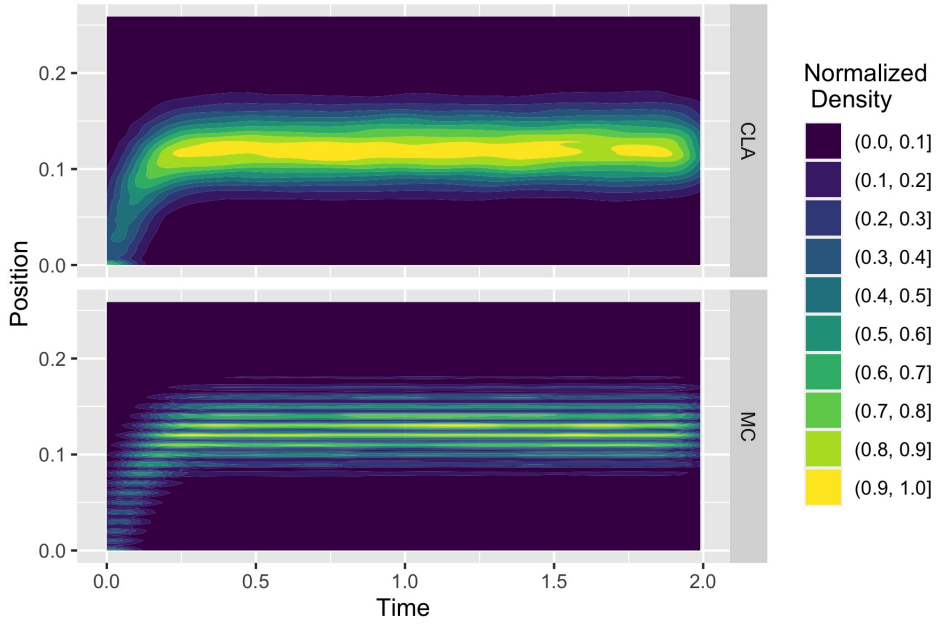

FIGURE 3: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 2$, for Example 4.3. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 1$, $\alpha_2 = 64$, $r = 100$, and $x_0^r = 0$. The figures were generated with the ggplot2 R package.

The above examples share the property of having $\sigma^2(x) > 0$ for $x \in \partial\mathcal{I}$. It is not hard to see that this implies

$$\int_0^\infty \mathbb{1}_{\{Z^r(s)\in\partial\mathcal{I}\}} ds = 0, \qquad \mathbb{P}^r - \text{a.s.}$$

for the CLA Z^r. When σ^2 vanishes at the boundary, Z^r can be absorbed at the boundary, as described in the next example.

Example 4.4. The SIS model, where SIS stands for Susceptible-Infected-Susceptible, is a stochastic epidemic model where an infection is transmitted and cured in a population of *susceptible* and *infected* individuals. Each susceptible individual acquires the infection at a rate proportional to the fraction of the total population that is infected. On the other hand, each infected individual has a chance of becoming cured at a constant rate. Once cured, the individual becomes susceptible again. This model can be described by the following chemical reactions:

$$S + I \xrightarrow{\alpha_1} 2I, \qquad I \xrightarrow{\alpha_2} S$$

where $\alpha_1, \alpha_2 > 0$ and I, S stand for infected and susceptible respectively. For a population size parameter $r > 0$, the process $\tilde{X}^r = \{(\tilde{X}_1^r(t), \tilde{X}_2^r(t)),\ 0 \le t < \infty\}$, which tracks the number of infected and susceptible individuals over time, is a continuous-time Markov chain taking values in $\mathbb{Z}_+^2$ with infinitesimal generator given by:

$$\tilde{q}^r_{\tilde{k},\tilde{k}+v} = \begin{cases} \alpha_1 \frac{k_1 k_2}{r} & \text{if } v = (1,-1), \\ \alpha_2 k_1 & \text{if } v = (-1,1), \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{k} = (k_1, k_2) \in \mathbb{Z}_+^2$, $v \in \mathbb{Z}^2 \setminus \{0\}$ and $\tilde{k} + v \in \mathbb{Z}_+^2$. Similar to Example 4.1, this model exhibits the following conservation of mass:

$$\tilde{X}_1^r(t) + \tilde{X}_2^r(t) = \tilde{X}_1^r(0) + \tilde{X}_2^r(0) \quad \text{for every } t \ge 0. \tag{4.6}$$

For $r \in \mathbb{Z}_{>0}$, suppose $\tilde{X}_1^r(0) + \tilde{X}_2^r(0) = r$ and define $X^r := \tilde{X}_1^r$. Then, X^r is a continuous-time Markov chain taking values in $\{0, 1, \dots, r\}$ which tracks the number of infected individuals. It has infinitesimal generator given by:

$$q^r_{k,k+\ell} = \begin{cases} \alpha_1 k \left(1 - \frac{k}{r}\right) & \text{if } \ell = 1, \\ \alpha_2 k & \text{if } \ell = -1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k \in \{0, 1, \dots, r\}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \{0, 1, \dots, r\}$. We choose $\mathcal{R} = \{r_n\}_{n \ge 1} = \mathbb{Z}_{>0}$. Under these conditions, $\{X^r\}_{r \in \mathcal{R}}$ is a DDF with $\mathcal{I} = [0, 1]$ and $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by:

$$\beta(x, \ell) = \begin{cases} \alpha_1 x(1-x) & \text{if } \ell = 1, \\ \alpha_2 x & \text{if } \ell = -1, \\ 0 & \text{otherwise.} \end{cases}$$

The coefficients $\{\lambda^r\}_{r \in \mathcal{R}}$ are given by $\lambda^r(x, \ell) = \beta(x, \ell)$. Consequently, $\{X^r\}_{r \in \mathcal{R}}$ satisfies Assumptions 2.1 and 3.1, with $\mathbf{L} = \{-1, 1\}$. The coefficients for the CLA are given by

$$\mu(x) = \alpha_1 x(1-x) - \alpha_2 x, \qquad \sigma(x) = \sqrt{\alpha_1 x(1-x) + \alpha_2 x}, \qquad x \in \mathcal{I}. \tag{4.7}$$

Again, existence and uniqueness in law holds for (2.9) as well as strong existence and uniqueness. Theorem 3.2 shows that we can realize versions of Z^r and $\overline{X}^r$ on the same

probability space satisfying (3.3) and (3.4). Our error estimates show Z^r and $\overline{X}^r$ are close over compact time intervals. However, hitting times can also be close as we show below.

In this model, σ vanishes at 0. Since μ also vanishes at 0, pathwise uniqueness for equation (2.9) (see Corollary 5.1) implies that 0 is an absorbing state for Z^r. The process $\overline{X}^r$ also has 0 as an absorbing state, which is the state of no infected individuals. The other boundary point of $\mathcal{I}$ is at 1, and this state is not absorbing for $\overline{X}^r$ or Z^r. Indeed, Z^r is reflected back into $\mathcal{I}$ at 1, whereas without reflection, the diffusion with coefficients given by (4.7) would exit $\mathcal{I}$ by going above 1.

In Figure 4 we show a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r for Example 4.4. As for prior examples, we simulated the Markov chain X^r using the Doob-Gillespie algorithm and the CLA Z^r using an Euler-Maruyama scheme where the discretized process is projected back to the interval in the event of escaping it. For Figure 4 we used an Euler step size of $h = 0.01$, which is comparable to the mean time until the next reaction for the Markov chain $\overline{X}^r$ at the initial level $x_0^r = 0.99$.

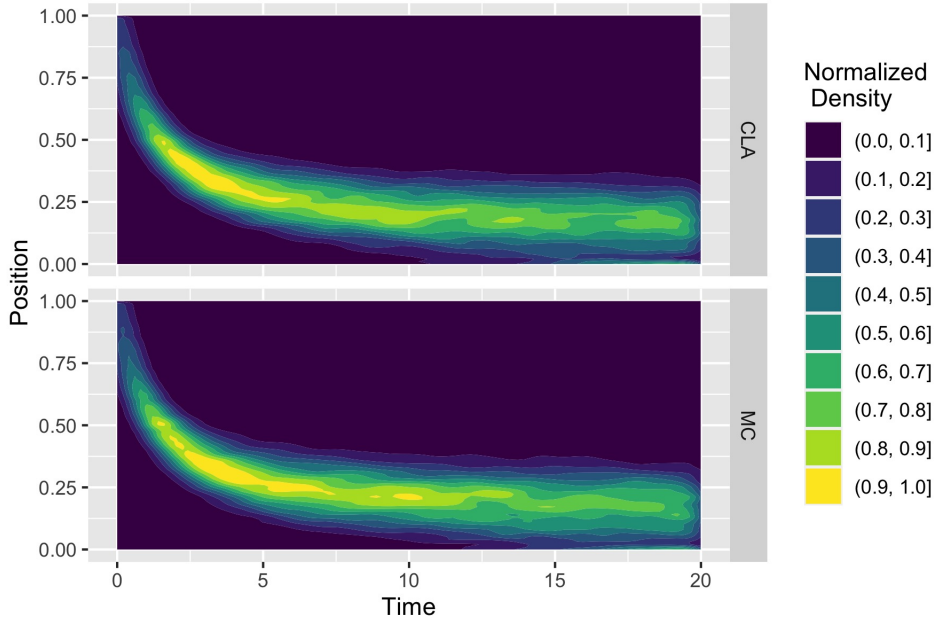

FIGURE 4: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 20$, for Example 4.4. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 1$, $\alpha_2 = 0.8$, $r = 100$, and $x_0^r = 0.99$. The figures were generated with the ggplot2 R package.

The Markov chain $\overline{X}^r$ has one transient class and one absorbing state, 0, which can be reached from the transient class. Therefore, $\overline{X}^r$ will hit 0 with probability 1, and in fact, one can show that the mean time for $\overline{X}^r$ to hit 0 will be finite when started from any of the states $\frac{1}{r}, \frac{2}{r}, \ldots, 1$. The time for Z^r to hit 0 is also finite, although this is not obvious a priori. In fact, as we show next, $T_0^r := \inf\{t \geq 0 \ : \ Z^r(t) = 0\}$ has

finite mean when $\mathbb{P}^r$-a.s. $Z^r(0) = x$, for $x \in \mathcal{I}$.

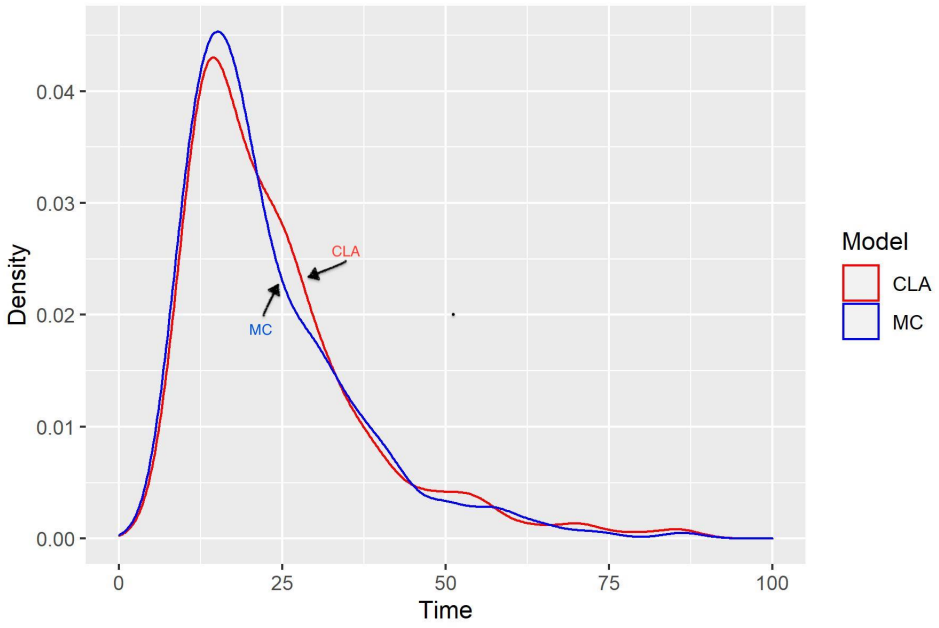

FIGURE 5: Empirical density for the hitting time to 0 (absorbing time) for the Markov chain $\overline{X}^r$ and the CLA Z^r. The plot was generated by 500 simulations of each process with parameters $\alpha_1 = 1$, $\alpha_2 = 0.95$, $r = 100$ and $x_0^r = 0.99$. The curves are kernel density estimates, generated with the ggplot2 R package.

Define the function $w(u) := 2r \int_{(0,u]} \frac{\mu(s)}{\sigma^2(s)} ds$ for $u \in (0,1]$ and $w(0) := 0$. Although $\sigma^2(0) = 0$, the function $s \in (0,1] \to \frac{\mu(s)}{\sigma^2(s)}$ can be continuously extended to $[0,1]$, making w a function of class $\mathcal{C}^1[0,1]$. Now, define for $u \in (0,1]$, $x \in [0,1]$,

$$g(u) := 2re^{-w(u)} \int_{(u,1]} \frac{e^{w(s)}}{\sigma^2(s)} ds, \quad f(x) := \int_0^x g(u) du. \tag{4.8}$$

The function g is nonnegative and, since $\sigma^2(s) = sh(s)$ for every $s \in [0,1]$, where $h : [0,1] \longrightarrow \mathbb{R}$ is continuous and doesn't vanish, g is integrable over $(0,1]$. With these facts, the reader may verify that f is of class $\mathcal{C}^2(0,1] \cap \mathcal{C}[0,1]$, with boundary values $f(0) = 0$ and $f'(1) = 0$, and such that $\mathcal{L}f = -1$ in $(0,1]$, where $\mathcal{L}f = \frac{\sigma^2}{2r} f'' + \mu f'$. We can use these properties, together with Itô's formula applied to a family of stopped processes, to conclude that for any $x \in \mathcal{I} = [0,1]$ and any CLA Z^r with associated tuple $(\Omega^r, \mathcal{F}^r, \{\mathcal{F}_t^r\}, \mathbb{P}^r, Z^r, W^r, L^r)$, where $Z^r(0) = x$, $\mathbb{P}^r$-a.s., we have $\mathbb{E}_x^r[T_0^r] = f(x) < \infty$.

In Figure 5 we show a comparison between the empirical distribution densities of the hitting time to 0 (absorbing time) for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r. As in Example 4.1, we simulate the Markov chain X^r using the Doob-Gillespie algorithm and the CLA Z^r using an Euler-Maruyama scheme where the discretized process Z_h^r is projected back to the interval in the event of escaping from it.

In this example though, the coefficient σ fails to be Lipschitz and therefore, the strong order of $h^{1/2-\varepsilon}$, for every $\varepsilon > 0$, is not guaranteed for the approximation Z_h^r to Z^r.

