Theoretical analysis of numerical schemes for stochastic differential equations

Luis Mario Chaparro Jáquez

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of Leeds
Department of Statistics

March 2025

Intellectual Property Statement

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Acknowledgements

I would like to express my gratitude to my supervisors Elena Issoglio and Jan Palczewski who, with immense patience, guided me through the world of mathematics and academia. I also thank CONACYT for their financial support, which enabled me to embark in this project.

Special thanks to my parents, Nicté and Francisco, who have supported me in all my questionable decisions. To my siblings, José Alonso and Ana Victoria, whose growth I have missed. To my friends in Mexico—Judith, David, Giselle, Chuma, Jordy, Tony—who are always there in one way or another, and to those in England—Anna, Rukia, Nils, Jacob—who have provided me with company in the last few years. And last but not least, to Joyce, for keeping me sane and happy.

Abstract

The present thesis deals with the design of an Euler-Maruyama method for stochastic differential equations (SDEs) with distributional drifts—generalized functions—and study its convergence rate. We obtain a lower bound for the convergence rate for SDEs, and for SDEs of McKean-Vlasov type using a transformation based on the mild solution of partial differential equations (PDEs). One novelty is the usage of the solution to the associated Fokker-Planck PDE in the solution of the McKean equation.

“We demand rigidly defined areas of doubt and uncertainty!”
—Vroomfondel in Douglas Adams’ *The Hitchhikers’ Guide to the Galaxy*

Contents

1 Introduction **1**

2 Literature review **5**

3 Setting of the problems and preliminary mathematics **15**

3.1 Function and distribution spaces . . . 15

3.2 The heat kernel and semigroup . . . 18

3.3 Some related PDEs . . . 19

3.4 Notions of solutions of the SDE . . . 22

3.5 Link between the martingale problem and virtual solutions . . . 27

4 SDEs with irregular drift **29**

4.1 The SDE and some motivation . . . 29

4.2 Intuition for the numerical scheme . . . 30

4.3 Numerical approximation of the regularized SDE . . . 31

4.4 Convergence of the regularized solution to the real solution . . . 43

4.5 The convergence rate . . . 51

4.6 Implementation of numerical methods . . . 54

4.7 Fokker-Planck equation and the law density of X . . . 62

5 McKean SDEs with irregular drift **67**

5.1 Introduction . . . 67

5.2 Problem statement . . . 67

5.3 Analytical results . . . 70

5.4 Regularity properties of ρ^N . . . 73

5.5 Three key bounds for B^N . . . 77

5.6 Convergence rate with an exact solution to the Fokker-Planck PDE . . . 79

5.7 Numerical approximation of ρ^N . . . 84

5.8 Implementation of numerical methods . . . 88

List of Figures

4.1 From left to right: $\mathcal{I}^N(y)$ for $y \in [-5, 5]$ and $1/N = 0.199494, 0.068111$ respectively. 57

4.2 Comparison of and fBm with Hurst parameter $H = 1 - \epsilon$ for $\epsilon = 10^6$ and the drifts generated with 2^{11} and 2^{17} time steps which translates into $N = 445$ and $N = 12407$ respectively. The drift b has regularity $-\beta$ for $\beta = \epsilon$. 60

4.3 Comparison of and fBm with Hurst parameter $H = 1/2 + \epsilon$ for $\epsilon = 10^6$ and the drifts generated with 2^{11} and 2^{17} time steps which translates into $N = 155$ and $N = 2448$. The drift b has regularity $-\beta$ for $\beta = 1/2 - \epsilon$. . . . 60

4.4 Average empirical convergence rates obtained from 40 Euler-Maruyama with drifts of the same regularity but generated for each run. Approximations have 10^4 sample paths for each β including 95% confidence interval for the average, the conjectured rate of $1/2 - \beta/2$ and the theoretical rate obtained in Theorem 37. 62

4.5 Law density of the SDE obtained by solving the associated Fokker-Planck PDE (green line) and the histogram of the empirical density (red bars) at time $t = T$ for an SDE with drift derived from a fBm with Hurst parameter $H = 1 - \epsilon$, or a drift in $C^{-\epsilon}$, for $\epsilon = 10^{-5}$. 64

4.6 Law density of the SDE obtained by solving the associated Fokker-Planck PDE (green line) and the histogram of the empirical density (red bars) at time $t = T$ for an SDE with drift derived from a fBm with Hurst parameter $H = 1/2 + \epsilon$, or a drift in $C^{-1/2-\epsilon}$, for $\epsilon = 10^{-5}$. 65

5.1 Example of convergence rate obtained for an SDE with a singular fixed drift $F(\hat{\rho}^N)b^N$ with $X_0 \sim \mathcal{N}(0, 1)$, hence we have $\rho_0(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$, and $F(x) = \sin(x)$. The error is computed over 40 runs, each with 10,000 samples, of the Euler-Maruyama scheme with different Brownian motion drivers. The empirical rate is compared with the theoretical rate and a 95% confidence interval is drawn around the empirical rate. 91

5.2 Example of convergence rate obtained for an SDE with multiple drifts $F(\hat{\rho}^N)b^N$ with $X_0 \sim \mathcal{N}(0,1)$, hence we have $\rho_0(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$, and $F(x) = \sin(x)$. The error is computed over 40 runs, each with 10,000 samples, of the Euler-Maruyama scheme with different Brownian motion drivers and 40 different drifts with the same regularity. The empirical rate is compared with the theoretical rate and a 95% confidence interval is drawn around the empirical rate. . . . 91
5.3 Effect of the product $F(\hat{\rho}^N)$ with b^N for the function $F(x) = \sin(x)$. Distribution approximations b^N with smoothing parameter $1/N$ for $N = 2448, 155$ and regularity $-\beta = -0.49$. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in x. . . . 93
5.4 Effect of the product $F(\hat{\rho}^N)$ with b^N for the function $F(x) = 1/(1 + \exp(-100(x-0.2)))$. Distribution approximations b^N with smoothing parameter $1/N$ for $N = 7926, 333$ and regularity $-\beta = -0.25$. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in x. . . . 93
5.5 Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.49}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$. . . . 95
5.6 Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.25}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$. . . . 96
5.7 Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.01}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$. . . . 97

List of Tables

4.1 Minimum amount of points $2M + 1$ needed in the discretization $\mathfrak{M}$ when $m = 2^{12}$ and $L = 5$. This varies according to the variance $1/N$, which in turn is a function of β 58

4.2 Average of empirical convergence rates obtained using 50 Euler-Maruyama approximations with 10^4 sample paths for each β, the conjecture rate of $1/2 - \beta/2$ and the theoretical rate from Theorem 37. All values rounded to 5 decimal places. Same values are plotted in Figure 4.4. 61

4.3 Results of the Kolmogorov-Smirnov test for the goodness of fit of the PDE approximation of the law density with respect to the empirical density for SDEs with drift in the space $C^{-1/4}$ and time steps $2^{11}, 2^9, 2^7$. 65

5.1 Results of the Kolmogorov-Smirnov test for the goodness of fit of the PDE approximation of the law density with respect to the empirical density for McKean-Vlasov SDEs with drift in the space $C^{-0.25}$ and time steps $2^{11}, 2^9, 2^7$ and $F(x) = \cos(x)$. 94

Chapter 1

Introduction

Stochastic differential equations (SDEs) have been a topic of interest in the stochastic calculus community since its beginning. Classical results on the existence and uniqueness of solutions to SDEs are used on a daily basis in multiple disciplines where stochastic calculus is paramount. Just as in the case of deterministic differential equations, SDEs often do not have explicit solutions and in order to use them in practice we must solve them numerically. Although classical results on SDE approximations date back to Maruyama's work [35] 70 years ago, there is currently great interest in relaxing the regularity assumptions of the coefficients and see how far the tried and tested iterative methods can take us. One of the most widely used references for numerical analysis of SDEs, the book by Kloeden and Platen [29], just before their result on the convergence rate of the Euler-Maruyama scheme they write:

> "the Euler scheme gives good numerical results when the drift and diffusion coefficients are nearly constant... however, it is not particularly satisfactory and the use of higher order schemes is recommended".