Example 4.5. We next consider the *crazy clock reaction*, discussed in Angius et al.[4]. This model can be described by the following chemical reactions:

$$A + B \xrightarrow{\alpha_1} 2B, \qquad A \xrightarrow{\alpha_2} B$$

where $\alpha_1, \alpha_2 > 0$. This example differs from Example 4.4 in the dynamics of the species A (species S in Example 4.4), which in this case monotonically decreases. For $r \in \mathbb{Z}_{>0}$, we let r be the total mass and, following [4], we define the process $X^r(t)$ to be the amount of species A at time t. The process X^r is a continuous-time Markov chain taking values in $\{0, 1, \dots, r\}$ with infinitesimal generator given by:

$$q^r_{k,k+\ell} = \begin{cases} \alpha_1 k \left(1 - \frac{k}{r}\right) + \alpha_2 k & \text{if } \ell = -1, \\ 0 & \text{otherwise,} \end{cases}$$

where $k \in \{0, 1, \dots, r\}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \{0, 1, \dots, r\}$. For $\mathcal{R} = \{r_n\}_{n\geq 1} = \mathbb{Z}_{>0}$, $\{X^r\}_{r\in\mathcal{R}}$ is a DDF with $\mathcal{I} = [0, 1]$, and $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by:

$$\beta(x, \ell) = \begin{cases} \alpha_1 x(1 - x) + \alpha_2 x & \text{if } \ell = -1, \\ 0 & \text{otherwise.} \end{cases}$$

Then Assumptions 2.1 and 3.1 hold with $\mathbf{L} = \{-1\}$ and so by Theorem 3.2, given an initial condition $x_0^r \in \overline{\mathcal{I}}^r$, Z^r and $\overline{X}^r$ can be realized on the same probability space so that (3.3) and (3.4) hold. The coefficients for the CLA are given by

$$\mu(x) = -\alpha_1 x(1 - x) - \alpha_2 x, \qquad \sigma(x) = \sqrt{\alpha_1 x(1 - x) + \alpha_2 x}, \qquad x \in \mathcal{I}.$$

The coefficients satisfy $\sigma(0) = \mu(0) = 0$ which, by pathwise uniqueness (see Corollary 5.1), implies that 0 is an absorbing state for Z^r. Similarly, the process $\overline{X}^r$ has 0 as an absorbing state, which will be attained with probability 1, and the mean time for $\overline{X}^r$ to reach 0 is finite starting from any of the states $\frac{1}{r}, \frac{2}{r}, \dots, 1$. The process $\overline{X}^r$ is decreasing, whereas Z^r is not monotone and is reflected at the upper boundary point 1 of $\mathcal{I}$. Using the same approach as in Example 4.4, for $T_0^r := \inf\{t \geq 0 \mid Z^r(t) = 0\}$, we can prove that T_0^r has finite mean when $\mathbb{P}^r$-a.s. $Z^r(0) = x$, for $x \in \mathcal{I}$. In fact, since $\frac{\mu}{\sigma^2}$ can be defined continuously up to 0 and $x \mapsto \int_x^1 \frac{du}{\sigma^2(u)} \leq -C \log x$ is integrable on $(0, 1]$, we can define f and g as in (4.8) and draw this conclusion.

In Figure 6 we show a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r for Example 4.5. As for prior examples, we simulated the Markov chain X^r using the Doob-Gillespie algorithm and the CLA Z^r using an Euler-Maruyama scheme where the discretized process is projected back to the interval in the event of escaping it. For Figure 6 we used an Euler step size of $h = 0.02$, which is comparable to the mean time until the next reaction for the Markov chain $\overline{X}^r$ at the initial level $x_0^r = 1$.

Example 4.6. Our final example shows how the long-run behaviour of the CLA Z^r can differ from that of the Markov chain $\overline{X}^r$. Consider an epidemic model as in

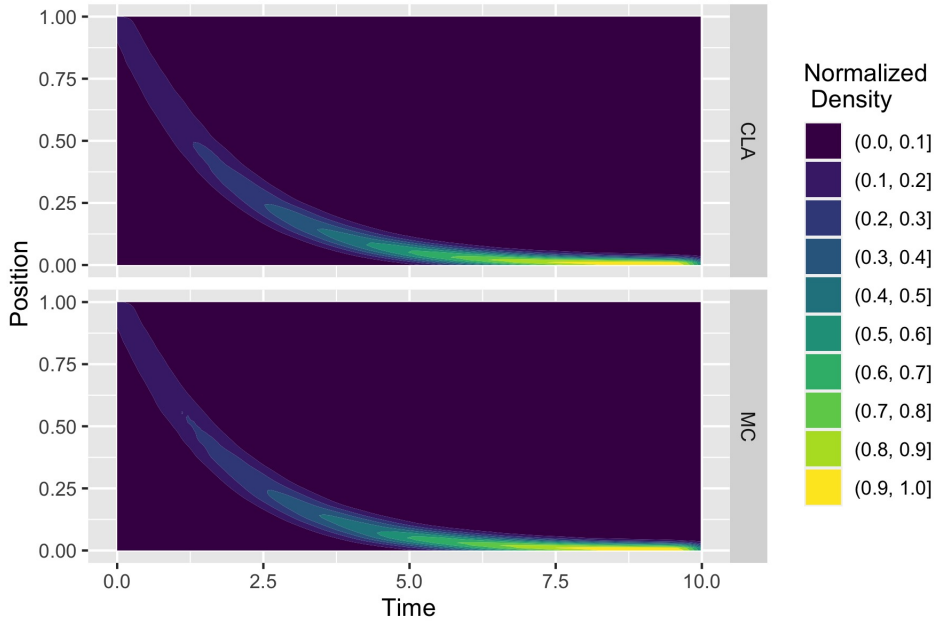

FIGURE 6: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 10$, for Example 4.5. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, $r = 100$, and $x_0^r = 1$. The figures were generated with the ggplot2 R package.

Example 4.4, with the difference that no recovery is allowed for infected individuals. This is called the SI model and can be described by the chemical reaction:

$$S + I \xrightarrow{\alpha_1} 2I$$

where $\alpha_1 > 0$ and I, S stand for infected and susceptible individuals, respectively. As in Example 4.4, we define the continuous-time Markov chain $\tilde{X}^r = (\tilde{X}_1^r, \tilde{X}_2^r)$. In this case, there are no transitions in the direction $v = (-1, 1)$. However, conservation of mass (4.6) holds in this case as well. For an integer-valued population size parameter $r \geq 1$, we consider the process $X^r := \tilde{X}_1^r$ which tracks the number of infected individuals for a total population of r individuals. This process is a continuous-time Markov chain taking values in $\{0, 1, \ldots, r\}$, with infinitesimal generator given by:

$$q_{k,k+\ell}^r = \begin{cases} \alpha_1 k \left(1 - \frac{k}{r}\right) & \text{if } \ell = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{4.9}$$

where $k \in \{0, 1, \ldots, r\}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $k + \ell \in \{0, 1, \ldots, r\}$. Equation (4.9) shows that, if at least one infected individual is present, the absence of recovery leads to a monotone increase of the infected population until everyone is infected. For $\mathcal{R} = \{r_n\}_{n \geq 1} = \mathbb{Z}_{>0}$, $\{X^r\}_{r \in \mathcal{R}}$ is a DDF with $\mathcal{I} = [0, 1]$ and function $\beta : \mathcal{I} \times (\mathbb{Z} \setminus \{0\}) \longrightarrow [0, \infty)$ given by:

$$\beta(x, \ell) = \begin{cases} \alpha_1 x(1 - x) & \text{if } \ell = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then Assumptions 2.1 and 3.1 hold with $\mathbf{L} = \{1\}$. For $r \in \mathcal{R}$, the coefficients for the CLA are given by

$$\mu(x) = \alpha_1 x(1-x), \qquad \sigma(x) = \sqrt{\alpha_1 x(1-x)}, \qquad x \in \mathcal{I}.$$

In this model, both $\overline{X}^r$ and Z^r have 0 and 1 as absorbing states. If $\overline{X}^r(0) \in (0,1)$, then $\overline{X}^r$ will monotonically increase up to 1. On the other hand, when $Z^r(0) \in (0,1)$, Z^r will not be monotonic. Moreover, as we show next, under the condition $Z^r(0) = x \in (0,1)$, the process Z^r will escape the interval $(0,1)$ $\mathbb{P}^r$-a.s. and it may do it through 0 or 1 with positive probability for each case. Although escaping through 0 is an undesirable property of Z^r in this example, in (4.11) below, we show that the actual probability of escaping through 0 decays exponentially as $r \to \infty$. Figure 7, has a (kernel) density estimate plot of sample path simulations for the continuous-time Markov chain $\overline{X}^r$ and the CLA Z^r for Example 4.6. Here, the same simulation methods as for the other examples were used, and we used an Euler step size of $h = 0.1$, which is comparable to the mean time until the next reaction for the Markov chain $\overline{X}^r$ at the initial level $x_0^r = 0.1$.

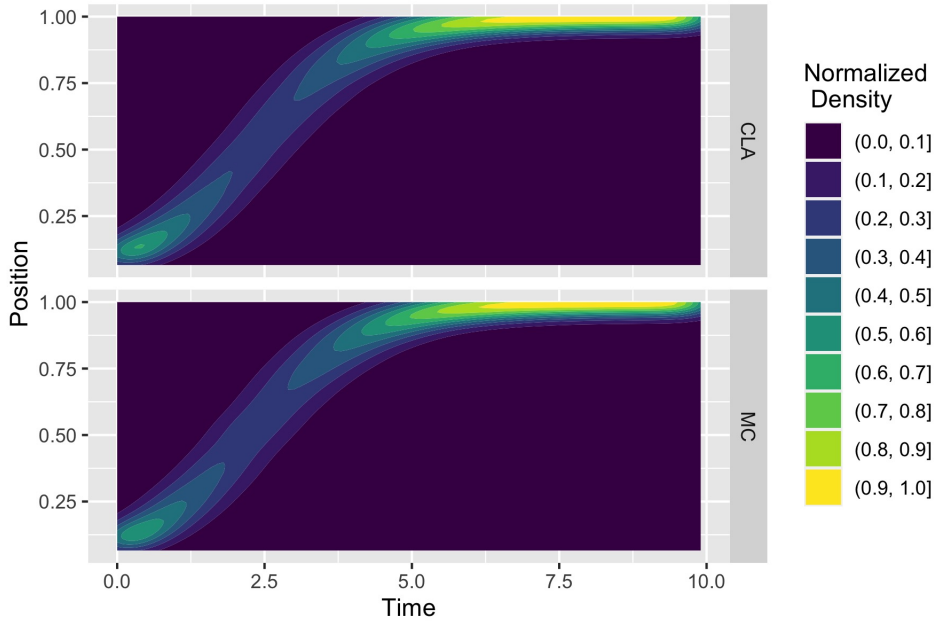

FIGURE 7: Kernel density estimate plot for the paths of the CLA Z^r (top panel) and the Markov chain $\overline{X}^r$ (bottom panel) from time 0 to $T = 10$, for Example 4.6. The plot was generated from 500 simulation runs of each process with parameters $\alpha_1 = 1$, $r = 100$, and $x_0^r = 0.1$. The figures were generated with the ggplot2 R package.

To justify (4.11), consider $x \in (0,1)$, $r \in \mathcal{R}$ and a CLA Z^r with $\mathbb{P}^r[Z^r(0) = x] = 1$. Let $T_{0,1}^r := \inf\{t \geq 0 \mid Z^r(t) \notin (0,1)\}$. Then, since μ and σ vanish on the set $\{0,1\}$,

$Z^r_{0,1} := \{Z^r(t \wedge T^r_{0,1}), 0 \le t < \infty\}$ is a solution of equation:

$$Z^r_{0,1}(t) = x + \int_0^t \mu(Z^r_{0,1}(s))ds + \frac{1}{\sqrt{r}} \int_0^t \sigma(Z^r_{0,1}(s))dW^r(s), \qquad t \ge 0. \tag{4.10}$$

The scale function p and speed measure $m(dy)$ of (4.10) are given by:

$$p(y) = \frac{e^r}{2r}\left(e^{-r} - e^{-2ry}\right), \qquad m(dy) = \left(\frac{2re^{-r}}{\alpha_1}\right)\frac{e^{2ry}}{y(1-y)}dy,$$

where $y \in (0,1)$ and where we choose $c = \frac{1}{2}$ as the value for which $p(c) = 0$. Note that p easily extends continuously to $[0,1]$. Let

$$v(y) := \int_{\frac{1}{2}}^y p'(z)\left(\int_{\frac{1}{2}}^z m(du)\right)dz = \frac{2r}{\alpha_1}\int_{\frac{1}{2}}^y e^{-2rz}\left(\int_{\frac{1}{2}}^z \frac{e^{2ru}}{u(1-u)}du\right)dz, \quad y \in (0,1).$$

The reader may verify that $v(0+), v(1-) < \infty$. By Proposition 5.5.32 in Karatzas & Shreve [10], we conclude that $\mathbb{E}^r_x[T^r_{0,1}] < \infty$. Finally, Proposition 5.5.22 in [10] yields the formula:

$$\mathbb{E}^r_x\left[\mathbb{1}_{\{Z^r(T^r_{0,1})=0\}}\right] = \frac{p(1)-p(x)}{p(1)-p(0)} = e^{-2rx}\left(\frac{1-e^{-2r(1-x)}}{1-e^{-2r}}\right) \le e^{-2rx}. \tag{4.11}$$

In the above examples, simulation of the Markov chain using the Doob-Gillespie algorithm and of the CLA using an Euler-Maruyama scheme generally produces similar results when the mean time until the next reaction for the Markov chain is comparable to the Euler step size in the CLA simulation (and thereby the computation times are of similar order). In examples with more reactions versus the number of species, the mean time until the next reaction for the Markov chain is typically reduced, requiring more simulation time, while the Euler step size for the CLA can be held fixed. The simulation advantage of the CLA then becomes more apparent as parameters for the CLA come from combining reactions, whereas simulation of the Markov chain involves simulating each reaction individually. Examples of this in one-dimension may be limited due to the restriction to one effective species, however, examples in higher dimensions (for which there are not error bounds yet), such as those in [2, 17], indicate that accurate results can be achieved with the CLA with a considerable reduction in computation time compared with that for the Markov chain. A further advantage of the CLA versus the Markov chain, especially in the one-dimensional case, is that to analytically compute closed form expressions for quantities of interest, for the CLA one uses differential equations, whereas for the Markov chain one uses difference equations, which tend to be more combinatorially complex. See [2] for some one-dimensional examples of such computations.

5. Existence and uniqueness for the CLA

In this section we study Stochastic Differential Equations with Reflection (SDERs). These equations are intimately related to the Constrained Langevin Approximation (CLA), since, as indicated in Remark 2.1, a CLA is (part of) a solution to a SDER. We start by giving the appropriate definitions regarding SDERs in Section 5.1, followed

by a result on uniqueness in Section 5.2. In Section 5.3 we provide an auxiliary result, Lemma 5.1. In Section 5.4 we give conditions for existence of solutions to SDERs, and follow with our main result for this Section 5, namely Theorem 5.2, where we give an existence and uniqueness result for the equation involving the CLA. Finally, in Section 5.5 we provide a time change result (Lemma 5.2) which will be useful in the proofs of Theorems 3.2 and 3.4.

5.1. SDER

As before, let $\mathcal{I} \subseteq \mathbb{R}_+$ be a closed interval of the form $[0, a]$, where $a > 0$, or the interval $[0, \infty)$. Let $d \geq 1$ be an integer and let $b : \mathcal{I} \longrightarrow \mathbb{R}$ and $\rho : \mathcal{I} \longrightarrow \mathbb{R}^d$ be Borel-measurable functions. Additionally, consider a real number $\alpha > 0$ and the function $\gamma : \mathcal{I} \longrightarrow \mathbb{R}$ given by (2.7).

Definition 5.1. Given $(b, \rho, \alpha, \gamma)$, we say that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ is a **solution of the following SDER**

$$Z(t) = Z(0) + \int_0^t b(Z(s))ds + \int_0^t \rho(Z(s))dW(s) + \alpha \int_0^t \gamma(Z(s))dL(s), \tag{5.1}$$

where

(i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_t\}$ is a filtration on $\mathcal{F}$ such that the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ satisfies the usual conditions,

(ii) $Z = \{Z(t),\ 0 \leq t < \infty\}$ is a continuous, $\{\mathcal{F}_t\}$-adapted process such that $\mathbb{P}$-a.s. $Z(t) \in \mathcal{I}$ for all $t \geq 0$,

(iii) $W = \{W(t) = (W_1(t), \ldots, W_d(t)),\ 0 \leq t < \infty\}$ is a d-dimensional standard Brownian motion and an $\{\mathcal{F}_t\}$-martingale under $\mathbb{P}$,

(iv) $L = \{L(t),\ 0 \leq t < \infty\}$ is a continuous, $\{\mathcal{F}_t\}$-adapted, one-dimensional, $\mathbb{P}$-a.s. non-decreasing process, with $\mathbb{P}[L(0) = 0] = 1$ and such that $\mathbb{P}$-a.s:

$$L(t) = \int_0^t \mathbb{1}_{\{Z(s) \in \partial\mathcal{I}\}} dL(s), \qquad \text{for every } t \geq 0, \tag{5.2}$$

(v) for every $t \geq 0$, $\mathbb{P}$-a.s.

$$\int_0^t |b(Z(s))|ds < \infty, \quad \text{and} \quad \int_0^t |\rho(Z(s))|^2 ds < \infty, \tag{5.3}$$

(vi) and the triple (Z, W, L) $\mathbb{P}$-a.s. satisfies:

$$Z(t) = Z(0) + \int_0^t b(Z(s))ds + \sum_{i=1}^d \int_0^t \rho_i(Z(s))dW_i(s) + \alpha \int_0^t \gamma(Z(s))dL(s) \tag{5.4}$$

for every $t \geq 0$.

For a Borel probability measure ν on $\mathcal{I}$, we will say that a solution to (5.1) has **initial distribution** ν, provided $Z(0)$ is distributed as ν, under $\mathbb{P}$.

The notion in Definition 5.1 is commonly referred to as a *weak* solution. For the sake of simplicity, we will just call it a solution in this paper. Additionally, we will say that Z is a solution of (5.1), meaning that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ is a solution of (5.1).

Remark 5.1. We introduce the parameter α for convenience. In fact, for a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ to (5.1) with coefficients $(b, \rho, \alpha, \gamma)$, we can define $\tilde{L} := \alpha L$ and get a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, \tilde{L})$ to (5.1) with coefficients $(b, \rho, 1, \gamma)$, and vice versa. The reason for introducing the parameter $\alpha > 0$ is to emphasize the order of the reflection term $\alpha \int_0^t \gamma(Z(s))dL(s)$ because in the context of the CLA (equation (2.9)), α takes the value $r^{-1/2}$, for $r \in \mathcal{R}$.

For a better understanding of the CLA, we explore further properties of SDERs. We start by describing two notions of uniqueness for equation (5.1).

Definition 5.2. We say that **uniqueness in law** or **weak uniqueness** holds for equation (5.1) if for every pair of solutions

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L) \text{ and } (\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{\mathbb{P}}, \tilde{Z}, \tilde{W}, \tilde{L})$$

with the same initial distribution ($Z(0) \overset{d}{=} \tilde{Z}(0)$), we have that (Z, L) and $(\tilde{Z}, \tilde{L})$ have the same law.

Definition 5.3. We will say that **pathwise uniqueness** holds for equation (5.1) if for every pair of solutions:

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L) \text{ and } (\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, \tilde{Z}, W, \tilde{L})$$

with common Brownian motion W, common filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and such that $\mathbb{P}[Z(0) = \tilde{Z}(0)] = 1$, we have that:

$$\mathbb{P}[Z(t) = \tilde{Z}(t)\ , L(t) = \tilde{L}(t) \text{ for every } 0 \leq t < \infty] = 1. \tag{5.5}$$

The relation between these two notions is stated in Proposition 5.1.

A second type of solution is the so-called *strong* solution. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a standard d-dimensional Brownian motion W defined there, together with a random variable ξ taking values in $\mathcal{I}$ such that ξ is independent of W. For the filtration $\{\mathcal{G}_t := \sigma(\xi, W(s); 0 \leq s \leq t) : 0 \leq t < \infty\}$ and the collection of null sets $\mathcal{N} := \{N \in \mathcal{F} \mid \mathbb{P}(N) = 0\}$, we define $\hat{\mathcal{F}}_t^W := \sigma(\mathcal{G}_t \cup \mathcal{N})$ for $t \geq 0$. Then $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P})$ satisfies the usual conditions. In addition, ξ is $\hat{\mathcal{F}}_0^W$-measurable and W is an $\{\hat{\mathcal{F}}_t^W\}$-martingale. We call $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$ a **setup**.

Definition 5.4. Given $(b, \rho, \alpha, \gamma)$ and a setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$, a **strong solution to the SDER** (5.1) with initial condition ξ, is a pair (Z, L) of one-dimensional processes defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

(i) $Z = \{Z(t),\ 0 \leq t < \infty\}$ is a continuous, $\{\hat{\mathcal{F}}_t^W\}$-adapted process such that $\mathbb{P}$- a.s. $Z(t) \in \mathcal{I}$ for all $t \geq 0$, and $\mathbb{P}[Z(0) = \xi] = 1$,

(ii) $L = \{L(t),\ 0 \leq t < \infty\}$ is a continuous, $\{\hat{\mathcal{F}}_t^W\}$-adapted, $\mathbb{P}$-a.s. non-decreasing process, with $\mathbb{P}[L(0) = 0] = 1$ and such that $\mathbb{P}$-a.s. (5.2) holds.

(iii) Conditions (5.3) and (5.4) hold.

In this article we will always emphasize *strong* when working with a strong solution. Note that in Definitions 5.1 and 5.4, Z is one-dimensional, whereas W can be multidimensional. This formulation will be used in constructing the CLA, in the proofs of Theorems 3.2 and 3.4. In Lemma 5.1 we will show how to *merge* the d stochastic integrals of (5.4) into one.

Definition 5.5. We say that **existence** holds for equation (5.1) if for every Borel probability measure ν on $\mathcal{I}$, there exists a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ to (5.1) with initial distribution ν. Additionally, we say that **strong existence** holds for equation (5.1) if for every setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$, a strong solution (Z, L) to equation (5.1) with initial condition ξ, exists. Finally, we say that **strong uniqueness** holds for equation (5.1) if for every setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$ and for every pair of strong solutions (Z, L) and $(\tilde{Z}, \tilde{L})$ with initial condition ξ, (5.5) holds.

A famous result of Yamada & Watanabe [24] can be readily generalized from SDEs to SDERs. Accordingly, we state the following proposition without proof.

Proposition 5.1. (**Yamada & Watanabe**.) *Pathwise uniqueness implies uniqueness in law for equation* (5.1). *Additionally, existence and pathwise uniqueness for equation* (5.1) *implies strong existence and uniqueness.*

Along the same lines, we state the next technical result, which will be used in the sequel. The proof follows in the same way as for Corollary 5.3.23 in Karatzas & Shreve [10] and it is therefore omitted.

Proposition 5.2. *Suppose pathwise uniqueness holds for equation* (5.1). *Let ν be a Borel probability measure on $\mathcal{I}$. If there exists a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ to equation* (5.1) *with initial distribution ν, then, for every setup $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t^{\tilde{W}}\}, \tilde{\mathbb{P}}, \tilde{W}, \tilde{\xi})$, where $\tilde{\xi}$ has distribution ν, there exists a strong solution $(\tilde{Z}, \tilde{L})$ to equation* (5.1) *with initial condition $\tilde{\xi}$.*

Equation (2.9) is of the form of (5.1) with $b = \mu$, $\rho = \frac{1}{\sqrt{r}}\sigma$, $\alpha = \frac{1}{\sqrt{r}}$, γ as in (2.7) and $d = 1$. In addition, we will also consider solutions of

$$\begin{aligned} Z^r(t) = Z^r(0) &+ \int_0^t \mu(Z^r(s))ds + \frac{1}{\sqrt{r}}\sum_{j=1}^n \int_0^t \ell_j \sqrt{\beta(Z^r(s), \ell_j)}dW_j(s) \\ &+ \frac{1}{\sqrt{r}}\int_0^t \gamma(Z^r(s))dL^r(s), \end{aligned} \tag{5.6}$$

for each $r \in \mathcal{R}$, which corresponds to (5.1) with $b = \mu$, $\rho_j = \frac{1}{\sqrt{r}}\ell_j\sqrt{\beta(\cdot, \ell_j)}$ for $j = 1, \ldots, n$, $\alpha = \frac{1}{\sqrt{r}}$, γ as in (2.7) and $d = n$. These equations are needed for technical reasons, as a key step in constructing the coupling in Theorems 3.2 and 3.4.

In the following, we study equations of the form (5.1), with an eye towards (2.9) and (5.6). We start by studying uniqueness in the next section.

5.2. Uniqueness

We could not find any place in the literature where the conditions in this section were given before for SDERs. We allow both the drift and dispersion coefficient to be more general than locally Lipschitz.

Theorem 5.1. *Consider equation (5.1) with coefficients $(b, \rho, \alpha, \gamma)$. Suppose that for every compact set $\mathcal{K} \subseteq \mathcal{I}$, there exist strictly increasing functions $h_{\mathcal{K}}, g_{\mathcal{K}} : [0, \infty) \longrightarrow [0, \infty)$ with $g_{\mathcal{K}}(0) = h_{\mathcal{K}}(0) = 0$, where $g_{\mathcal{K}}$ is continuous and concave, and such that:*

$$\int_{(0,\varepsilon)} \frac{1}{g_{\mathcal{K}}(u)} du = \int_{(0,\varepsilon)} \frac{1}{h^2_{\mathcal{K}}(u)} du = \infty, \qquad \text{for every } \varepsilon > 0, \tag{5.7}$$

and

$$|b(x) - b(y)| \leq g_{\mathcal{K}}(|x - y|), \qquad |\rho_i(x) - \rho_i(y)| \leq h_{\mathcal{K}}(|x - y|) \tag{5.8}$$

for every $x, y \in \mathcal{K}$ and $i = 1, \ldots, d$. Then, pathwise uniqueness holds for (5.1). In particular, it holds for $g_{\mathcal{K}}(u) = A_{\mathcal{K}} u$ and $h_{\mathcal{K}}(u) = A_{\mathcal{K}} u^{\alpha}$ with $\alpha \geq 1/2$, where $A_{\mathcal{K}} > 0$ is a constant.

Yamada in [23] proved this result for the case $\mathcal{I} = [0, \infty)$ and $d = 1$, under the assumptions $g_{\mathcal{K}}(u) = Au$, where $A > 0$ is a constant, and $h_{\mathcal{K}}$ being the same function for every compact set $\mathcal{K} \subseteq \mathcal{I}$. The technique used in [23] can be extended to the more general case described above, as we now show.

Proof. We first consider the case where $\mathcal{I} = [0, a]$, for some $a > 0$. Because of this, we only consider the functions $h_{\mathcal{I}}$ and $g_{\mathcal{I}}$ and omit the subscript $\mathcal{I}$. Under condition (5.7) we can construct an increasing sequence $\{\psi_m\}_{m \geq 1}$ of real-valued $\mathcal{C}^2(\mathbb{R})$ functions such that the following properties (i)-(iii) hold for every $m \geq 1$ (for an illustration of how to construct such an increasing sequence see Proposition 5.2.13 in Karatzas & Shreve [10]) :

(i) ψ_m is even, $\psi_m \geq 0$, $\psi_m(0) = 0$, $\psi_m(x) \leq \psi_{m+1}(x)$ and $\lim_{m \to \infty} \psi_m(x) = |x|$ for every $x \in \mathbb{R}$,

(ii) $|\psi'_m| \leq 1$, $\psi'_m(x) \geq 0$ for $x \geq 0$ and $\psi'_m(x) \leq 0$ for $x \leq 0$,

(iii) $\psi''_m \geq 0$ and $\psi''_m(x) h^2(|x|) \leq \frac{2}{m}$ for every $x \in \mathbb{R}$.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z_1, W, L_1)$ and $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z_2, W, L_2)$ be two solutions of (5.1) with common Brownian motion W, common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, common filtration $\{\mathcal{F}_t\}$ and such that $\mathbb{P}[Z_1(0) = Z_2(0)] = 1$. We define $\Delta(t) := Z_1(t) - Z_2(t)$ for $t \geq 0$. For $m \geq 1$, we apply Itô's formula and obtain that $\mathbb{P}$-a.s. for all $t \geq 0$, $\psi_m(\Delta(t))$ equals

$$\begin{aligned}
&\int_0^t \psi'_m(\Delta(s))(b(Z_1(s)) - b(Z_2(s)))ds + \alpha \int_0^t \psi'_m(\Delta(s))\gamma(Z_1(s))dL_1(s) \\
&- \alpha \int_0^t \psi'_m(\Delta(s))\gamma(Z_2(s))dL_2(s) + \frac{1}{2}\sum_{i=1}^{d} \int_0^t \psi''_m(\Delta(s))(\rho_i(Z_1(s)) - \rho_i(Z_2(s)))^2 ds \\
&+ \sum_{i=1}^{d} \int_0^t \psi'_m(\Delta(s))(\rho_i(Z_1(s)) - \rho_i(Z_2(s)))dW_i(s).
\end{aligned} \tag{5.9}$$

Fix $t \geq 0$. We claim that

$$\alpha \int_0^t \psi'_m(\Delta(s))\gamma(Z_1(s))dL_1(s) \leq 0, \qquad \mathbb{P} - \text{a.s.} \tag{5.10}$$