This statement is partially true, however a phenomenal effort has been made since that book was written to prove that the Euler scheme suffices to solve SDEs with a much larger family of coefficients, and in particular for very rough drifts and an additive noise as we will study in the present thesis.

We note that the Euler scheme is often the default choice of numerical method to solve SDEs in practice, and in fact since many of the theoretical questions deal with the case of additive noise and very rough drifts—for instance where the motivation is regularization by noise to solve ODEs—following the advice in [29] will take us away from the more widely used Euler and Milstein schemes, of order 1/2 and 1 respectively, because those two are equivalent in that case which will make us resort to schemes of order 3/2 and higher. However, the implementation of higher order methods requires the usage of derivatives of the coefficients, and increasing the roughness of coefficients implies that these derivatives will not exist in the classical sense and if needed they must be obtained in non-standard ways. Hence, the simplicity of the Euler-Maruyama scheme is a great motivation to find out if it can be used in broader scenarios.

The interest on SDEs with distributional coefficients comes from both theoretical and applied questions. For instance some authors have found applications in the development of models for random irregular media. Such is the work of Russo and Trutnau [43] where they give a probabilistic representation to a parabolic PDE, with a diffusion process given by an SDE with a drift which is the distributional derivative of a continuous function, which it is ultimately used in the study of SPDEs. Another work is that of Hu et al. [22] have used SDEs with this kind of drifts—in particular for distributional derivatives of a two-sided Brownian motion—to represent diffusions in random media such as Bronx diffusions. There is also the work of Cannizzaro and Chouk [8], which is a multidimensional generalization of [15], where they use a martingale problem approach to prove the well-posedness of SDEs with distributional drifts in the Hölder-Besov space of negative order, which they ultimately use to study behaviour of the polymer measure. The works mentioned above are of a theoretical nature, trying to find well-posedness of the equations they approach. However, in these cases one is not able to obtain exact solutions for the SDEs under study, and here is where questions of convergence of numerical methods arise. Addressing such questions, although in different settings, we mention the work by De Angelis et al. [14] and Goudenège et al. [19]. In the former, the authors study an SDE with additive Brownian noise and a drift in belonging to of fractional negative Sobolev spaces and design numerical methods providing a lower bound for the rate of convergence. Whereas in the latter, they implement numerical methods for an SDE with a fractional Brownian motion as a noise and drifts in a Besov space of negative order.

One can see that the implementation of numerical methods for irregular drifts, can sometimes lag behind the theoretical results, and although one can benefit of simply knowing that convergence can be attained and leave to other authors the implementation of the corresponding numerical methods, making the design of implementations equal in importance to the theoretical questions can give light to other interesting approaches to the study of the SDEs. Additionally, once the convergence results have been obtained, the numerical implementations can pose an extra significant challenge since coefficients of irregular nature are sometimes not trivial to implement. This is especially noticeable for distributional drifts, where numerical techniques can be quite challenging due to the fact that it is not possible to simulate a distribution numerically—at least not directly—and one must come up with clever techniques to accomplish such task.

In this work we develop estimates for the numerical error for approximations of SDEs and McKean-Vlasov SDEs with a distributional drift and an additive Brownian noise. In particular, the drifts we study belong to a negative Besov space, which one can think of an extension to negative regularity of the Hölder spaces. We design and implement robust numerical methods to solve said equations and test them in order to compare the theoretical and empirical rate of convergence. We note that the obtained convergence rate is a lower bound and, as such, can be tightened using different techniques. Our work is similar in spirit to the aforementioned work by De Angelis et al. [14], but it parts from the study by Issoglio and Russo in the papers on SDEs and McKean SDEs with distributional coefficients [27, 26] where the drifts belong to negative Besov spaces. And in comparison with the work of Goudenège et al. [19] we are able to address the case of a standard

Brownian noise rather than having extra regularity obtained by a Hurst parameter in the noise which is smaller than 1/2.

This thesis is organized as follows:

1. In Chapter 2 we present a literature review where we explore the work done by multiple authors in the field of SDEs with irregular coefficients. We see the progression in the loss of regularity and how these problems are addressed for different regimes, starting with the more restrictive settings involving only discontinuous functions and gradually moving towards the most general cases with distributional coefficients. We place special emphasis on the study of numerical methods for these kinds of equations.

2. Chapter 3 contains some preliminary theory on the well-posedness of the problems we address numerically. We set the stage for the coming numerical work presented in the next chapters by presenting the fundamental tools that we use and which have been developed prior and in parallel to the present work.

3. Chapter 4 covers the first problem we tackle: the convergence of an Euler-Maruyama scheme for an SDE driven by additive Brownian motion and with a drift taking values in a negative Besov space. We use the concepts of mild solutions to PDEs and estimates for them that allow us to propose a bound for the strong error in L^1 between the approximated solution and the real solution. Additionally, we implement the numerical methods and obtain an empirical rate that improves from the bound found theoretically and study of the law density of the solution to the SDE in terms of the associated Fokker-Planck equation.

4. Finally, in Chapter 5 we use the techniques that we learned in the development of Chapter 4 for an SDE of the McKean-Vlasov class. We add an extra step to the numerical methods of solving the one dimensional Fokker-Planck PDE associated to the SDE to tackle the need of the density of the solution, for which we need to adjust our estimates for the strong error in L^1, still obtaining the same convergence rate as for the SDEs in Chapter 4. It also provides some insights into the numerical study of the Fokker-Planck equation associated with the McKean SDE with a distributional drift, and moreover, facilitated the development of numerical tools to verify accuracy of the iterative approximations.

Chapter 2

Literature review

The classical setting on existence and uniqueness of solutions of SDEs and furthermore, for numerical solutions, requires strong regularity conditions which for modern applications of stochastic calculus can severely limit the class of SDEs that one can study. For this case the reader is referred to any textbook on stochastic calculus such as Øksendal's [38], Baldi's [4] or Karatzas and Shreve's [28]. Because of the limitations imposed classically, a vibrant research area has sprouted in an attempt to extend the well posedness of SDEs into more generalized scenarios.

For an SDE it is well known that strong existence and uniqueness is guaranteed when the coefficients are Lipschitz continuous and have linear growth, c.f. Theorem 5.2.1 from [38]. Additionally, the results on convergence of numerical methods for classical SDEs are a direct consequence of the existence and uniqueness of strong solutions, in fact the conditions to obtain a convergence rate of iterative schemes such as the Euler-Maruyama method[1], are not very different from the conditions to find strong solutions. See for example Theorem 10.2.2 from [29]:

Theorem 1. *Let us have an SDE*

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \sigma(s, X_s)dW_s,$$

for a one-dimensional Brownian motion W *and* $t \in [0, T]$*, and an iterative Euler-Maruyama approximation with uniform time steps* $t_0, \dots, t_N$

$$Y_{n+1}^{\Delta} = Y_n^{\Delta} + \mu(n, Y_n^{\Delta})\Delta + \sigma(n, Y_n^{\Delta})\Delta W_n,$$

where $\Delta = t_{n+1} - t_n$ *for all* $n \in [0, N]$ *is the length of a time step and* $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ *is the* n*-th increment of the Brownian motion* W*. Suppose that the following conditions hold:*

1. $\mathbb{E}|X_0|^2 < \infty$,
2. *There exists* K_1 *such that* $\mathbb{E}(|X_0 - Y_0^{\Delta}|^2)^{1/2} \leq K_1\Delta^{1/2}$,

[1] We may also call it simply Euler scheme or method.

3. *There exists K_2 such that for all $t \in [0,T]$ and $x, y \in \mathbb{R}$ we have $|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K_2|x-y|$,*

4. *There exists K_3 such that for all $t \in [0,T]$ and $x, y \in \mathbb{R}$ we have $|\mu(t,x)| + |\sigma(t,x)| \leq K_3(1+|x|)$,*

where the constants K_1, K_2, K_3 are independent of Δ. Then, for the Euler-Maruyama approximation Y^Δ, the estimate

$$\mathbb{E}|X_T - Y_N^\Delta| \leq K_4 \Delta^{1/2}$$

holds, for a constant K_4 independent of Δ.