For this, observe first that

$$\alpha \int_0^t \psi_m'(\Delta(s))\gamma(Z_1(s))dL_1(s) = \alpha \int_0^t \psi_m'(\Delta(s))\mathbb{1}_{\{Z_1(s)=0\}}dL_1(s) - \alpha \int_0^t \psi_m'(\Delta(s))\mathbb{1}_{\{Z_1(s)=a\}}dL_1(s).$$

Now, for $0 \leq s \leq t$, if $Z_1(s) = 0$, then $\Delta(s) = -Z_2(s) \leq 0$ and $\psi_m'(\Delta(s)) \leq 0$. So, $\mathbb{P}$-a.s., $\alpha \int_0^t \psi_m'(\Delta(s))\mathbb{1}_{\{Z_1(s)=0\}}dL_1(s) \leq 0$. On the other hand, for $0 \leq s \leq t$, if $Z_1(s) = a$, then $\Delta(s) = a - Z_2(s) \geq 0$ and $\psi_m'(\Delta(s)) \geq 0$. Then, $\mathbb{P}$-a.s. $-\alpha \int_0^t \psi_m'(\Delta(s))\mathbb{1}_{\{Z_1(s)=a\}}dL_1(s) \leq 0$. Thus, (5.10) holds. In a similar fashion, we can show that $-\alpha \int_0^t \psi_m'(\Delta(s))\gamma(Z_2(s))dL_2(s) \leq 0$, $\mathbb{P}$-a.s. Combining these results with the martingale property of the stochastic integrals, the bound of one on $|\psi_m'|$, using (5.8), on taking expectations in (5.9) we obtain

$$\begin{aligned}\mathbb{E}[\psi_m(\Delta(t))] &\leq \mathbb{E}\left[\int_0^t \psi_m'(\Delta(s))(b(Z_1(s)) - b(Z_2(s)))ds\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^d \int_0^t \psi_m''(\Delta(s))(\rho_i(Z_1(s)) - \rho_i(Z_2(s)))^2 ds\right] \\ &\leq \int_0^t \mathbb{E}[g(|\Delta(s)|)]ds + \frac{td}{m} \leq \int_0^t g(\mathbb{E}[|\Delta(s)|])ds + \frac{td}{m},\end{aligned}$$

where the last inequality follows by the concavity of g and Jensen's inequality. On letting $m \to \infty$ and applying Lemma 1.4.1. in Agarwal & Lakshmikantham [1] (related to Osgood's criterion), we obtain that for every $t \geq 0$, $\mathbb{E}[|\Delta(t)|] = 0$. This implies that $\mathbb{P}[Z_1(t) = Z_2(t), \text{ for every } t \geq 0] = 1$.

For $k = 1, 2$ define $Y_k(t) := \int_0^t \gamma(Z_k(s))dL_k(s)$, for $t \geq 0$. Since both (Z_1, L_1) and (Z_2, L_2) satisfy (5.4) with common Brownian motion, and $Z_1 = Z_2$ $\mathbb{P}$-a.s. by the previous analysis, we must have $\mathbb{P}[Y_1(t) = Y_2(t), \text{ for every } t \geq 0] = 1$. Next, observe that $\mathbb{P}$-a.s. for every $t \geq 0$ and $k = 1, 2$: $\int_0^t \gamma(Z_k(s))dY_k(s) = \int_0^t \gamma^2(Z_k(s))dL_k(s) = \int_0^t \mathbb{1}_{\{Z_k(s)\in\partial\mathcal{I}\}}dL_k(s) = L_k(t)$, where we used (5.2). We conclude that $\mathbb{P}[L_1(t) = L_2(t), \text{ for every } t \geq 0] = 1$.

For the case $\mathcal{I} = [0, \infty)$, consider $K > 0$ and define $T_K = \inf\{t \geq 0 \mid Z_1(t) \geq K \text{ or } Z_2(t) \geq K\}$. This case follows by choosing $\mathcal{K} = [0, K]$, constructing the family $\{\psi_m^{\mathcal{K}}\}_{m\geq 1}$ related to $h_{\mathcal{K}}$, replacing t with $t \wedge T_K$ in (5.9), estimating in a similar manner to the previous case and then letting $K \to \infty$. □

Corollary 5.1. *Suppose Assumption 3.1 or 3.2 holds. Then, pathwise uniqueness (and uniqueness in law) holds for equation (2.9) and also for (5.6).*

Proof. Under either Assumption 3.1 or 3.2, the functions $\beta(\cdot, \ell_j)$, $j = 1, \ldots, n$ are locally Lipschitz continuous on $\mathcal{I}$. Therefore, the functions μ and σ^2 are locally Lipschitz continuous on $\mathcal{I}$. By using the fact that $x \mapsto x^{1/2}$ is Hölder continuous of order $1/2$ on $[0, \infty)$, we see that condition (5.8) holds for (2.9) and (5.6) and pathwise uniqueness follows for these equations by Theorem 5.1. By Proposition 5.1, uniqueness in law holds for (2.9) and (5.6). □

5.3. Merging stochastic integrals

Equations (2.9) and (5.6) are related in the following way. Denote the dispersion coefficient of (5.6) by $\rho^r := (\rho_1^r, \ldots, \rho_n^r)$, where $\rho_j^r(x) := \frac{\ell_j}{\sqrt{r}}\sqrt{\beta(x,\ell_j)}$ for $j = 1, \ldots, n$ and $x \in \mathcal{I}$. Then,

$$\sum_{j=1}^{n}(\rho_j^r(x))^2 = \left(\frac{1}{\sqrt{r}}\sigma(x)\right)^2 \qquad \text{for } x \in \mathcal{I}. \tag{5.11}$$

The next lemma implies that we can merge the n stochastic integrals of (5.6) into a single stochastic integral in (2.9). We state this in the general context of equation (5.1) and the equation:

$$Z(t) = Z(0) + \int_0^t b(Z(s))ds + \int_0^t \Upsilon(Z(s))d\hat{W}(s) + \alpha\int_0^t \gamma(Z(s))dL(s), \tag{5.12}$$

where $\Upsilon(x) := \left(\sum_{i=1}^d \rho_i^2(x)\right)^{1/2}$ for $x \in \mathcal{I}$, and $\hat{W}$ is a one-dimensional standard Brownian motion. The following is similar to results in the literature on the representation of continuous local-martingales, such as Theorem 3.4.2 in Karatzas & Shreve [10].

Lemma 5.1. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ be a solution of equation (5.1). Then, there exists a one-dimensional standard Brownian motion $\hat{W} = \{\hat{W}(t),\ 0 \le t < \infty\}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, which is an $\{\mathcal{F}_t\}$-martingale there and such that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, \hat{W}, L)$ is a solution of equation (5.12).*

Proof. The process $M = \{M(t),\ 0 \le t < \infty\}$, defined by

$$M(t) := \sum_{i=1}^{d}\int_0^t \rho_i(Z(s))dW_i(s), \qquad t \ge 0,$$

is a continuous $\{\mathcal{F}_t\}$-local martingale such that $\mathbb{P}[M(0) = 0] = 1$. Its quadratic variation process is given by $\langle M\rangle(t) = \int_0^t \Upsilon^2(Z(s))ds$ for $t \ge 0$. Consider the process

$$\hat{W}(t) := \int_0^t \frac{1}{\Upsilon(Z(s))}\mathbb{1}_{\{\Upsilon(Z(s))\neq 0\}}dM(s) + \int_0^t \mathbb{1}_{\{\Upsilon(Z(s))=0\}}dW_1(s), \qquad t \ge 0. \tag{5.13}$$

Then $\hat{W} = \{\hat{W}(t),\ 0 \le t < \infty\}$ is a continuous $\{\mathcal{F}_t\}$-local martingale, such that $\mathbb{P}[\hat{W}(0) = 0] = 1$ and with quadratic variation process given by

$$\begin{aligned}\langle\hat{W}\rangle(t) &= \left\langle \int_0^\bullet \frac{1}{\Upsilon(Z(s))}\mathbb{1}_{\{\Upsilon(Z(s))\neq 0\}}dM(s)\right\rangle(t) + \left\langle \int_0^\bullet \mathbb{1}_{\{\Upsilon(Z(s))=0\}}dW_1(s)\right\rangle(t)\\ &= \int_0^t \mathbb{1}_{\{\Upsilon(Z(s))\neq 0\}}ds + \int_0^t \mathbb{1}_{\{\Upsilon(Z(s))=0\}}ds = t,\end{aligned}$$

for every $t \ge 0$. By Levy's characterization, $\hat{W}$ is a standard Brownian motion and an

$\{\mathcal{F}_t\}$-martingale. Moreover, for $t \geq 0$,

$$\begin{aligned}\int_0^t \Upsilon(Z(s))d\hat{W}(s) &= \int_0^t \Upsilon(Z(s))\frac{\mathbb{1}_{\{\Upsilon(Z(s))\neq 0\}}}{\Upsilon(Z(s))}dM(s) \\ &\quad + \int_0^t \Upsilon(Z(s))\mathbb{1}_{\{\Upsilon(Z(s))=0\}}dW_1(s) \\ &= \int_0^t \mathbb{1}_{\{\Upsilon(Z(s))\neq 0\}}dM(s).\end{aligned}$$

Since $\langle M\rangle(t) = \int_0^t \Upsilon^2(Z(s))ds$, it follows that $\int_0^t \mathbb{1}_{\{\Upsilon(Z(s))=0\}}dM(s) = 0$ for every $t \geq 0$. Hence, $\int_0^t \Upsilon(Z(s))d\hat{W}(s) = M(t)$ for every $t \geq 0$ and we can conclude the desired result. □

Corollary 5.2. *For $r \in \mathcal{R}$, if $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W, L^r)$ is a solution to (5.6), then there exists a standard Brownian motion W^r such that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W^r, L^r)$ is a solution to (2.9).*

5.4. Existence and uniqueness

Tanaka in [22] studied SDERs like (5.1) in a more general setting. Namely, he considered the process Z to be multi-dimensional, taking values in a convex set and with reflection at the boundary given by the inward pointing normal. The following is a direct consequence of Theorem 4.2 in [22] applied in the one-dimensional case.

Proposition 5.3. (**Tanaka**.) *Let $x_0 \in \mathcal{I}$. If b and ρ are bounded and continuous, then there exists a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ to equation (5.1) such that $Z(0) = x_0$ $\mathbb{P}$-a.s.*

Using this result as a starting point, we proceed to establish strong existence for equations (2.9) and (5.6). We include strong uniqueness and uniqueness in law in the statement below, which comes from Corollary 5.1.

Theorem 5.2. *Suppose Assumption 3.1 or 3.2 holds. Then, existence and uniqueness in law hold for equations (2.9) and (5.6), as well as strong existence and uniqueness.*

Proof. It suffices to prove existence for (5.6). Indeed, by Corollary 5.2, a solution to (5.6) with initial distribution ν will generate a solution to (2.9) with initial distribution ν. Then pathwise uniqueness (established in Corollary 5.1) and Proposition 5.1 enable us to conclude that strong existence holds for both (2.9) and (5.6).

In order to prove existence for (5.6), first consider the case $\mathcal{I} = [0, a]$. Under Assumption 3.1, Proposition 5.3 implies the existence of solutions to (5.6) for initial distributions charging a single point. Now, consider an initial distribution ν on $\mathcal{I}$ and a setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$, where W is an n-dimensional Brownian motion and ξ has distribution ν. Then, for $r \in \mathcal{R}$, for each integer $k \geq 1$, let $\xi_k := \sum_{m=0}^{\infty} \frac{m}{2^k}\mathbb{1}_{\{\frac{m}{2^k} \leq \xi < \frac{m+1}{2^k}\}}$. Then, $\xi_k \nearrow \xi$ as $k \to \infty$. By Proposition 5.2, for each pair of integers $k \geq 1$ and $0 \leq m \leq \lfloor a2^k \rfloor$ there is a strong solution $(Z_k^{r,m}, L_k^{r,m})$ to equation (5.6) on $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$, with $Z_k^{r,m}(0) = \frac{m}{2^k}$. Then

$$Z_k^r(t) := \sum_{m=0}^{\lfloor a2^k \rfloor} Z_k^{r,m}(t)\mathbb{1}_{\{\xi_k = \frac{m}{2^k}\}}, \quad L_k^r(t) := \sum_{m=0}^{\lfloor a2^k \rfloor} L_k^{r,m}(t)\mathbb{1}_{\{\xi_k = \frac{m}{2^k}\}}, \qquad t \geq 0, \quad (5.14)$$

is a strong solution (Z_k^r, L_k^r) on $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W)$ to equation (5.6) with initial condition ξ_k. The sequence $\{(Z_k^r, L_k^r, W)\}_{k\geq 1}$ can be proved to be $\mathcal{C}$-tight and moreover, it can be proved that every subsequential weak limit generates a solution to (5.6) with initial distribution ν. By weak uniqueness, the sequence actually converges weakly and generates a (weak) solution to (5.6) with initial distribution ν. By Proposition 5.1, there exists a strong solution to (5.6) on $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W)$.

For the case $\mathcal{I} = [0, \infty)$, the coefficients in (5.6) are not necessarily bounded and, we cannot apply Proposition 5.3 directly. For any integer $k \geq 1$, define $\mu_k(x) := \mu(x \wedge k)$ and $\rho_{j,k}^r(x) := \rho_j^r(x \wedge k)$ for $x \in \mathcal{I}$ and $1 \leq j \leq n$, where $\rho_j^r(x) := \frac{\ell_j}{\sqrt{r}}\sqrt{\beta(x, \ell_j)}$ for $x \in \mathcal{I}$ and $1 \leq j \leq n$. Additionally, consider the following equation for $t \geq 0$:

$$\begin{aligned} Z_k^r(t) = Z_k^r(0) + \int_0^t \mu_k(Z_k^r(s))ds + \sum_{j=1}^n \int_0^t \rho_{j,k}^r(Z_k^r(s))dW_j(s) \\ + \frac{1}{\sqrt{r}} \int_0^t \gamma(Z_k^r(s))dL_k^r(s). \end{aligned} \tag{5.15}$$

For each $k \geq 1$, the coefficient μ_k is Lipschitz continuous and bounded, while the $\rho_{j,k}^r$ are Hölder continuous of order $1/2$, and bounded, for every $1 \leq j \leq n$. For $x_0 \in \mathcal{I}$, consider a setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$ where $\mathbb{P}[\xi = x_0] = 1$. By combining Theorem 5.1, Proposition 5.3 and Proposition 5.2 we obtain a sequence of processes $\{(Z_k^r, L_k^r)\}_{k\geq 1}$, where for every $k \geq 1$, (Z_k^r, L_k^r) is a strong solution to (5.15) with initial condition ξ on the setup $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t^W\}, \mathbb{P}, W, \xi)$. Following an approach similar to the proof of Theorem 10.6 in Chung & Williams [8], and by using the linear growth condition (3.7), we can show that $\mathbb{P}$-a.s. the limits $Z^r(t) := \lim_{k\to\infty} Z_k^r(t)$ and $L^r(t) := \lim_{k\to\infty} L_k^r(t)$ exist for every $t \geq 0$ and they define a strong solution to equation (5.6) with initial condition $Z^r(0) = x_0$.

For an arbitrary initial distribution ν, we prove existence in a similar manner to that for the case of a bounded interval $\mathcal{I}$. □

5.5. Time change

We end this section with a result needed in the proofs of Theorems 3.2 and 3.4.

Lemma 5.2. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W, L^r)$ $r \in \mathcal{R}$, be a family of solutions of (5.6) and, by extending the probability space if needed, let $\tilde{W}$ be an n-dimensional standard Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that is independent of Z^r, W, L^r. Then, for each $r \in \mathcal{R}$, there exists an n-dimensional standard Brownian motion $B^r = (B_1^r, \ldots, B_n^r)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}$-a.s., $Z^r(t)$ is equal to*

$$Z^r(0) + \int_0^t \mu(Z^r(s))ds + \sum_{j=1}^n \frac{\ell_j}{r} B_j^r\left(r\int_0^t \beta(Z^r(s), \ell_j)ds\right) + \frac{1}{\sqrt{r}}\int_0^t \gamma(Z^r(s))dL^r(s) \tag{5.16}$$

for every $t \geq 0$. Furthermore, $\sigma(B) \subseteq \sigma(Z^r, W, \tilde{W}, \mathcal{N})$ where $\mathcal{N}$ is the collection of null sets in $(\Omega, \mathcal{F}, \mathbb{P})$.

The statement of Lemma 5.2 is similar to that for Theorem 6.5.3.b in Ethier & Kurtz [9], except that here, (5.16) has an additional reflection term. The underlying idea of the proof is similar, namely, a continuous local martingale can be time-changed to a

Brownian motion. Moreover, for a vector of continuous local martingales with cross variation equal to zero, we can time-change each component and obtain, by means of Knight's Theorem (Theorem V.1.9 and V.1.10 in Revuz & Yor [19]), a multidimensional Brownian motion. For completeness, we provide a proof of Lemma 5.2.

Proof. For $j = 1, \dots, n$ define:

$$M_j^r(t) := \int_0^t \sqrt{r\beta(Z^r(s), \ell_j)} dW_j(s), \qquad t \geq 0. \tag{5.17}$$

Then, $M^r := (M_1^r, \dots, M_n^r)$ is a vector of continuous $\{\mathcal{F}_t\}$-local martingales with $\langle M_j^r \rangle(t) = r \int_0^t \beta(Z^r(s), \ell_j) ds$ for $t \geq 0$, $M_j^r(0) = 0$ $\mathbb{P}$-a.s., and with cross variation $\langle M_i^r, M_j^r \rangle(t) = 0$ for $i \neq j$ and $t \geq 0$. By Proposition 1.26 in Chapter IV of Revuz & Yor [19], $\lim_{t\to\infty} M_j^r(t)$ exists $\mathbb{P}$-a.s. on $\{\langle M_j^r \rangle(\infty) < \infty\}$ and we let $M_j^r(\infty)$ be a random variable that is $\mathbb{P}$-a.s. equal to this limit on $\{\langle M_j^r \rangle(\infty) < \infty\}$. Now, for $s \geq 0$ and $j = 1, \dots, n$, define

$$T_j^r(s) := \inf\{t \geq 0 \mid \langle M_j^r \rangle(t) > s\}$$

and

$$B_j^r(s) := \begin{cases} M_j^r(T_j^r(s)) & \text{if } s < \langle M_j^r \rangle(\infty), \\ M_j^r(\infty) + \tilde{W}_j(s - \langle M_j^r \rangle(\infty)) & \text{if } s \geq \langle M_j^r \rangle(\infty). \end{cases} \tag{5.18}$$

By Knight's Theorem (as in Theorem V.1.9 and V.1.10 in Revuz & Yor [19]), the process $B^r := (B_1^r, \dots, B_n^r)$ is an n-dimensional Brownian motion. For $j = 1, \dots, n$, $\mathbb{P}$-a.s. $B_j^r\left(\langle M_j^r \rangle(t)\right) = M_j^r(t)$ for every $t \geq 0$. We conclude by using these identities in (5.6). □

6. Proofs of main results

In this section we give the proofs for our main results. We start by providing an outline.

6.1. Outline for the proofs of Theorems 3.2 and 3.4

(i) Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with an n-dimensional Brownian motion W defined there, for each $r \geq 8$ in $\mathcal{R}$ and $x_0^r \in \overline{\mathcal{I}}^r$, consider a strong solution (Z^r, L^r) to equation (5.6). Existence is guaranteed by Theorem 5.2.

(ii) Using the solution of (5.6), we construct a solution $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z^r, W^r, L^r)$ to (2.9) and thereby, obtain a CLA Z^r. The construction is given in Lemma 5.1.

(iii) Using a time change, we obtain an n-dimensional Brownian motion B^r, and Z^r can be described as a solution to equation (5.16) involving B^r.

(iv) From the Brownian motion B^r, and independent uniform random variables, we construct independent Poisson processes $N_1^r, \dots, N_n^r$ coupled appropriately to B^r (see Theorem A.1).

(v) Using these Poisson processes we generate the process X^r, as a solution to equation (6.2), following Theorem 6.4.1 in Ethier & Kurtz [9].

(vi) Having constructed both processes on the same space, we compare their paths. A critical property is the Lipschitz continuity of the Skorokhod map (see in Appendix B).

6.2. Proof of Theorem 3.2

Let β and $\{\epsilon^r\}_{r\in\mathcal{R}}$ be the functions associated to a given NDDF under Assumptions 2.1 and 3.1. Consider a complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ on which there is a standard n-dimensional Brownian motion $W=\{W(t),\ 0\le t<\infty\}$, a standard n-dimensional Brownian motion $\tilde{W}=\{\tilde{W}(t),\ 0\le t<\infty\}$ and n double sequences $U^1=(U^1_{i,j})_{i,j\ge1},\dots,U^n=(U^n_{i,j})_{i,j\ge1}$ of i.i.d. random variables with common uniform distribution on $(0,1)$, such that the elements $W,\tilde{W},U^1,\dots,U^n$ are independent. Consider $\{\mathcal{F}^W_t\}$ to be the natural filtration for W where $\mathcal{F}^W_t=\sigma(W(s)\,;\,0\le s\le t)$ for each $t\ge0$, and let $\mathcal{N}$ denote the null sets in $(\Omega,\mathcal{F},\mathbb{P})$. Define the filtration $\{\mathcal{F}_t\}$ as $\mathcal{F}_t:=\sigma(\mathcal{F}^W_t\cup\mathcal{N})$ for $t\ge0$. This filtration is such that the quadruple $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P})$ satisfies the usual conditions.

Fix an $r\ge8$ in $\mathcal{R}$ and $x^r_0\in\overline{\mathcal{I}}^r$. Define the random variable $\xi_{x^r_0}(\cdot):=x^r_0$. The σ-algebra $\sigma(\xi_{x^r_0})$ is trivial and therefore, independent from $\sigma(W(s):s\ge0)$. As for Definition 5.4, we can construct the filtration $\{\hat{\mathcal{F}}^W_t\}$ and obtain a setup $(\Omega,\mathcal{F},\{\hat{\mathcal{F}}^W_t\},\mathbb{P},W,\xi_{x^r_0})$. Since $\sigma(\xi_{x^r_0})$ is trivial, $\{\hat{\mathcal{F}}^W_t\}=\{\mathcal{F}_t\}$. By Theorem 5.2, there is a strong solution (Z^r,L^r) of (5.6) with initial condition $\xi_{x^r_0}$. By Corollary 5.2, there is a one-dimensional Brownian motion $W^r=\{W^r(t),\mathcal{F}_t,\ 0\le t<\infty\}$ on $(\Omega,\mathcal{F},\mathbb{P})$ so that $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P},Z^r,W^r,L^r)$ is a solution to (2.9). In other words, Z^r is a CLA associated with $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P},Z^r,W^r,L^r)$. In addition, $\mathbb{P}$-a.s. $Z^r(0)=x^r_0$.

Since $\tilde{W}$ is independent of W, and for each $r\in\mathcal{R}$, (Z^r,L^r) is a strong solution for (5.6) driven by W, then $\tilde{W}$ will be independent of W,Z^r,L^r. By Lemma 5.2, for each $r\in\mathcal{R}$, we can construct an n-dimensional standard Brownian motion $B^r=(B^r_1,\dots,B^r_n)$ such that $\mathbb{P}$-a.s. the time-change equation (5.16) holds.

We continue by coupling B^r to a vector of n independent Poisson processes. First, note that by Lemma 5.2, $\sigma(B^r)\subseteq\sigma(W,\tilde{W},\mathcal{N})$ and so B^r is independent of $U^1=(U^1_{i,j})_{i,j\ge1},\dots,U^n=(U^n_{i,j})_{i,j\ge1}$. By Theorem A.1, there are n mutually independent unit rate Poisson processes $N^r_1,\dots,N^r_n$ defined on $(\Omega,\mathcal{F},\mathbb{P})$ such that for each $1\le j\le n$,

$$\mathbb{P}\left[\sup_{0\le s\le t}|N^r_j(s)-s-B^r_j(s)|>(C+\theta)\log t+x\right]\le Kt^{-\theta\lambda}e^{-\lambda x} \tag{6.1}$$

for every $t>1$, $x>0$ and $\theta>0$, where $C,K,\lambda>0$ are universal constants. Recall the definition of $\lambda^r(\cdot,\cdot)$ from (2.1). Consider the following equation:

$$X^r(t)=rx^r_0+\sum_{j=1}^n\ell_jN^r_j\left(r\int_0^t\lambda^r\left(\frac{X^r(s)}{r},\ell_j\right)ds\right),\qquad t\ge0. \tag{6.2}$$

By Theorem 6.4.1 of Ethier & Kurtz [9], this equation has a unique solution $X^r=\{X^r(t),\ 0\le t<\infty\}$ on $(\Omega,\mathcal{F},\mathbb{P})$, and the solution is a continuous-time Markov chain taking values in $\mathcal{I}^r$ with infinitesimal generator given by Q^r. Then, $\overline{X}^r=\frac{1}{r}X^r$ satisfies:

$$\overline{X}^r(t)=x^r_0+\sum_{j=1}^n\frac{\ell_j}{r}N^r_j\left(r\int_0^t\lambda^r\left(\overline{X}^r(s),\ell_j\right)ds\right),\qquad\text{for all } t\ge0.$$

Define $\nu^r(x):=\sum_{j=1}^n\ell_j\epsilon^r(x,\ell_j)$ for $x\in\mathcal{I}$, and $Y^r_j(t):=N^r_j(t)-t$ for $1\le j\le n$ and

$t \geq 0$. Then, for $t \geq 0$:

$$\begin{aligned}\overline{X}^r(t) &= x_0^r + \sum_{j=1}^n \frac{\ell_j}{r}\left[Y_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right) + r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right] \\ &= x_0^r + \sum_{j=1}^n \frac{\ell_j}{r}\left[Y_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right) - B_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right)\right] \\ &+ \sum_{j=1}^n \frac{\ell_j}{r} B_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right) + \int_0^t \mu(\overline{X}^r(s))ds + \int_0^t \nu^r(\overline{X}^r(s))ds.\end{aligned}$$

On the other hand, note that $Z^r(t) = \chi^r(t) + \frac{1}{\sqrt{r}}\int_0^t \gamma(Z^r(s))dL^r(s)$ for every $t \geq 0$, where

$$\chi^r(t) := x_0^r + \int_0^t \mu(Z^r(s))ds + \sum_{j=1}^n \frac{\ell_j}{r} B_j^r\left(r\int_0^t \beta(Z^r(s),\ell_j)ds\right), \quad t \geq 0. \tag{6.3}$$

Now, fix $T \geq 1$. Unless indicated otherwise, the following inequalities and estimates hold $\mathbb{P}$-a.s. for all $0 \leq t \leq T$,

$$\begin{aligned}|\overline{X}^r(t) - \chi^r(t)| \leq\ & \underbrace{\frac{1}{r}\sum_{j=1}^n |\ell_j|\left|Y_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right) - B_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right)\right|}_{I_1(t)} \\ &+ \underbrace{\frac{1}{r}\sum_{j=1}^n |\ell_j|\left|B_j^r\left(r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds\right) - B_j^r\left(r\int_0^t \beta(Z^r(s),\ell_j)ds\right)\right|}_{I_2(t)} \\ &+ \underbrace{\left|\int_0^t \mu(\overline{X}^r(s)) - \mu(Z^r(s))ds\right|}_{I_3(t)} + \underbrace{\left|\int_0^t \nu^r(\overline{X}^r(s))ds\right|}_{I_4(t)}.\end{aligned} \tag{6.4}$$

In the following, we estimate each element $I_1, \ldots, I_4$ by using a different method. In the following, define

$$M_\beta := \sup_{1\leq j\leq n}\sup_{x\in\mathcal{I}} \beta(x,\ell_j). \tag{6.5}$$

For I_1 we use (6.1). Note first that for $s \in [0,T]$,

$$r\lambda^r(\overline{X}^r(s),\ell_j) = r[\beta(\overline{X}^r(s),\ell_j) + \epsilon^r(\overline{X}^r(s),\ell_j)] \leq r[M_\beta + M_{\mathcal{I}}], \tag{6.6}$$

where $M_{\mathcal{I}}$ is defined in (3.2). Then, for $0 \leq t \leq T$, $r\int_0^t \lambda^r(\overline{X}^r(s),\ell_j)ds \leq r[M_\beta + M_{\mathcal{I}}]T$. By (6.1) and Lemma C.1, choosing $v := [M_\beta + M_{\mathcal{I}}]T$, there exist a constant $C(v)$ and nonnegative random variables $\{\eta_{j,1}^r\}_{j=1}^n$ such that for each $1 \leq j \leq n$:

$$\sup_{0\leq u\leq rv} |Y_j^r(u) - B_j^r(u)| \leq C(v)\log r + \eta_{j,1}^r$$

and

$$\mathbb{P}[\eta^r_{j,1} > x] \leq Kr^{-2}e^{-\lambda x}, \qquad \text{for every } x > 0. \tag{6.7}$$

Define the constants $C_{T,1} := C([M_\beta + M_\mathcal{I}]T)$ and $K_{T,1} := K$. Then,

$$\begin{aligned} I_1(t) &= \frac{1}{r}\sum_{j=1}^{n} |\ell_j| \left| Y^r_j\left(r\int_0^t \lambda^r(\overline{X}^r(s), \ell_j)ds\right) - B^r_j\left(r\int_0^t \lambda^r(\overline{X}^r(s), \ell_j)ds\right)\right| \\ &\leq \frac{1}{r}\sum_{j=1}^{n} |\ell_j| \sup_{0\leq u\leq r(M_\beta+M_\mathcal{I})T} |Y^r_j(u) - B^r_j(u)| \\ &\leq \frac{1}{r}\sum_{j=1}^{n} |\ell_j|(C_{T,1}\log r + \eta^r_{j,1}). \end{aligned} \tag{6.8}$$