The exponent 1/2 in the last equation of the theorem above is what we call the *convergence rate* of the Euler-Maruyama method. Some authors prefer to use instead of Δ the parameter N which signifies the number of time steps, in that case the notation for the last inequality would be

$$\mathbb{E}|X_T - Y_N^\Delta| \leq K_6 N^{-1/2}.$$

Note that conditions 3 and 4 of the above theorem are exactly what we have discussed before for existence and uniqueness of solutions, Lipschitz continuity and linear growth, therefore this result will also break apart when we deviate from this pair of very strong regularity assumptions, and enter into the realm of SDEs with irregular coefficients.

The first attempts to study SDEs with irregular coefficients are fairly modern, even in comparison with the study of stochastic calculus. In the literature, one will find that 'irregular coefficients' for SDEs can mean very different types of regularity or other conditions, and this of course starts the moment one steers away from Lipschitz continuity and/or sub-linear growth. Considered one of the seminal works in the area, the article by Zvonkin [49], with its legacy reflected in the myriad of implementations of their techniques, proposed a transformation of an SDE with drift being merely measurable and bounded, whereas the diffusion was continuous in its two variables. Said transformation—which comes from the infinitesimal generator of the SDE—has the powerful effect of removing the drift of the SDE, and by its nature requires of some PDE theory to build their arguments. After some years, Veretennikov [48] expanded Zvonkin's work in to the multidimensional case, finding that for additive noise and a simply bounded drift, they can attain a strong solution. Both of these works consider the driver of the equation being a Brownian motion and set the stage for the work that has been done in more recent years.

Let us start with a slightly less irregular setting, which is partially motivated by applications in energy management, this is for SDEs that have coefficients which simply exhibit some discontinuities. Among this line, the work of Leobacher et al. [34] attempts to use alternative methods to those of [48] and [49] which rely on the ellipticity of the SDE generator, so they refer to SDEs with a potentially 'degenerate' diffusion in the sense of degenerate ellipticity. They prove that this is possible by making a transformation of the SDE removing the discontinuities and having a new SDE in terms of which the solution to the time homogeneous SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{2.1}$$

can be found for μ that is discontinuous at a singular point, and sufficiently smooth at the rest, and where σ is locally Lipschitz—component-wise.

Then in a sequel, Leobacher and Szölgyenyi [31] allow for a finite number of discontinuities in the drift, provided in between those discontinuities the Lipschitz condition still holds, and using again a transformation G that they can invert to recover the solution of the original SDE (2.1). Once the solution Z of the transformed SDE is obtained, it is then possible to use the Euler-Maruyama scheme for it, to get the approximation $\bar{Z}$ and define the approximation for (2.1) as $\bar{X} = G^{-1}(\bar{Z})$. This result is in one dimension, but it admits a non-trivial multidimensional extension presented in [32] which requires some conditions on the surface of discontinuities and furthermore, the transformation G has to be inverted numerically incurring in a high cost of computation. In both the one-dimensional and multidimensional case, the convergence rate of the Euler-Maruyama approximation is of order $1/2$. Notice that equation (2.1) is time homogeneous, but in [44] they extend the result on the existence of a unique strong solution into the time inhomogeneous case in order to solve an optimal control problem in the setting of energy storage, using the same transformation techniques as in the prior works. In an attempt to find a simpler method to solve to multidimensional SDEs, Leobacher and Szölgyenyi [33] propose a clever trick to calculate the error and simply using the Euler-Maruyma approximation of the transformed SDE for $Z = G(X)$, i.e. Z^δ without the need to invert G. Considering the SDE has a solution X and an Euler approximation X^δ they can see the error of the approximation as

$$\|X - X^\delta\|_2 = \|G^{-1}(Z) - G^{-1}(Z^\delta)\|_2$$

and then split it into the terms $\|Z - Z^\delta\|_2$ which is the error of the numerical approximation to the new transformed SDE, and $\|Z^\delta - G(X^\delta)\|_2$, which is difference of the approximation to the new transformed SDE and the transformation of the approximation to X. For this simplified method they obtain a strong convergence rate of $1/4$ saving themselves a few practical challenges. Another one dimensional case in which the convergence rate of $1/2$ is attained is when the SDE contains Poisson jumps, this is an equation of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + \rho(X_t)dN_t$$

with $X_0 \in \mathbb{R}$ and $(N_t)_{t\geq 0}$ a Poisson process with intensity λ. This equation is studied in [40] and they impose a few restrictions in the real-valued coefficients, namely the drift μ is piecewise Lipschitz with discontinuities on $\zeta_1, \ldots, \zeta_m$, the diffusion coefficient σ is Lipschitz and non-zero on the collection discontinuity points of μ and finally the so-called jump coefficient ρ is Lipschitz. All of them additionally have sublinear growth. They propose a similar technique of transformation of the SDE obtaining a new SDE which is solved numerically and has an optimal strong rate of convergence.

Moving away from merely discontinuous coefficients, we recall the above-cited article by

Veretenikov [48] tackles the problem of existence of strong solutions for SDEs with bounded measurable drifts and multiplicative noise. However, this result was only theoretical and it was not until the Gyöngy and Krylov [21] found that indeed well-posedness of the Euler scheme for the SDEs proposed by Veretennikov holds and there is convergence in probability of the numerical solutions. The last result was only qualitative, and more recently Dareiotis et al. [13] take on the task of providing convergence rates to Gyöngy and Krylov's result, obtaining in this case a convergence rate $1/2-\epsilon$ for some small $\epsilon>0$. The authors discuss that for an additive drift the rate of convergence is not ideal, since in the classical case this rate must be 1, so they plan to pursue the regularity condition for the drift which will yield a convergence rate akin to the one for the Milstein scheme. Moreover, when the noise is only additive they are able to find a convergence rate for a drift with Sobolev regularity of order $\alpha\in(0,1)$. This result can be used as a benchmark for the convergence rates of schemes for SDEs with distributional coefficients in the sense that we would expect that if the drift has a negative regularity γ, as $\gamma\to 0$ from the left we would expect to have a convergence rate that agrees with a convergence rate in [13] when $\alpha\to 0$ from the right. In a different work Butkovsky et al. [7] use stochastic sewing lemma, exploiting the effects of regularization by noise, to find a convergence rate $((1/2+\alpha H)\wedge 1)$ for SDEs with additive noise, but this time they work on SDEs driven by fractional Brownian motion with Hurst parameter $H\in(0,1)$ and a drift living in a Hölder-Besov space C^α for $\alpha\in[0,1]\cap(1-1/(2H),1]$. Notice that this is indeed one step in the pursuit of a Milstein like convergence rate that Dareiotis et al. mentions in [13], albeit being aided by the regularity properties of a fractional noise. In the same work for the case of multiplicative standard Brownian noise—$H=1/2$ —they reiterate the convergence rate obtained in [13].

Losing more regularity, we step into the realm of distributional coefficients. Some of the starting works on this regard is by Flandoli, Russo and Wolf [17, 18] where the authors discuss an SDE in one dimension and with a drift that is defined as the derivative of a one dimensional function g that is only continuous, this is $f(t,x)=g'(x)$, and so they have time homogeneous one-dimensional SDEs of the type

$$X_t = x_0 + \int_0^t \sigma(X_s)dW_s + \int_0^t g'(X_s)ds,$$

where σ and g are merely continuous functions. Solutions to this equation are considered as a singular martingale problem or in the sense of solutions in law or weak—an approach that we will find familiar in the coming chapters.[2] For the equation above, we have the formal operator

$$\mathcal{L}f = \frac{\sigma^2}{2}f'' + g'f'$$

and in [17, Proposition 3.11 and Corollary 3.12] the authors find that if the operator $\mathcal{L}$ is close to divergence form, this is $\mathcal{L}^d = f = (\frac{\sigma^2}{2}f')' + (g^d)'f$ where $g^d = g - \frac{\sigma^2}{2}$, and if X

[2] In order to avoid confusions with weak solutions of PDEs, when talking about weak solutions of SDEs we will often prefer to call them solutions in law.

solves the martingale problem with respect to $\mathcal{L}$ then X is a solution in law to the SDE

$$X_t = X_0 + \int_0^t \sigma(X_s)dW_s + A(b)$$

for $b = \frac{\sigma^2}{2} + \beta$, where A is the extension onto $C^0(\mathbb{R})$ of the map $l \mapsto \int_0^{\cdot} l'(X_s)ds$ which is defined on $C^1(\mathbb{R})$. In the sequel [18] the same SDE is explored, but now there is a development of an Itô formula when the drift is L^∞_{loc} and the functions f applied are in a Sobolev space $W^{1,2}_{\text{loc}}$. Furthermore, they find functions f such that $f(X)$ is a semi-martingale.

An example of works where the drift is a significantly different type of distribution was written by Bass and Chen [5] where the SDE of interest is

$$dX_t = dW_t + dA_t$$

where $A^i_t = \int_0^t f_i(X_s)ds$ for some n-dimensional vector $A_t = (A^1_t, \dots, A^n_t)$ and $f_t(x)dx$ is the Revuz measure of A^i_t for f bounded. In this case the goal is to replace $f_i(x)dx$ by some more generic signed measures π_i. In particular, measures of the Kato class were considered, i.e.

$$K_\alpha = \left\{\pi(dx) : \limsup_{\epsilon\to 0} \sup_{x\in\mathbb{R}^d} \int_{B(x,\epsilon)} |x-y|^{-\alpha}|\pi|(dy) = 0\right\}$$

where $|\pi|$ is the total variation of the measure π. Then they say that $f \in K_\alpha$ if $f(x)dx \in K_\alpha$. The idea is then to find a process in $\mathbb{R}^3$ that behaves as a Brownian motion outside the domain Γ but inside of it drifts upwards, where $\Gamma = A \times \mathbb{R}$ and A is the Sierpinski triangle in $\mathbb{R}^2$. The main result tells us that we can indeed find a solution in law to the SDE of interest, and although the paper presents it for $d \geq 3$ it also holds for $d \geq 2$.

With an SDE again with a unit diffusion Delarue and Diel [15] have a drift $\partial_x Y_t(X_t)$ where the paths $(\mathbb{R} \ni x \mapsto Y_t(x))_{t\geq 0}$ are continuous. This is a one dimensional problem and the derivative $\partial_x Y_t(X_t)$ is in the sense of generalized functions and since Y_t is time dependent, so is its derivative. They also go for the approach of the Martingale problem and use the associated PDEs with the appropriate operators. Then Cannizzaro and Chouk [8] generalize this work for d-dimension using paracontrolled distributions as per the theory in [20].

In one of the main steps towards the work in this thesis, Flandoli, Issoglio and Russo [16] deal with the SDE

$$\begin{cases} dX_t = b(t, X_t)dt + dW_t \\ X_0 = x_0 \end{cases}$$

for a d-dimensional Brownian motion W_t and a drift b taking values in a fractional Sobolev space, in particular they say that $b \in L^\infty([0,T]; H^{-\beta}_{\tilde{q},q})$ for $\beta \in (0, 1/2)$, $q \in (\frac{d}{1-\beta}, \frac{d}{\beta})$ and $\tilde{q} := \frac{d}{1-\beta}$. They introduce the concept of virtual solution, which is very useful to create bounds in terms of the solutions to a class of associated PDEs.

The present work draws much inspiration from the work of De Angelis et al. [14], in fact the initial stages of the project where a mere generalization of their tools for a different type of drift. In [14], they work in the same setting for the drifts as [16] but in the one-dimensional case. One key difference is that in dimension one, this type of equations have a strong solution. They approach the problem by observing that the space of negative regularity $H^{-\beta}_{q,\tilde{q}}$ will have as a base the family of Haar functions and approximate the drift b as a truncated series of such functions. Furthermore, b^N is the regularized version of the truncated series that is obtained taking a convolution of the heat kernel in what they call a randomization procedure to control the convergence rate appropriately. Their usage of Haar functions is also for the sake of numerical efficiency in their implementations. For a drift $b \in C^\kappa([0,T]; H^\beta_{\tilde{q},q})$ for $\kappa \in (1/2, 1)$, and $\beta \in (0, 1/4)$, $q \in (4, 1/\beta)$, $\tilde{q} = 1/(1-\beta)$, and a convergence rate in L^1 of the numerical approximation of the form

$$-\frac{(\frac{1}{2}-\beta)(\gamma-\frac{1}{2})}{2(\frac{3}{4}-\beta(\gamma-\frac{1}{2}))}+\epsilon,$$

for $\gamma = 1-\beta-\frac{1}{q}$. Doing some substitution one can see that this rate ranges between $\frac{1}{6}$ for $\beta \sim 0$ and 0 for $\beta \sim 1/4$.

The theoretical framework that we take the most advantage from is done by Issoglio and Russo in three articles, namely [27, 25, 26]. The first pair of articles are part of the same underlying theory, and are the base of the work on SDEs with distributional drift which is presented in Chapter 4, whereas the third one on McKean-Vlasov SDEs which will be discussed on Chapter 5. All of those papers will also be discussed in greater length in Chapter 3, where we talk precisely about some of the results on well-posedness of the problems we attack numerically.

In [27], they study the SDE

$$dX_t = b(t, X_t)dt + dW_t,$$

with $X_0 \sim \mu$, where $X_t \in \mathbb{R}^d$, and $(W_t)_{t\geq 0}$ is a d-dimensional Brownian motion, μ is a probability measure and the drift $b(t,\cdot) \in C^{(-\beta)+}$. In Chapter 4 we will get deeply familiar with the one-dimensional version of this SDE, but we may recall that this equation is also intimately related to the ones in [17, 18] with a different kind of drift, and in the multidimensional case as we have seen there is the paper [16]. Their approach uses the interpretation of the mentioned SDE as a singular martingale problem which, we will discuss in Section 3.4, and hence they rely on the usage of a PDE

$$\begin{cases} \partial_t f + \frac{1}{2}\Delta f + \nabla f b = \lambda f + g \\ f(T) = f_T, \end{cases}$$

for $\lambda = 0$, which is studied in the sibling paper [25] for a $\lambda \in \mathbb{R}$, where the authors are concerned with the questions of existence and uniqueness for its solutions. Note that the left-hand side of the equation above is the infinitesimal generator of the process X, and that well posedness of said PDE allows them to provide a meaning for the martingale problem, which depends on that operator. They also provide the dynamics followed by the solution of the SDE, and with some considerations find equivalent results to the ones by Stroock and

Varadhan [45] by finding an equivalence between the solution to the martingale problem and a suitable notion of solution of the SDE. In terms of the study of the PDE above, which is fundamental to solve the SDE problem, in [25] the authors develop a meaning to this class of PDEs. Variations of the parameter λ will lead to different equations which can be useful in the study of their related SDEs. They investigate the action of the heat semigroup on the PDEs, in particular Schauder and Bernstein inequalities of which we will take big advantage, in our work, when finding bounds for the convergence of numerical methods. In one of the most beneficial techniques for us, they give the notion of weak and mild solutions of that PDE and their equivalence. Then prove that mild solutions are unique using the properties they develop of the application of the heat semigroup.