For I_2, we use modulus of continuity estimates for Brownian motion. Define for $1 \leq j \leq n$:

$$\Lambda^r_j(z) := \sup_{0\leq u\leq rv} \sup_{0\leq s\leq z\log r} |B^r_j(u+s) - B^r_j(u)|, \qquad 0 \leq z \leq rv, \tag{6.9}$$

$$\overline{\Lambda}^r_j := \sup_{0\leq z\leq rv} (z+1)^{-1/2}\Lambda^r_j(z). \tag{6.10}$$

By the proof of Theorem 11.3.1, in Ethier & Kurtz [9], there exist nonnegative random variables $\{\eta^r_{j,2}\}_{j=1}^n$ and constants $C_{T,2}, K_{T,2} > 0$, depending only on $M_\beta, M_\mathcal{I}$ and T such that for every $1 \leq j \leq n$:

$$\overline{\Lambda}^r_j \leq C_{T,2}\log r + \eta^r_{j,2} \qquad \mathbb{P}\text{- a.s.} \tag{6.11}$$

and

$$\mathbb{P}[\eta^r_{j,2} > x] \leq K_{T,2}r^{-2}\exp\left[-\frac{x^2}{18\log r}\right] \qquad \text{for every } x > 0. \tag{6.12}$$

To derive these estimates from [9], scrutiny of (3.11)-(3.17) in [9] reveals that these equations can apply to any Brownian motion and furthermore $\overline{\beta}_l T$ in [9] can be replaced by our v, and the constants K_l and C_l in [9] can be taken to not depend on l. In this proof, we have dropped a term from the estimate (3.17) in [9] when writing (6.12) since this term would not improve our bounds. Now, for each $1 \leq j \leq n$, for every $0 \leq t \leq T$:

$$\begin{aligned} &\left|B^r_j\left(r\int_0^t \lambda^r(\overline{X}^r(s), \ell_j)ds\right) - B^r_j\left(r\int_0^t \beta(Z^r(s), \ell_j)ds\right)\right| \\ &\leq \Lambda^r_j\left(\frac{r}{\log r}\int_0^t |\lambda^r(\overline{X}^r(s), \ell_j) - \beta(Z^r(s), \ell_j)|ds\right), \end{aligned}$$

where we used (6.6) and the fact that $r\int_0^t \beta(Z^r(s), \ell_j)ds \leq rv$ for $t \in [0, T]$. By (6.10), for every $0 \leq z \leq rv$, we have that $\Lambda^r_j(z) \leq \overline{\Lambda}^r_j(1+z)^{1/2}$. Now, since $r \geq 8$, $\log r \geq 2$ and

$$\frac{r}{\log r}\int_0^t |\lambda^r(\overline{X}^r(s), \ell_j) - \beta(Z^r(s), \ell_j)|ds \leq \frac{r}{\log r}2[M_\beta + M_\mathcal{I}]T \leq r[M_\beta + M_\mathcal{I}]T = rv.$$

Therefore,

$$\begin{aligned}
&\Lambda_j^r\left(\frac{r}{\log r}\int_0^t |\lambda^r(\overline{X}^r(s),\ell_j)-\beta(Z^r(s),\ell_j)|ds\right)\\
&\le \overline{\Lambda}_j^r\left(1+\frac{r}{\log r}\int_0^t |\lambda^r(\overline{X}^r(s),\ell_j)-\beta(Z^r(s),\ell_j)|ds\right)^{1/2}\\
&\le \overline{\Lambda}_j^r\left(1+\frac{r}{\log r}\int_0^t \left(|\beta(\overline{X}^r(s),\ell_j)-\beta(Z^r(s),\ell_j)|+|\epsilon^r(\overline{X}^r(s),\ell_j)|\right)ds\right)^{1/2}\\
&\le \overline{\Lambda}_j^r\left(1+\frac{r}{\log r}\int_0^t A|\overline{X}^r(s)-Z^r(s)|+\frac{M_{\mathcal{I}}}{r}ds\right)^{1/2}\\
&\le \left(C_{T,2}\log r+\eta_{j,2}^r\right)\left(1+\frac{TM_{\mathcal{I}}}{\log r}+\frac{Ar}{\log r}\int_0^t |\overline{X}^r(s)-Z^r(s)|ds\right)^{1/2},
\end{aligned}$$

where in the third inequality we used (3.2), the fact that all $\beta(\cdot,\ell_j)$ are Lipschitz continuous with Lipschitz constant $A>0$, and in the last inequality we used (6.11). Hence,

$$I_2(t)\le \frac{1}{r}\sum_{j=1}^n |\ell_j|\left(C_{T,2}\log r+\eta_{j,2}^r\right)\left(1+\frac{TM_{\mathcal{I}}}{\log r}+\frac{Ar}{\log r}\int_0^t |\overline{X}^r(s)-Z^r(s)|ds\right)^{1/2}. \tag{6.13}$$

We estimate I_3 using the Lipschitz property of μ. The Lipschitz constant for μ will be called A_μ and it can be taken as $A\sum_{j=1}^n|\ell_j|$. Then, for $0\le t\le T$,

$$I_3(t)\le \int_0^t |\mu(\overline{X}^r(s))-\mu(Z^r(s))|ds\le A_\mu\int_0^t |\overline{X}^r(s)-Z^r(s)|ds. \tag{6.14}$$

Finally, for I_4 we have

$$I_4(t)=\left|\int_0^t \nu^r(\overline{X}^r(s))ds\right|\le\int_0^t |\nu^r(\overline{X}^r(s))|ds\le\int_0^t\frac{M_{\mathcal{I}}}{r}\sum_{j=1}^n|\ell_j|ds\le\frac{1}{r}\Big(TM_{\mathcal{I}}\sum_{j=1}^n|\ell_j|\Big). \tag{6.15}$$

For $s,t\in[0,T]$, let $\zeta^r(s):=\frac{r|\overline{X}^r(s)-Z^r(s)|}{\log r}$ and $\overline{\zeta}^r(t)=\sup_{0\le s\le t}\zeta^r(s)=\frac{r}{\log r}\|\overline{X}^r-Z^r\|_t$. Since T is fixed, lets denote $\overline{\zeta}^r:=\overline{\zeta}^r(T)$. Combining (6.4) with (6.8),(6.13),(6.14) and (6.15), we obtain for each $0\le t\le T$,

$$\begin{aligned}
\frac{r|\overline{X}^r(t)-\chi^r(t)|}{\log r}&\le\frac{1}{\log r}\sum_{j=1}^n|\ell_j|(C_{T,1}\log r+\eta_{j,1}^r)\\
&\quad+\frac{1}{\log r}\sum_{j=1}^n|\ell_j|\left(C_{T,2}\log r+\eta_{j,2}^r\right)\left(1+\frac{TM_{\mathcal{I}}}{\log r}+A\int_0^t\zeta^r(s)ds\right)^{1/2}\\
&\quad+A_\mu\int_0^t\zeta^r(s)ds+\frac{1}{\log r}\left(TM_{\mathcal{I}}\sum_{j=1}^n|\ell_j|\right).
\end{aligned}$$

Hence, $\mathbb{P}$-a.s. for each $0 \le t \le T$,

$$\frac{r|\overline{X}^r(t) - \chi^r(t)|}{\log r} \le G^r + A_\mu \int_0^t \overline{\zeta}^r(s)ds, \tag{6.16}$$

where

$$G^r := \sum_{j=1}^n |\ell_j| \left(C_{T,1} + \frac{TM_{\mathcal{I}}}{\log r} + \frac{\eta_{j,1}^r}{\log r} \right) + \sum_{j=1}^n |\ell_j| \left(C_{T,2} + \frac{\eta_{j,2}^r}{\log r} \right) \left(1 + \frac{TM_{\mathcal{I}}}{\log r} + AT\overline{\zeta}^r \right)^{1/2}.$$

Notice that G^r does not depend on t. Since the right side of (6.16) is monotone in t it follows that $\mathbb{P}$-a.s. for all $t \in [0,T]$,

$$\sup_{0\le u\le t} \frac{r|\overline{X}^r(u) - \chi^r(u)|}{\log r} = \frac{r}{\log r}\|\overline{X}^r - \chi^r\|_t \le G^r + A_\mu \int_0^t \overline{\zeta}^r(s)ds. \tag{6.17}$$

Now, consider the Skorokhod Map $\Gamma_{0,a}$ on $[0,a]$ defined in Section B. From the discussion there, we observe that $\mathbb{P}$-a.s. $\Gamma_{0,a}(\chi^r) = Z^r$. Also, we note that since $\overline{X}^r(t) \in \mathcal{I}$ for every $t \ge 0$, then $\Gamma_{0,a}(\overline{X}^r) = \overline{X}^r$. By the Lipschitz continuity of the Skorokhod Map (B.4), $\mathbb{P}$-a.s. for all $0 \le t \le T$:

$$\|\overline{X}^r - Z^r\|_t = \|\Gamma_{0,a}(\overline{X}^r) - \Gamma_{0,a}(\chi^r)\|_t \le 4\|\overline{X}^r - \chi^r\|_t.$$

Consequently, $\mathbb{P}$-a.s. for every $0 \le t \le T$,

$$\overline{\zeta}^r(t) = \frac{r}{\log r}\|\overline{X}^r - Z^r\|_t \le \frac{4r}{\log r}\|\overline{X}^r - \chi^r\|_t \le 4G^r + 4A_\mu \int_0^t \overline{\zeta}^r(s)ds.$$

Then, by Gronwall's inequality, $\overline{\zeta}^r(t) \le e^{4A_\mu t} 4G^r$ for all $t \in [0,T]$. In particular, $\mathbb{P}$-a.s.:

$$\begin{aligned}\overline{\zeta}^r &\le e^{4A_\mu T} 4E^r = 4e^{4A_\mu T} \sum_{j=1}^n |\ell_j| \left(C_{T,1} + \frac{TM_{\mathcal{I}}}{\log r} + \frac{\eta_{j,1}^r}{\log r} \right) \\ &+ 4e^{4A_\mu T}(AT)^{1/2} \left(\frac{1}{AT} + \frac{M_{\mathcal{I}}}{A\log r} + \overline{\zeta}^r \right)^{1/2} \sum_{j=1}^n |\ell_j| \left(C_{T,2} + \frac{\eta_{j,2}^r}{\log r} \right).\end{aligned}$$

This is an inequality of the type: $\gamma \le a + (c+\gamma)^{1/2} b$ which implies $\gamma \le c + 2a + b^2$. Thus, $\mathbb{P}$-a.s.,

$$\begin{aligned}\overline{\zeta}^r &\le \frac{1}{AT} + \frac{M_{\mathcal{I}}}{A\log r} + 8e^{4A_\mu T} \sum_{j=1}^n |\ell_j| \left(C_{T,1} + \frac{TM_{\mathcal{I}}}{\log r} + \frac{\eta_{j,1}^r}{\log r} \right) \\ &+ 16e^{8A_\mu T} AT \left(\sum_{j=1}^n |\ell_j| \left(C_{T,2} + \frac{\eta_{j,2}^r}{\log r} \right) \right)^2.\end{aligned}$$

Define $\|\ell\|_\infty = \sup_{1\le j\le n} |\ell_j|$. Since $\log r \ge 2$, we have $\mathbb{P}$-a.s.,

$$\begin{aligned}
\overline{\zeta}^r &\le \frac{2+M_{\mathcal{I}}T}{2AT} + 8e^{4A_\mu T}\sum_{j=1}^n |\ell_j|\left(C_{T,1} + \frac{TM_{\mathcal{I}}}{2}\right) + 8e^{4A_\mu T}\sum_{j=1}^n |\ell_j|\frac{\eta^r_{j,1}}{\log r} \\
&\quad + 16e^{8A_\mu T}AT\left(\sum_{j=1}^n |\ell_j| C_{T,2} + \sum_{j=1}^n |\ell_j|\frac{\eta^r_{j,2}}{\log r}\right)^2 \\
&\le \frac{2+M_{\mathcal{I}}T}{2AT} + 8e^{4A_\mu T} n\|\ell\|_\infty \left(C_{T,1} + \frac{TM_{\mathcal{I}}}{2}\right) + \frac{8e^{4A_\mu T}}{\log r}\sum_{j=1}^n |\ell_j|\eta^r_{j,1} \\
&\quad + 32e^{8A_\mu T}AT\left[\left(\sum_{j=1}^n |\ell_j| C_{T,2}\right)^2 + \left(\sum_{j=1}^n |\ell_j|\frac{\eta^r_{j,2}}{\log r}\right)^2\right] \\
&\le \underbrace{\frac{2+M_{\mathcal{I}}T}{2AT} + 8e^{4A_\mu T} n\|\ell\|_\infty \left(C_{T,1} + \frac{TM_{\mathcal{I}}}{2}\right) + 32e^{8A_\mu T}ATC^2_{T,2}n^2\|\ell\|^2_\infty}_{C_T} \\
&\quad + \underbrace{\frac{8e^{4A_\mu T}\|\ell\|_\infty}{\log r}\sum_{j=1}^n \eta^r_{j,1} + \frac{32e^{8A_\mu T}AT\|\ell\|^2_\infty}{(\log r)^2}\left(\sum_{j=1}^n \eta^r_{j,2}\right)^2}_{\eta^r_T} \\
&= C_T + \eta^r_T =: \Theta^r_T.
\end{aligned}$$

Thus, we have proved that for every $T \ge 1$, (3.3) holds.

Now, we establish the tail bound (3.4). First, notice that C_T has already been defined and depends only on $T, \mathcal{I}, \mathbf{L}, M_{\mathcal{I}}$, and β. We will prove the following and then (3.4) will follow.