A generalization of SDEs are the so called McKean-Vlasov equations, where the coefficients can depend on the distribution of the solution to the SDE. Issoglio and Russo [27] study a McKean-Vlasov SDE whose drift has the same regularity as in [27] with a very distinct dependency of the drift on the density of the solution with an additive Brownian noise. They study the d-dimensional McKean-Vlasov equation

$$X_t = \int_0^t F(\rho(s, X_s))b(s, X_s)ds + W_t,$$

where $\rho(t, \cdot)$ is the law density of X_t, for a random initial condition X_0 with density ρ_0. They assume that $b(t, \cdot) \in C^{(-\beta)+}$ and the function F is nonlinear bounded and with bounded derivative, so the overall regularity of the drift is encoded in b and is what poses the problem of assigning meaning to the SDE. As in [27], this paper also uses analytical tools but this time in the form of the Fokker-Planck PDE

$$\begin{cases} \partial_t \rho = \frac{1}{2}\Delta\rho - \operatorname{div}(\rho F(\rho) b) \\ \rho(0) = \rho_0, \end{cases}$$

for which they find a unique solution $\rho \in C([0, T]; C^\beta)$ in the sense of distributions, which they use to find their main result on the existence and uniqueness of solution of the McKean SDE.

On a related work by Issoglio et al. [24] the authors prove existence and uniqueness for a singular martingale problem associated to a kinetic McKean-Vlasov system of equations for the velocity V_t and position X_t of a particle in the phase-space. They key difference with [26] is the kinetic property of the system, which means that the Brownian motion only drives the term which represents the velocity. They do it by thinking about their martingale problem as a case of a class of linear, rather than McKean, martingale problems; which requires knowing the density of the solution to the equation since it has to be an input of the linear martingale problem. Then they find a solution to this linear singular martingale problem, which they take in ensuring that it coincides with the classical Stroock-Varadhan problem in the smooth case. Finally, since the above represents a solution by construction they verify that the construction indeed solves the problem at hand.

Another approach in the setting of distributional drifts with more generic noises, the work by Anzeletti et al. [1] deals with the one dimensional SDE

$$dX_t = b(X_t)dt + dB_t$$

where the driver noise is a Fractional Brownian motion with Hurst index $H \in (0,1)$ and the drift b belongs to the space $B^{\beta}_{p,\infty}$, which is written as B^{β}_{p} for convenience.[3] Proving that a Weak solution exist given certain conditions for β, p, H. In particular, they show that if $b = a\delta_0$ where, $a \in \mathbb{R}$ and δ_0 is the Dirac delta centred at zero then a weak solution exists for any $H < \sqrt{2} - 1$. Their technique to prove existence of a weak solution require of considering another, more general, SDE

$$dX_t = b(X_t)dt + dZ_t$$

for some continuous stochastic process Z, which they solve for a fixed trajectory of the noise $(Z_t(\omega))_t$ by rewriting said equation into a random ODE of the form

$$dY_t = b(Y_t + Z_t)$$

where $Y_t = X_t - Z_t$. They study the regularity of the operator $T^Z_t b : y \mapsto \int_0^t b(y, Z_s)ds$ and find out that in some cases the transformation $T^Z_t b$ is more regular than b, and then the ODE admits a solution $Y_t = Y_0 + \int_0^t T^Z_{ds} b(Y_s)$. Then, they prove that this solution to the ODE is equivalent to the solution to their original SDE, so they prove that the ODE has a solution of the form above. Additionally, they show that a weak solution always exists if the drift is in the space $B^{1/4}_1$. Then, using stability arguments and regularity estimates powered by the stochastic sewing lemma by Lê [30], they go on to prove that pathwise uniqueness holds for weak solutions provided that b is in $B^{\beta}_{p,\infty}$ with $\beta - 1/p \geq 1 - 1/2H$. And when b is a measure they are able to prove that strong solutions have sufficient Hölder regularity. And obtain strong existence and uniqueness for $b = a\delta_0$ and $H \leq 1/4$.

The work that, to our knowledge, resembles the closest the results of the present text is [19] where the authors study the multidimensional SDE

$$X_t = X_0 + \int_0^t b(X_t)ds + \int_0^t dB_t$$

which has as a driver the fractional Brownian motion $(B_t)_{t\geq 0}$ in d dimension with Hurst coefficient $H < 1/2$, and a time homogeneous drift $b \in B^{\gamma}_{p}$ with a *regularity assumption* $1 - \frac{1}{2H} \leq \gamma - \frac{d}{p} < 0$. It is worth noticing that the requirement for the Hurst coefficient is paramount for this approach, because when the threshold $H = 1/2$ is reached then it is no longer possible to handle distributional drifts but only functional. In that case, however, they prove that they extend the regime studied by [7]. They make an extension of the existence result by [1] for any dimension $d \geq 1$ by claiming that for $b \in B^{\gamma}_{p,\infty}$ and satisfying their regularity assumption there is a weak solution to the corresponding SDE. Then they

[3] See Section 3.1 for the definition of Besov spaces.

perform a standard Euler-Maruyama scheme approximation

$$X_t^{h,n} = X_0 + \int_0^t b^n(X_t^{h,n})ds + \int_0^t dB_t$$

for some time step h—which is a carefully chosen function of n—and an approximation $(b^n)_{n\in\mathbb{N}}$ of b in the appropriate space. On this setting they obtain a convergence rate

$$\frac{1}{2(1-\gamma+\frac{d}{p})} - \epsilon$$

in the space $C^{1/2}_{[0,1]}L^m(\Omega)$ for $m \geq 2$. Notably, if $p = \infty$ and $d = 1$ the convergence rate $\frac{1}{2(1-\gamma)} - \epsilon$, which is the setting we will explore in depth in the present work, we will make further comparisons of this in the coming chapters. Additionally, they study the limiting case when $\gamma - \frac{d}{p} = 1 - \frac{1}{2H}$, which leads to a non-explicit convergence rate $1 - \frac{1}{2H}$.

Another common case of extension of the noise is to use α-stable noises as the driver. There is the work by Athreya et al. [2] for a one dimensional time homogeneous SDE with a symmetric α-stable process L_t with $\alpha \in (1, 2)$. They see the drift as the limit of the integral $\int_0^t b^n ds$, for a sequence b^n which approximates the drift b in C^β for a negative β as seen as well in the work by [1], and their solution is of the form

$$X_t = x + A_t + L_t$$

where the term A_t satisfies

$$A_t^n = \int_0^t b^n(X_s)ds \to A_t.$$

They prove pathwise uniqueness for $\beta \in (\frac{1-\alpha}{2}, 0]$, in a way extending the result by Tanaka et al. [46] into distributional drifts. Additionally prove strong existence on the solution when

$$\mathbb{E}|A_t - A_s|^2 \leq C|t-s|^{2k}$$

for any $k \in (0, 1]$.

In the inhomogeneous drift, and multi-dimensional setting we can see among others the work [11], where the SDE on focus has an α-stable additive noise $\mathcal{W}_t$ and the drift is in the regime of Besov spaces, with only integrability required for the time parameter, specifically the drift belongs to the space $L^r([0,T]; B^{\gamma-1}_{p,q}(\mathbb{R}^d, \mathbb{R}^d))$. The stability parameter of the noise is $\alpha \in (1, 2]$, which includes the Brownian motion case and the jump case, notably it does not include either the lower bound of the so-called *sub-critical* case—where $\alpha \geq 1$—nor the *super-critical* case—where $\alpha < 1$—which is included in works such as [39] and [12] respectively but only for Hölder continuous drifts. The questions concerning [11] are theoretical in nature, in particular to establish conditions for well-posedness of the equation but, contrast with the works [16, 17, 18, 5], here the drift is obviously more generic, and they have to control the parameters p, q, γ, r, α. As the aforementioned works, they also rely on a Zvonkin type transformation and develop Schauder type estimates for the mild solution of the associated PDEs. In doing so, the prove the well-posedness of the martingale problem provided that $p, q, r \geq 1$ and $\gamma \in (1/2, 1)$ for all $\alpha \in (1, 2]$. Additionally, for some slightly stronger conditions they derive the dynamics of the martingale problem, which allows them to build a weak solution, extending the result from [2], which is one dimensional. They make some hints about potential usage of their tools for numerical approximations, but do not delve into that problem.

Chapter 3

Setting of the problems and preliminary mathematics

The purpose of this chapter is to stablish the setting that is common for the two problems we solve. This includes results and derivations that are helpful for both problems.

We will be working with a class of SDEs of the form

$$X_t = X_0 + \int_0^t b(s, X_s) ds + W_t, \tag{3.1}$$

where $(W_t)_{t\geq 0}$ is a Brownian motion and $b(t, \cdot)$ belongs to the space of Schwartz distributions $\mathcal{S}'(\mathbb{R})$ for all $t \in [0, T]$. More precisely, the map $t \mapsto b(t)$ is $\frac{1}{2}$-Hölder continuous on $[0, T]$ with values in the Hölder-Zygmund space $C^{-\gamma}$ of negative order $-\gamma < 0$, which we denote by $b \in C_T^{1/2} C^{-\gamma}$.

3.1 Function and distribution spaces

We will start by setting notation that will hold for the rest of the document and furthermore defining the spaces on which the problem is posed.

In what follows we will use standard notation for derivatives in "time" and "space", making slight changes depending on convenience, and specially on the type of function we are observing. For single variable functions, $x \mapsto f(x)$ and $t \mapsto g(t)$, that are regular enough, we will denote space derivatives as f', f'', and so on; and time derivatives as $\dot{g}$, $\ddot{g}$, etc. Occasionally we may denote an arbitrary a-th derivative of f as $f^{(a)}$, in these cases we will make an extra effort to be clear as to which variable is being used when required. Whereas if we have two variable functions $h(t, x)$ we will prefer the notation h_x, h_{xx} and h_t, h_{tt}. Where it makes it more readable, or we are applying the derivative to a composite function, we extend this notation as follows $h_x = \partial_x h$, respectively for each case. Furthermore, an a-th partial derivative in x of h will be denoted $\partial_x^a h$, similarly for an a-th derivative in t. The later notation may be also used for single variable functions on occasion.

Definition 2 (Schwartz functions and tempered distributions)**.** Let $\mathcal{S}(\mathbb{R})$ be the space of

Schwartz functions on $\mathbb{R}$, i.e. the space of all C^∞ functions on $\mathbb{R}$ such that for any $k, a \in \mathbb{N}$ there exists $C_{k,a}$ such that

$$\sup_{x\in\mathbb{R}} |x|^k |\phi^{(a)}(x)| \leq C_{k,a}.$$

A sequence $(\phi_n)_{n\in\mathbb{N}}$ converges to ϕ in $\mathcal{S}(\mathbb{R})$ if the derivatives of all orders of the sequence ϕ_n converge uniformly to those of ϕ and

$$\sup_{x\in\mathbb{R}} \sup_{n\in\mathbb{N}} |x|^k |\phi_n^{(a)}(x)| \leq C_{k,a},$$

holds for some constants $C_{k,a}$ which are independent of n.

A *tempered distribution* (also called *Schwartz distribution*) on $\mathbb{R}$ is a linear mapping $\phi \mapsto (f, \phi)$ from $\mathcal{S}(\mathbb{R})$ to $\mathbb{R}$ such that $(f, \phi_n) \to (f, \phi)$ if $\phi_n \to \phi$ in $\mathcal{S}(\mathbb{R})$. We denote the set of tempered distributions by $\mathcal{S}'(\mathbb{R})$. We say that $f_n \to f$ in $\mathcal{S}'(\mathbb{R})$ if $(f_n, \phi) \to (f, \phi)$ for every $\phi \in \mathcal{S}(\mathbb{R})$.

In what follows we will write $\mathcal{S}$ for $\mathcal{S}(\mathbb{R})$ and $\mathcal{S}'$ for $\mathcal{S}'(\mathbb{R})$. In fact this will be the case for any function or distribution spaces on $\mathbb{R}$. Below we will require the definition of the Fourier transform for distributions.

Definition 3 (Fourier transform)**.** The *Fourier transform* of a continuous, absolutely integrable function $f : \mathbb{R} \to \mathbb{C}$ is defined by

$$\mathcal{F}[f](\xi) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-i\xi x} f(x) dx \tag{3.2}$$

In particular this defines the Fourier transform for functions in $\mathcal{S}$, and we have the following result. If we do not need to specify the dependency of a function ϕ to which we apply the Fourier transform, instead of denoting it by $\mathcal{F}[\phi]$ we will shorten the notation it by writing $\mathcal{F}\phi$, the same will be done for the inverse Fourier transform.

Theorem 4 ([41, Theorem 5.63])**.** *If $\phi \in \mathcal{S}$, then $\mathcal{F}\phi \in \mathcal{S}$. Moreover, the mapping $\mathcal{F}$ is continuous from $\mathcal{S}$ into itself.*

The theorem below is about the inversion property for the Fourier transform on distributions.

Theorem 5 ([41, Theorem 5.64])**.** *Let $\phi \in \mathcal{S}$. Then there is a unique $\phi \in \mathcal{S}$ such that $\phi = \mathcal{F}\psi$. Moreover, the inverse Fourier transform of ϕ, i.e. ψ is given by*

$$F^{-1}[\phi](x) := \psi(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} = e^{i\xi x} g(\xi) d\xi \tag{3.3}$$

For distributions, the Fourier transform, and its inverse are defined similarly to the way in which distributions themselves are defined.

Definition 6. Let $f \in \mathcal{S}'$, then the Fourier transform of f is defined by

$$(\mathcal{F}f, \phi) = (f, \mathcal{F}^{-1}\phi), \tag{3.4}$$

for all $\phi \in \mathcal{S}$.

For $\gamma \in \mathbb{R}$, the Hölder-Zygmund spaces are defined by

$$C^{\gamma}(\mathbb{R}) := \left\{ f \in \mathcal{S}' : \|f\|_{\gamma} := \sup_{j \in \mathbb{N}} 2^{j\gamma} \left\| \mathcal{F}^{-1} \left[\phi_j \mathcal{F} f \right] \right\|_{L^{\infty}} < \infty \right\}, \tag{3.5}$$

where $(\phi_j)_{j \in \mathbb{Z}}$ is a partition of unity[1] on $\mathbb{R}$. Hölder-Zygmund spaces $C^{\gamma}(\mathbb{R})$ are a special case of the Besov spaces $B^{\gamma}_{p,q}(\mathbb{R})$ where $p = q = \infty$. For details on the general spaces in our context see [8, Section 2], and for a deeper dive into those spaces in a more analytical context see [3, 47]. See that C^{γ} will denote $C^{\gamma}(\mathbb{R})$. The norm under which the spaces C^{γ} are defined will rarely (if ever) be used by us in our tasks of making estimates, instead we note several equivalent norms for different cases of the parameter γ. Note that for $\gamma \in \mathbb{R}^{+} \setminus \mathbb{N}$, C^{γ} coincides with the classical Hölder spaces, therefore the norm $\| \cdot \|_{\gamma}$ is equivalent to the classical γ-Hölder norm. Recall that if $\gamma \in (0,1)$ and $f \in C^{\gamma}$ the classical γ-Hölder norm is

$$\|f\|_{C^{\gamma}} = \|f\|_{L^{\infty}} + \sup_{\substack{x \neq y \\ |x-y|<1}} \frac{|f(x) - f(y)|}{|x-y|^{\gamma}}, \tag{3.6}$$

and if $\gamma \in (1,2)$ and $f \in C^{\gamma}$ we have

$$\|f\|_{C^{\gamma}} = \|f\|_{L^{\infty}} + \left\|f'\right\|_{L^{\infty}} + \sup_{\substack{x \neq y \\ |x-y|<1}} \frac{|f'(x) - f'(y)|}{|x-y|^{\gamma}}, \tag{3.7}$$

where $\|f\|_{L^{\infty}} = \operatorname{esssup}_{x \in \mathbb{R}} |f(x)|$. Let us also denote the sup norm $\|f\|_{\infty} = \sup_{x \in \mathbb{R}} |f(x)|$ and note that the norm in (3.6) is equivalent to the one from the Hölder-Zygmund spaces from (3.5), hence the abuse of notation.

We will often make use of the following estimate for functions in Hölder-Zygmund spaces which comes as a corollary of [20, Lemma A.1].