For every $T \ge 1$ there exist constants $\lambda_T, K_T > 0$, depending only on $T, \mathcal{I}, \mathbf{L}, M_{\mathcal{I}}$, and β, such that

$$\mathbb{P}[\eta^r_T > x] \le K_T r^{-2}\exp\left(-\lambda_T x \log r\right) \qquad \text{for every } x > 0. \tag{6.18}$$

In order to prove this, first define $a_{T,1} := 8e^{4A_\mu T}\|\ell\|_\infty$ and $a_{T,2} := 32e^{8A_\mu T}AT\|\ell\|^2_\infty$ which are constants depending only on $T, \mathcal{I}, \mathbf{L}, M_{\mathcal{I}}$, and β. Then, for an arbitrary

$x > 0$:

$$\begin{aligned}\mathbb{P}[\eta_T^r > x] &= \mathbb{P}\left(\frac{a_{T,1}}{\log r}\sum_{j=1}^n \eta_{j,1}^r + \frac{a_{T,2}}{(\log r)^2}\left(\sum_{j=1}^n \eta_{j,2}^r\right)^2 > x\right)\\ &\le \mathbb{P}\left(\frac{a_{T,1}}{\log r}\sum_{j=1}^n \eta_{j,1}^r > \frac{x}{2}\right) + \mathbb{P}\left(\frac{a_{T,2}}{(\log r)^2}\left(\sum_{j=1}^n \eta_{j,2}^r\right)^2 > \frac{x}{2}\right)\\ &= \underbrace{\mathbb{P}\left(\sum_{j=1}^n \eta_{j,1}^r > \frac{x\log r}{2a_{T,1}}\right)}_{J_1} + \underbrace{\mathbb{P}\left(\sum_{j=1}^n \eta_{j,2}^r > \frac{\sqrt{x}\log r}{\sqrt{2a_{T,2}}}\right)}_{J_2}.\end{aligned}$$

For J_1, equation (6.7) yields:

$$\begin{aligned}J_1 &\le \sum_{j=1}^n \mathbb{P}\left(\eta_{j,1}^r > \frac{x\log r}{2a_{T,1}n}\right) \le \sum_{j=1}^n K_{T,1} r^{-2}\exp\left(-\lambda\left(\frac{x\log r}{2a_{T,1}n}\right)\right)\\ &= nK_{T,1}r^{-2}\exp\left(-\lambda_{T,1}x\log r\right),\end{aligned}$$

where $\lambda_{T,1} = \frac{\lambda}{2a_{T,1}n} > 0$. Similarly for J_2, equation (6.12) yields:

$$\begin{aligned}J_2 &\le \sum_{j=1}^n \mathbb{P}\left(\eta_{j,2}^r > \frac{\sqrt{x}\log r}{\sqrt{2a_{T,2}n}}\right) \le \sum_{j=1}^n K_{T,2} r^{-2}\exp\left(-\frac{1}{18\log r}\left(\frac{\sqrt{x}\log r}{\sqrt{2a_{T,2}n}}\right)^2\right)\\ &= nK_{T,2}r^{-2}\exp\left(-\lambda_{T,2}x\log r\right),\end{aligned}$$

where $\lambda_{T,2} = (36n^2 a_{T,2})^{-1} > 0$. Putting these two bounds together, we obtain

$$\mathbb{P}[\eta_T^r > x] \le K_T r^{-2}\exp\left(-\lambda_T x\log r\right),$$

where $K_T = 2n(K_{T,1} \vee K_{T,2}) > 0$ and $\lambda_T = \lambda_{T,1} \wedge \lambda_{T,2} > 0$. This proves the claim.

□

6.3. Proof of Corollary 3.1

For $r \ge 8$ in $\mathcal{R}$, let $\overline{X}^r$ and Z^r be the processes described in Theorem 3.2. For each $T \ge 1$, the law of the pair $(\{\overline{X}^r(t),\ 0 \le t \le T\}, \{Z^r(t),\ 0 \le t \le T\})$, denoted by π_T^r, is a coupling of the laws P_T^r and $\tilde{P}_T^r$, i.e., π_T^r is a probability measure on $(\mathcal{D}[0,T] \times \mathcal{D}[0,T], \mathcal{M}_T \otimes \mathcal{M}_T)$ such that the marginals of π_T^r are given by P_T^r and $\tilde{P}_T^r$ respectively. In the notation of Section 1.3, $\pi_T^r \in \Pi(P_T^r, \tilde{P}_T^r)$.

Let d_0 be the metric defined by (see (12.16) of [6]),

$$d_0(x,y) := \inf_{\lambda\in\Lambda}\{\|\lambda\|^\circ \vee \|x - y\circ\lambda\|_T\} \qquad x, y \in \mathcal{D}[0,T], \tag{6.19}$$

where Λ is the set of continuous, onto and strictly increasing mappings $\lambda : [0,T] \longrightarrow [0,T]$ and $\|\lambda\|^\circ = \sup_{0\le s<t\le T} |\log\frac{\lambda(t)-\lambda(s)}{t-s}|$ for $\lambda \in \Lambda$. Under this metric, $(\mathcal{D}[0,T], d_0)$

is a complete and separable metric space which induces the J_1-topology on $\mathcal{D}[0,T]$. Observe furthermore that $d_0(x,y) \le \|x-y\|_T$ for any $x, y \in \mathcal{D}[0,T]$. For $p \ge 1$:

$$\begin{aligned}\mathcal{W}_p(P_T^r, \tilde{P}_T^r) &= \inf_{\pi \in \Pi(P_T^r, \tilde{P}_T^r)} \left(\int_{\mathcal{D}[0,T]\times\mathcal{D}[0,T]} d_0^p(x,y) d\pi(x,y) \right)^{1/p} \\ &\le \left(\int_{\mathcal{D}[0,T]\times\mathcal{D}[0,T]} d_0^p(x,y) d\pi_T^r(x,y) \right)^{1/p} = \left(\mathbb{E}\left[d_0^p\left(\overline{X}^r, Z^r \right) \right] \right)^{1/p} \\ &\le \left(\mathbb{E}\left[\|\overline{X}^r - Z^r\|_T^p \right] \right)^{1/p} \le (\mathbb{E}[(\Theta_T^r)^p])^{1/p} \left(\frac{\log r}{r} \right),\end{aligned}$$

where we used (3.3) in the last inequality. In the proof of Theorem 3.2, we have shown that $\Theta_T^r = \eta_T^r + C_T$ and η_T^r is a nonnegative random variable such that $\mathbb{P}[\eta_T^r > x] \le K_T r^{-2} \exp(-\lambda_T x \log r)$ for every $x > 0$. Since $(\Theta_T^r)^p \le 2^{p-1}((\eta_T^r)^p + C_T^p)$, to bound $\mathbb{E}[(\Theta_T^r)^p]$, we compute:

$$\begin{aligned}\mathbb{E}[(\eta_T^r)^p] &= \int_0^\infty p x^{p-1} \mathbb{P}[\eta_T^r > x] dx \le \frac{pK_T}{r^2} \int_0^\infty x^{p-1} e^{-\lambda_T x \log r} dx \\ &= \frac{pK_T}{r^2 (\lambda_T \log r)^p} \int_0^\infty u^{p-1} e^{-u} du \le \frac{p\Gamma(p) K_T}{\lambda_T^p},\end{aligned}$$

where we used the fact that $r^2 (\log r)^p \ge 1$. Combining the above, we obtain that (3.5) holds with $C := 2^{1-1/p} \left(\frac{p\Gamma(p)K_T}{\lambda_T^p} + C_T^p \right)^{1/p}$.

□

6.4. Proof of Theorem 3.4

The proof follows by similar arguments as for Theorem 3.2. In particular, the construction of the processes is essentially the same and takes into account that the processes do not explode (see Theorem 5.2 for the case $\mathcal{I} = [0,\infty)$). To prove the bounds, we follow the proof of Theorem 3.2 but with $t \wedge \tau_{\mathcal{K}}^r$ in place of t, and where the Skorokhod map Γ_0 is used in place of $\Gamma_{0,a}$.

□

Remark 6.1. One can show that $\frac{1}{\sqrt{r}} \int_0^{t\wedge\tau_{\mathcal{K}}^r} \gamma(Z^r(s)) dL^r(s) \le \Theta_{T,\mathcal{K}}^r \frac{\log r}{r}$, i.e., the "reflection term" is of the same order as the error term. However, despite the small size of this "reflection term", it can have a substantial effect on the nonlinear dynamics.

Appendix A. Komlós-Major-Tusnády type approximation

Theorem A.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which are defined:*

(i) a standard n-dimensional Brownian motion $B = (B_1, \dots, B_n)$,

(ii) a family of n double sequences $U^1 = (U_{i,j}^1)_{i,j\ge1}, \dots, U^n = (U_{i,j}^n)_{i,j\ge1}$ of i.i.d. random variables with common uniform distribution on $(0,1)$, where all of the random variables are mutually independent and independent of B.

Then, there exist n unit rate Poisson processes $N_1, \ldots, N_n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, mutually independent and such that:

$$\mathbb{P}\left[\sup_{0\le t\le T} |N_\ell(t) - t - B_\ell(t)| > (C+\theta)\log T + x\right] \le KT^{-\theta\lambda}e^{-\lambda x} \tag{A.1}$$

for every $1 \le \ell \le n$, $T > 1$, $x > 0$ and $\theta > 0$. Here, $C, K, \lambda > 0$ are universal constants.

The proof of Theorem A.1 follows ideas of Komlós, Major & Tusnády [11, 12] and Ethier & Kurtz [9]. More precisely, Theorem A.1 is a variation of Corollary 7.5.3 in Ethier & Kurtz [9]. There, the authors showed that for a (one-dimensional) Lévy process X, with finite exponential moments for $X(1)$, one can construct on the same probability space a realization of X and a process $\{at + bB(t),\ 0 \le t < \infty\}$, where $a, b \in \mathbb{R}$ and B is a Brownian motion, such that

$$\mathbb{P}\left[\sup_{0\le t\le T} |X(t) - at - bB(t)| > (C'+\theta)\log T + x\right] \le K'T^{-\theta\lambda'}e^{-\lambda' x}$$

for all $T > 1$, $x > 0$ and $\theta > 0$, with constants $C', K', \lambda' > 0$. The results in Theorem A.1 are for the case where X is a unit rate centered Poisson process. Our result is different from the one of Ethier & Kurtz in the sense that, instead of simply showing that a coupling exists, we explicitly construct it on a space where a prescribed Brownian motion already lives. For this construction, only extra independent uniform random variables are needed besides the given Brownian motion. The ability to construct the processes $N_1, \ldots, N_n$ on any such probability space will be important for our proofs of Theorems 3.2 and 3.4. More precisely, it will let us address step (iv) in Section 6.1.

In order to prove Theorem A.1, we start by recalling the results of Komlós, Major & Tusnády [11, 12]. Roughly speaking, these authors showed that if F is a distribution function on the real line with finite exponential moments, i.e., $\int e^{\alpha u} dF(u) < \infty$ for $|\alpha| < \alpha_0$ for some $\alpha_0 > 0$, it is possible to couple an i.i.d. sequence of random variables $(\mathcal{X}_i)_{i\ge1}$ with distribution F to an i.i.d. sequence $(\mathcal{Y}_i)_{i\ge1}$ of standard normals in such a way that $|S(m) - T(m)| = O(\log m)$, where $S(m) = \sum_{i=1}^m \mathcal{X}_i$ and $T(m) = \sum_{i=1}^m \mathcal{Y}_i$. With this idea we prove the following.

Lemma A.1. *Let $B = \{B(t),\ 0 \le t < \infty\}$ be a standard one-dimensional Brownian motion defined on a space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, there exist universal constants $C, K, \lambda > 0$ and an i.i.d. sequence $(\xi_i)_{i\ge1}$ of Poisson random variables of mean 1, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, such that:*

$$\mathbb{P}\left[\sup_{1\le k\le m} |S(k) - k - B(k)| > C\log m + x\right] \le Ke^{-\lambda x} \tag{A.2}$$

for every $m \ge 1$ and $x > 0$, where $S(k) = \sum_{i=1}^k \xi_i$.

Proof. Define, $\mathcal{Y}_i := B(i) - B(i-1)$ for $i \ge 1$. Then, $(\mathcal{Y}_i)_{i\ge1}$ is a sequence of i.i.d. standard normal random variables. Consider the distribution function F of a random variable $Z - 1$, where Z is Poisson distributed with mean 1. Then, F defines a distribution with mean 0, variance 1 and finite exponential moments. Moreover, F is a lattice distribution (as defined in [11]). By Theorem 1 in [11] (the KMT

approximation), there exist constants $C, K, \lambda > 0$ depending only on F, and functions $\{f_i(y_1, \ldots, y_{2i})\}_{i\geq 1}$ such that

$$\mathcal{X}_i := f_i(\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_{2i}), \qquad i \geq 1$$

defines an i.i.d. sequence of random variables with distribution function F, for which

$$\mathbb{P}\left[\sup_{1\leq k\leq m} |\tilde{S}(k) - \tilde{T}(k)| > C\log m + x\right] \leq Ke^{-\lambda x} \tag{A.3}$$

holds for every $m \geq 1$ and $x > 0$, where $\tilde{S}(k) = \sum_{i=1}^{k} \mathcal{X}_i$ and $\tilde{T}(k) = \sum_{i=1}^{k} \mathcal{Y}_i$, $1 \leq k \leq m$. Note that $\tilde{T}(k) = \sum_{i=1}^{k}(B(i) - B(i-1)) = B(k) - B(0) = B(k)$, $\mathbb{P}-$a.s. Additionally, if we define the variables $\xi_i := \mathcal{X}_i + 1$ for $i \geq 1$, then $(\xi_i)_{i\geq 1}$ will be an i.i.d. sequence of Poisson random variables with mean 1. Define $S(k) := \sum_{i=1}^{k} \xi_i$ for $1 \leq k \leq m$. Then, $\mathbb{P}$-a.s., $S(k) - k - B(k) = \tilde{S}(k) - \tilde{T}(k)$, and (A.2) follows from (A.3). □

Remark A.1. The sequence $\xi = (\xi_i)_{i\geq 1}$ is such that $\xi(\omega) = \Psi(B(\omega, \cdot))$ for every $\omega \in \Omega$, where Ψ is a deterministic mapping. To be precise, $\Psi = \Psi_2 \circ \Psi_1$ where $\Psi_1 : \mathcal{C} \longrightarrow \mathbb{R}^{\mathbb{N}_+}$ is given by $\Psi_1(w) = \{w(i) - w(i-1)\}_{i\geq 1}$ and $\Psi_2 : \mathbb{R}^{\mathbb{N}_+} \longrightarrow \mathbb{R}^{\mathbb{N}_+}$ is given by $\Psi_2(\{y_i\}_{i\geq 1}) = \{f_i(y_1, \ldots, y_{2i}) + 1\}_{i\geq 1}$ where $(f_i)_{i\geq 1}$ are the functions described in Theorem 1 in [11].

Proof of Theorem A.1. Lemma A.1 allow us to construct, on $(\Omega, \mathcal{F}, \mathbb{P})$, n i.i.d. sequences $(\xi_i^\ell)_{i\geq 1}$, for $1 \leq \ell \leq n$, with common Poisson distribution of mean 1 in such a way that, for each $1 \leq \ell \leq n$, equation (A.2) holds with B_ℓ, $(\xi_i^\ell)_{i\geq 1}$ and $S_\ell(k) = \sum_{i=1}^{k} \xi_i^\ell$ in place of B, $(\xi_i)_{i\geq 1}$ and $S(k)$, respectively, with universal constants $C', K', \lambda' > 0$. From Remark A.1, $(\xi_i^\ell)_{i\geq 1}$ is a function of B_ℓ for each ℓ. For $1 \leq \ell \leq n$ define the process $N_\ell = \{N_\ell(t),\ 0 \leq t < \infty\}$ by $N_\ell(0) := 0$ and

$$N_\ell(i - 1 + t) := S_\ell(i-1) + \sum_{j=1}^{\xi_i^\ell} \mathbb{1}_{\{U_{i,j}^\ell \leq t\}} \tag{A.4}$$

where $i \geq 1$ and $0 < t \leq 1$. Thus, ξ_i^ℓ gives the number of jumps of N_ℓ in $(i-1, i]$ and the jump points are given by ξ_i^ℓ i.i.d. uniform random variables on $(i-1, i]$. For each $1 \leq \ell \leq n$, N_ℓ is a unit rate Poisson process such that $N_\ell(k) = S_\ell(k)$ for each $k \geq 1$ $\mathbb{P}$-a.s.