Lemma 7 (Bernstein inequality)**.** *For any $\gamma \in \mathbb{R}$ there is $c > 0$ such that*

$$\|f'\|_{C^{\gamma}} \leq c \|f\|_{C^{\gamma+1}}$$

for all $f \in C^{\gamma+1}$.

For a Banach space $(E, \| \cdot \|_E)$, we denote by $C_T E$ the space $C([0,T]; E)$ of continuous functions on $[0,T]$ taking values in E with the norm

$$\|f\|_{C_T E} = \sup_{t \in [0,T]} \|f(t)\|_E.$$

Similarly, for $\kappa \in (0,1)$, we write $C^{\kappa}_T E$ for the space $C^{\kappa}([0,T]; E)$ of κ-Hölder continuous functions on $[0,T]$ taking values in E with the following norm

$$\|f\|_{C^{\kappa}_T E} = \|f\|_{C_T E} + [f]_{\kappa, E},$$

[1] This is a family of functions $(\phi_j(x))_{j \in \mathbb{Z}}$ such that there is a neighbourhood of x where all but finitely many points are non-zero, and such that $\sum_{j \in \mathbb{Z}} \phi_j(x) = 1$ for some $x \in \mathbb{R}$. For a detailed explanation see [8].

where

$$[f]_{\kappa,E} = \sup_{t\in[0,T],t\neq s} \frac{\|f(t)-f(s)\|_E}{|t-s|^\kappa}. \tag{3.8}$$

In the same spaces C_TE we have a family of equivalent norms for $\rho \geq 0$ defined as

$$\|f\|^{(\rho)}_{C_TE} := \sup_{t\in[0,T]} e^{-\rho(T-t)}\|f(t)\|_E. \tag{3.9}$$

Indeed, it is easy to see that

$$e^\rho \|f\|^{(\rho)}_{C_TE} \leq \|f\|_{C_TE} \leq e^{\rho T}\|f\|^{(\rho)}_{C_TE}. \tag{3.10}$$

Finally, for any given $\gamma \in \mathbb{R}$ we denote $C^{\gamma+} := \cup_{\alpha>\gamma} C^\alpha$, and $C^{\gamma-} := \cap_{\alpha<\gamma} C^\alpha$. Similarly, we write $C_TC^{\gamma+}$ which means that there is some $\alpha > \gamma$ such that $f \in C_TC^\alpha$. Additionally, for any γ, the space denoted by DC^γ represents the space of differentiable functions such that their derivative belongs to C^γ, so we have that $DC^\gamma \supset C^{\gamma+1}$.

3.2 The heat kernel and semigroup

We will need to apply the heat kernel as a smoothing procedure in some instances which will include applying it on distributions. Let us introduce a few results and definitions to aid our way.

Recall that we have said in Definition 2 that distributions are linear functionals and in fact one can associate a distribution to a continuous function f via the integral

$$\langle f, \phi\rangle = \int_{\mathbb{R}} f(x)\phi(x)dx \tag{3.11}$$

for any Schwartz function ϕ.

For a distribution T that does not come from a function by associating an integral as above, we will often just call it T after specifying which space it belongs to, e.g. $T \in \mathcal{S}'$, which will automatically tells us over which functions it operates. However, there may be instances where the notation $\langle\cdot,\cdot\rangle$ notation may be preferred even for pure distributions. An example of this is the identity which defines the product of a distribution and a function. This is, for any function $a \in \mathcal{S}$ and a distribution $T \in \mathcal{S}'$ the product aT is understood as

$$aT = \langle aT, \phi\rangle = \langle T, a\phi\rangle \tag{3.12}$$

for any Schwartz function ϕ.

Definition 8 (Heat kernel and semigroup)**.** The function

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{|x|^2}{2t}} \tag{3.13}$$

is called *heat kernel* with parameter t, and it is the fundamental solution to the heat equation. The operator[2] acting as a convolution of the heat kernel with a (generalized) function, is called heat semigroup, and it is denoted by $(P_t)_{t\geq 0}$. For any $g \in \mathcal{S}$ and $t \geq 0$ we have

$$(P_t g)(y) = (p_t * g)(y) = \int_{\mathbb{R}} p_t(x) g(y-x) dx. \tag{3.14}$$

Note that the heat semigroup is defined for functions in $\mathcal{S}$, but we want to apply it on distributions in $\mathcal{S}'$. The extension of this operator is made by dual pairing. Now let $f \in \mathcal{S}'$, the map $P_t f$ is a linear map from $\mathcal{S}'$ to $\mathcal{S}'$ defined for any $\phi \in \mathcal{S}$ as $\phi \mapsto (P_t f)(\phi) := \langle f, P_t \phi \rangle$.

Applying the semigroup on a distribution will improve its regularity, this is seen by the so called *Schauder estimates*.

Lemma 9 (Schauder estimates)**.**

1. *For any $\xi \in \mathbb{R}$, $\zeta \geq 0$ and $t' > 0$, there is a constant $c_S^1 = c_S^1(\xi, \zeta, t')$ such that for any $f \in C^{\xi - 2\zeta}$*
$$\|P_t f\|_{C^\xi} \leq c_S^1 t^{-\zeta} \|f\|_{C^{\xi-2\zeta}}, \quad t \in (0, t']. \tag{3.15}$$

2. *For $\xi \in \mathbb{R}$, $\zeta \in (0,1)$ and $t' > 0$, there exists a constant $c_S^2 = c_S^2(\xi, \zeta, t')$ such that for $f \in C^{\xi+2\zeta}$*
$$\|P_t f - P_s f\|_{C^\xi} \leq c_S^2 (t-s)^\zeta \|f\|_{C^{\xi+2\zeta}}, \quad 0 \leq s < t \leq t', \tag{3.16}$$
in particular, when $s = 0$, we obtain
$$\|P_t f - f\|_{C^\xi} \leq c_S^2 t^\zeta \|f\|_{C^{\xi+2\zeta}} \tag{3.17}$$

Note that ξ and ζ in point 1 are unrelated to ξ and ζ in point 2.

For a proof of the above lemma see [24, Theorem 2.10] with $d = 0$ and $N = 1$.

Remark 10. In this work, we will take $t = 1/N$ for $N \geq 1$, so it is enough to consider the above estimates with $t' = 1$. We will therefore omit the dependence on t' when we refer to constants c_S^1 and c_S^2.

3.3 Some related PDEs

Related to the concept of solutions of the SDE there is a family of PDEs with useful bounds which are discussed in [25] and recalled here. We will use those bounds later to find the convergence rate of the numerical methods we implement. Firstly, the class of PDEs we will discuss is of the form

$$\begin{cases} u_t + \frac{1}{2} u_{xx} + b u_x = \lambda u + g, \text{ on } [0,T] \times \mathbb{R} \\ u(T) = u_T, \end{cases} \tag{3.18}$$

where g is a continuous function in time taking values in $C_T C^{-\beta}$, $u_T \in CD^{(1-\beta)-}$ and λ is a real number.

[2]Recall that semigroups are families of operators, we choose to call them operators as a shorthand.

Before stating the definitions we care about, one more thing to mention is the way the product bu_x is defined. Notice that this is potentially the product of two distributions, so let us define a meaning for it.

Definition 11 (Pointwise product of distributions)**.** Suppose that $f, g \in \mathcal{S}'$. Choose a function $\varphi \in \mathcal{S}$ such that $0 \le \varphi(x) \le 1$ for all $x \in \mathbb{R}$, and that for $|x| \le 1$ we have $\varphi(x) = 1$, and for $|x| \ge 3/2$ then $\varphi(x) = 0$. Let us consider the approximation $S^j h$ for $h \in \mathcal{S}'$ and $j \in \mathbb{N}$ defined by

$$S^j h(x) := \mathcal{F}^{-1}\Big[\varphi\Big(\frac{\xi}{2^j}\Big)\mathcal{F}[h]\Big](x).$$

The *pointwise product* fg is defined as

$$fg := \lim_{j\to\infty} S^j f S^j g \tag{3.19}$$

if the limit exists in $\mathcal{S}'$.