The estimate (A.1) now follows as in the proof of Corollary 7.5.3 of Ethier & Kurtz [9] by considering $\{N_\ell(t) - t,\ 0 \leq t < \infty\}$ as the process $X = \{X(t),\ 0 \leq t < \infty\}$ there, for each $1 \leq \ell \leq n$. The constants $C, K, \lambda > 0$ in (A.1) are given by $C = 3C'$, $\lambda = \min\{\frac{\lambda'}{3}, \frac{1}{3C'}\}$ and $K = K' + 2\exp(\psi(1/C') + \psi(-1/C')) + 4\exp(1/2(C')^2)$ where $\psi(x) = \exp(e^x - 1 - x)$, $x \geq 0$. Therefore, the constants $C, K, \lambda > 0$ are universal.

Notice that each N_ℓ is constructed from (B_ℓ, U^ℓ). More concretely, there exists a measurable function $\Phi : \mathcal{C} \times \mathbb{R}^{\mathbb{N}_+} \times \mathbb{R}^{\mathbb{N}_+} \longrightarrow \mathcal{D}$ such that $N_\ell(\omega, \cdot) = \Phi(B_\ell(\omega, \cdot), U^\ell(\omega))$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Since $(B_1, U^1), \ldots, (B_n, U^n)$ are independent, we get that the processes $N_1, \ldots, N_n$ are independent. □

Remark A.2. By examining the proof of Theorem A.1, we observe that the Poisson processes $N_1, \ldots, N_n$ are not necessarily adapted to $\{\mathcal{F}_t^B \vee \sigma(U^1, \ldots, U^n)\}$ where $\{\mathcal{F}_t^B\}$ is the filtration generated by the Brownian motion B.

Appendix B. The Skorokhod problem

In this section we introduce the so called *Skorokhod Problem* and its associated *Skorokhod Map* in one-dimension. We are particularly interested in the Lipschitz continuity of the Skorokhod Map (see (B.3) and (B.4)) which is key for the proofs of Theorems 3.2 and 3.4. This property is a special feature of our one-dimensional scenario, since Lipschitz continuity is not known in higher dimensions for the smoothly varying reflection field on $\partial\mathbb{R}^d_+$ associated with the CLA [17].

Let $\mathcal{I} = [0,\infty)$ or $[0,a]$ and γ be as in (2.7). Recall $\mathcal{D}$ is the set of functions $x : [0,\infty) \longrightarrow \mathbb{R}$ which are right-continuous on $[0,\infty)$, with finite left-hand limits on $(0,\infty)$. Define $\mathcal{D}_\mathcal{I}$ as the family of $x \in \mathcal{D}$ with $x(0) \in \mathcal{I}$ and $\mathcal{C}_\mathcal{I}$ as the set of continuous functions in $\mathcal{D}_\mathcal{I}$.

Let $x \in \mathcal{D}_\mathcal{I}$, a pair $(z,y) \in \mathcal{D}_\mathcal{I} \times \mathcal{D}$ will be called a solution of the **Skorokhod problem on** $\mathcal{I}$ for x if the following holds:

(i) $z(t) = x(t) + \int_0^t \gamma(z(s))dy(s)$, for every $t \geq 0$,

(ii) $z(t) \in \mathcal{I}$, for every $t \geq 0$,

(iii) y is a non-decreasing function such that $y(0) = 0$ and

$$\int_{[0,\infty)} \mathbb{1}_{\{z(s)\notin\partial\mathcal{I}\}} dy(s) = 0.$$

In 1961, Skorokhod [21] introduced this problem for $\mathcal{I} = [0,\infty)$ and showed that for each $x \in \mathcal{D}_\mathcal{I}$, there exists a unique solution $(z,y) \in \mathcal{D}_\mathcal{I} \times \mathcal{D}$ to the Skorokhod problem for x given by:

$$y(t) = \sup_{0\leq s\leq t} x^-(s) \quad \text{and} \quad z(t) = x(t) + y(t), \quad t \geq 0. \tag{B.1}$$

Here $x^-(t) = \max\{-x(t), 0\}$ is the negative part of $x(t)$. Define the **Skorokhod map on** $[0,\infty)$ as the function $\Gamma_0 : \mathcal{D}_\mathcal{I} \longrightarrow \mathcal{D}_\mathcal{I}$ given by $x \mapsto \Gamma_0(x) = z$ where (z,y) is the solution of the Skorokhod problem for x. From (B.1) one can deduce that Γ_0 maps $\mathcal{C}_\mathcal{I}$ into $\mathcal{C}_\mathcal{I}$ and for every $x, \tilde{x} \in \mathcal{D}_\mathcal{I}$ and $T > 0$,

$$\|\Gamma_0(x) - x - (\Gamma_0(\tilde{x}) - \tilde{x})\|_T \leq \|x - \tilde{x}\|_T, \tag{B.2}$$

and

$$\|\Gamma_0(x) - \Gamma_0(\tilde{x})\|_T \leq 2\|x - \tilde{x}\|_T. \tag{B.3}$$

For $\mathcal{I} = [0,a]$ and $x \in \mathcal{D}_{[0,a]}$, there exists a unique solution $(z,y) \in \mathcal{D}_\mathcal{I} \times \mathcal{D}$ to the Skorokhod Problem on $[0,a]$ for x. This follows from Anulova & Liptser [5]. Define the **Skorokhod map on** $[0,a]$ as the function $\Gamma_{0,a} : \mathcal{D}_{[0,a]} \longrightarrow \mathcal{D}_{[0,a]}$ given by $x \mapsto \Gamma_{0,a}(x) = z$ where (z,y) is the solution of the Skorokhod problem on $[0,a]$ for x.

Kruk, Lehoczky, Ramanan & Shreve [13], gave the following explicit formula for $\Gamma_{0,a}$: $\Gamma_{0,a} = \Lambda_a \circ \Gamma_0$ where $\Lambda_a(x)(t) = x(t) - \sup_{s\in[0,t]} \left[(x(s) - a)^+ \wedge \inf_{u\in[s,t]} x(u)\right]$ for $x \in \mathcal{D}_\mathcal{I}, t \geq 0$. Using this formula, these authors showed that the Skorokhod map on $[0,a]$ maps $\mathcal{C}_\mathcal{I}$ into $\mathcal{C}_\mathcal{I}$ and for $x, \tilde{x} \in \mathcal{D}_{[0,a]}$ and $T > 0$,

$$\|\Gamma_{0,a}(x) - \Gamma_{0,a}(\tilde{x})\|_T \leq 4\|x - \tilde{x}\|_T. \tag{B.4}$$

We can relate the Skorokhod map to SDERs. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, Z, W, L)$ be a solution of equation (5.1) for $\mathcal{I} = [0, \infty)$ or $\mathcal{I} = [0, a]$. Define

$$\chi(t) := Z(0) + \int_0^t b(Z(s))ds + \int_0^t \rho(Z(s))dW(s), \qquad t \geq 0. \tag{B.5}$$

Then $\mathbb{P}$-a.s., $(Z(\cdot), \alpha L(\cdot))$ is the solution of the Skorokhod problem for $\chi(\cdot)$ on $\mathcal{I}$, and $\Gamma_0(\chi(\cdot)) = Z(\cdot)$ if $\mathcal{I} = [0, \infty)$, or $\Gamma_{0,a}(\chi(\cdot)) = Z(\cdot)$ if $\mathcal{I} = [0, a]$.

Appendix C. A technical lemma

Lemma C.1. *Let Y be a centered unit rate Poisson process and W be a standard one-dimensional Brownian motion such that*

$$\mathbb{P}\left[\sup_{0\leq s\leq t} |Y(s) - W(s)| > (C + \theta)\log t + x\right] \leq Kt^{-\theta\lambda}e^{-\lambda x} \tag{C.1}$$

for every $t > 1$, $x > 0$, $\theta > 0$, where $C, K, \lambda > 0$ are constants. Then, for each $r \geq 3$ and $v > 0$, there is a nonnegative random variable η_v^r and a constant $C(v) > 0$, not depending on r, such that:

$$\sup_{0\leq s\leq rv} |Y(s) - W(s)| \leq C(v)\log r + \eta_v^r$$

and

$$\mathbb{P}[\eta_v^r > x] \leq Kr^{-2}e^{-\lambda x} \qquad \text{for every } x > 0. \tag{C.2}$$

Lemma C.1 is a slightly improved version of a result pointed out by Ethier & Kurtz [9] in their proof of Theorem 11.3.1 (see (3.9) and (3.10)). Here, we explicitly link the parameters λ and K in (C.2) with the same parameters in (C.1). In particular, λ does not depend on v.

Proof. Let $x > 0$ and $\theta > 0$ be arbitrary. Using (C.1) for $t = r(v \vee 1) \geq 3$:

$$\begin{aligned}
&\mathbb{P}\left[\sup_{0\leq s\leq rv} |Y(s) - W(s)| > (C + \theta)\log(r(v \vee 1)) + x\right] \\
&\leq \mathbb{P}\left[\sup_{0\leq s\leq r(v\vee 1)} |Y(s) - W(s)| > (C + \theta)\log(r(v \vee 1)) + x\right] \leq K(r(v \vee 1))^{-\theta\lambda}e^{-\lambda x}.
\end{aligned}$$

Let $\theta = \frac{2}{\lambda}$. Then, the RHS becomes $K(r(v \vee 1))^{-2}e^{-\lambda x} \leq Kr^{-2}e^{-\lambda x}$. Then, for

$$\eta_v^r := \max\left\{\sup_{0\leq s\leq rv} |Y(s) - W(s)| - \left(C + \frac{2}{\lambda}\right)\log(r(v \vee 1)), 0\right\},$$

we have $\mathbb{P}[\eta_v^r > x] \leq Kr^{-2}e^{-\lambda x}$ for every $x > 0$. Finally, since $\log r \geq 1$ and $\log(v\vee 1) \geq 0$, we have

$$\begin{aligned}
\sup_{0\leq s\leq rv} |Y(s) - W(s)| &\leq \eta_v^r + \left(C + \frac{2}{\lambda}\right)\log(r(v \vee 1)) \\
&= \eta_v^r + \left(C + \frac{2}{\lambda}\right)[\log r + \log(v \vee 1)] \leq \eta_v^r + C(v)\log r
\end{aligned}$$

where $C(v) := \left(C + \frac{2}{\lambda}\right)[1 + \log(v \vee 1)]$. □

Acknowledgements

We are grateful to Saul Leite for sharing a preliminary version of the code used for the simulations provided in the Examples section. We thank Yi Fu for assistance in adapting code to produce these simulation results. We also thank an anonymous referee for several helpful suggestions concerning our examples.

Funding information

Research supported in part by NSF Grants DMS-1712974 and DMS-2153866, and the Charles Lee Powell Foundation.

Competing interests

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

References

[1] Agarwal, R.P. and Lakshmikantham, V. (1993). *Uniqueness And Nonuniqueness Criteria For Ordinary Differential Equations.* World Scientific Publishing Company.

[2] Anderson, D., Higham, D., Leite, S. and Williams, R.J. (2019). Constrained Langevin equations and (bio)chemical reaction networks. *SIAM Multiscale Modeling & Simulation*, **Vol. 17,** No. 1, 1-30.

[3] Anderson, D. and Kurtz, T. (2015). *Stochastic Analysis of Biochemical Systems.* Springer International Publishing.

[4] Angius, A., Balbo, G., Beccuti, M., Bibbona, E., Horvath, A. and Sirovich, R. (2015). Approximate analysis of biological systems by hybrid switching jump diffusion. *Theoretical Computer Science.* **587,** 49-72.

[5] Anulova, S.V., Liptser, B.Sh. (1990). Diffusional approximation for processes with the normal reflection. *Theory Probab. Appl.* **Vol 35, No. 3,** 411-423.

[6] Billingsley, P. (1999). *Convergence of Probability Measures.* 2nd Edition. Wiley Series in Probability and Statistics.

[7] Bossy, M., Gobet, E. and Talay, D. (2004). A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *Journal of Applied Probability*, **41(3),** 877-889.

[8] Chung, K.L. and Williams, R.J. (1990) *Introduction to Stochastic Integration.* Second Edition. Birkhäuser Boston.

[9] Ethier, S. and Kurtz, T. (1986). *Markov Processes: Characterization and Convergence.* John Wiley, New York.

[10] Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

[11] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **Vol. 32, No. 1**, 111–131.

[12] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **Vol. 34, No. 1**, 33–58.

[13] Kruk, L., Lehoczky, J., Ramanan, K. and Shreve, S. (2007). An explicit formula for the Skorokhod map on $[0, A]$. *Ann. Probab.* 35 **no. 5,** 1740–1768.

[14] Kurtz, T. (1970). Solutions of Ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, **7(1),** 49-58.

[15] Kurtz, T. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, **8(2),** 344-356.

[16] Kurtz, T. (1976) Limit theorems and diffusion approximations for density dependent Markov chains. *In: Wets R.J.B. (eds) Stochastic Systems: Modeling, Identification and Optimization, I. Mathematical Programming Studies*, **vol 5**. Springer, Berlin, Heidelberg. 67-78.

[17] Leite, S. and Williams, R.J. (2019) A constrained Langevin approximation for chemical reaction networks. *Ann. Appl. Probab.* **29, no. 3,** 1541–1608.

[18] Linetsky, V. (2005). On the transition densities for reflected diffusions. *Advances in Applied Probability*, **37(2),** 435-460.

[19] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion.* 3rd Edition. Springer, Berlin, Heidelberg.

[20] Slominski, L. (1994). On approximation of solutions of multidimensional SDE's with reflecting boundary conditions. *Stochastic Processes and their Applications* Elsevier, **vol. 50(2),** 197-219.

[21] Skorokhod, A. V. (1961). Stochastic Equations for Diffusion Processes in a Bounded Region. *Theory of Probability and its Applications*, Philadelphia **Vol. 6, Iss. 3,** :11.

[22] Tanaka, H. (1979). Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.* **9, no. 1,** 163–177.

[23] Yamada, T. (1976). On the uniqueness of solutions of stochastic differential equations with reflecting barrier conditions. *Séminaire de probabilités de Strasbourg.* **10,** pp. 240-244.

[24] Yamada, T., and Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, **11, no. 1,** 155-167.