One can see that this definition is not operational unless we prove that the pointwise product exists in the space where we are working, so we recall the argument from [27, Section 2.2] where they claim that for $f \in C^\alpha$ and $g \in C^{-\beta}$ such that $\alpha - \beta > 0$ and $\alpha, \beta > 0$ then the pointwise product fg is an element of $C^{-\beta}$. Additionally, there exist a constant $c > 0$ such that

$$\|fg\|_{C^{-\beta}} \le c\|f\|_{C^\alpha}\|g\|_{C^{-\beta}}. \tag{3.20}$$

This property is termed *point-wise product* and is a consequence of the so called Bony paraproduct, for a detailed description see [20, Section 2.1].

The following two assumptions will stand for the PDE (3.18).

Assumption 12. $g \in C_T C^{(-\beta)+}$ and $u_T \in DC^{(1-\beta)+}$ for $\beta \in (0, 1/2)$.

Assumption 13. $g \in C_T C^{(-\beta)+}$ and $u_T \in C^{(2-\beta)+}$ for $\beta \in (0, 1/2)$.

For ease of notation, we define an operator $\mathcal{L}$ as follows[3]:

Definition 14. Let $b \in C_T^{1/2} C^{(-\beta)+}$ for $\beta \in (0, 1/2)$. The operator $\mathcal{L}$ is defined as

$$\mathcal{L} : \mathcal{D}^0_{\mathcal{L}} \to C_T\mathcal{S}',$$

which maps

$$f \mapsto \mathcal{L}f := \partial_t f + \frac{1}{2}\partial_{xx} f + b\partial_x f, \tag{3.21}$$

where the product $b\partial_x f$ is understood in the sense of (3.19) which exists because $\partial_x f \in C^\beta$ and $b \in C^{(-\beta)+}$ by (3.20) and the domain

$$\mathcal{D}^0_{\mathcal{L}} := C_T DC^\beta \cap C^1_T \mathcal{S}'$$

is contained in $\mathcal{D}_{\mathcal{L}}$.

[3]This operator can, and will, be defined on smaller domains where it is needed.

It is worth mentioning that [16, 25, 27] deal with equations in $\mathbb{R}^d$ whereas we work in $\mathbb{R}$. We define a mild solution of (3.18) in $C_T DC^\beta$ as follows:

Definition 15. Let $u \in C_T DC^\beta$ for $\beta \in (0, 1/2)$. We say that u is a mild solution of (3.18) if u satisfies

$$u(t) = P_{T-t}u_T + \int_t^T P_{s-t}(u_x(s)b(s))ds - \int_t^T P_{s-t}(\lambda u + g)ds \tag{3.22}$$

for all $t \in [0, T]$.

The product $u_x(s)b(s)$ above is understood as in (3.19) and well-defined thanks to (3.20). Existence and uniqueness of the mild solution to PDE (3.18) is proved in [25, Theorem 4.7]. We recall the theorem below.

Theorem 16 ([25, Theorem 4.7])**.** *Let $b \in C_T^{1/2} C^{-\beta}$ for $\beta \in (0, 1/2)$.*

1. *Let u_T and g satisfy Assumption 12. Then there exist a mild solution u to (3.18) in $C_T DC^{(1-\beta)-}$ which is unique in $C_T DC^\beta$.*

2. *Let u_T and g satisfy Assumption 13, and u_T be bounded. Then the unique mild solution u of (3.18) is also bounded and $u \in C_T C^{(2-\beta)-}$.*

In this theorem we can see that the regularity of the terminal condition to equation (3.18) will determine the regularity of the solution itself. More importantly, notice that under Assumption 12 we have that existence of a mild solution holds in a smaller space $C_T DC^{(1-\beta)-}$ and uniqueness in the larger. This is a result of the usage of Banach fixed point theorem on the map of the mild solution of (3.18), that is

$$\mathcal{T}u := P_{T-t}u_T + \int_t^T P_{s-t}(u_{xx}(s)b(s))ds - \int_t^T P_{s-t}(\lambda u_x(s) + g(s))ds. \tag{3.23}$$

The theorem proves that the map $\mathcal{T}$ is a contraction for any $\alpha \in (\beta, 1-\beta)$ in the space $C_T C^\alpha$ (point 1) and in $C_T DC^{1+\alpha}$ (point 2). Moreover, it is possible to see that it works also for $\alpha = \beta$, with a small adjustment in the bound.

In order to define a virtual solution to SDE (3.1), let us see a particular case of PDE (3.18), we set $g = -b$ and $u_T = 0$, which clearly satisfy Assumptions 12 and 13 and the equation reads

$$\begin{cases} \mathcal{L}v = \lambda v - b \\ v(T) = 0. \end{cases} \tag{3.24}$$

Since that equation is a special case of (3.18), its mild solution exists in $C_T C^{(2-\beta)-}$ and is unique in $C_T DC^\beta$. Moreover, by [25, Proposition 4.13] the solution v of (3.24) is bounded in (t, x), i.e. there is a constant c such that

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}} |v(t, x)| \le c. \tag{3.25}$$

Also, we choose λ such that for a constant $C(\beta, \epsilon)$ the following holds

$$\lambda \geq [C(\beta, \epsilon) \max\{\sup_N \|b^N\|_{C_T C^{-\beta+\epsilon}}, \|b\|_{C_T C^{-\beta+\epsilon}}\}]^{\frac{1}{1-\theta}}, \tag{3.26}$$

where[4] $b^N = p_{1/N} * b$. So we have

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}} |v_x(t,x)| \leq \frac{1}{2}. \tag{3.27}$$

In the following, instead of referring to equation (3.26) we may simply say that $\lambda > 0$ is *large enough*. By [25, Theorem 4.14] we have the following auxiliary function

$$\phi(t,x) := v(t,x) + x, \tag{3.28}$$

where v is the solution to (3.24). And by [25, Proposition 4.16] we have extra properties for ϕ that we will benefit from, namely

1. It holds that $\phi \in C^{0,1}$ and $\phi_x \in C_T C^{(1-\beta)-}$, and ϕ_x is uniformly bounded. So it follows that ϕ is Lipschitz with constant $K_\phi = \sup_{(t,x)} |\phi_x| = \sup_{(t,x)} |v_x + 1| \leq 3/2$.

2. For λ large enough, we have that $\phi(t,\cdot)$ is invertible for all $t \in [0,T]$ and its space inverse is denoted by

$$\psi(t,\cdot) := \phi^{-1}(t,\cdot). \tag{3.29}$$

   Additionally, $\psi \in C^{0,1}$ and $\psi_x C_T C^{(1-\beta)-}$ and ψ_x is uniformly bounded, so is Lipschitz with constant $K_\psi = 2$.

We additionally argue that ψ is 2-Lipschitz, which is proved as follows. By the definition of ψ as the space inverse of ϕ, we have $z = \psi(t,z) + u(t,\psi(t,z))$. Hence,

$$\begin{aligned} |\psi(t,y) - \psi(t,y')| &\leq |u(t,\psi(t,y)) - u(t,\psi(t,y'))| + |y - y'| \\ &\leq \frac{1}{2}|\psi(t,y) - \psi(t,y')| + |y - y'|, \end{aligned}$$

where the last inequality uses again that $u(t,\cdot)$ is $\frac{1}{2}$-Lipschitz.

3.4 Notions of solutions of the SDE

In this section we discuss the martingale problem associated to this SDE and its links to the concept of virtual solutions. We will see that within the setting of our problem—importantly, the fact that is in one dimension—the virtual solutions will be equivalent to strong solutions of the SDE.

The martingale problem with distributional drift

The celebrated martingale problem proposed by Stroock and Varadhan [45] arises as a natural formulation on the construction of diffusion processes. We will give a brief

[4] We talk more about this object in Section 4.2

explanation of the classical problem in order to motivate the formulation of the martingale problem with singular drift that [27] propose.

Notice that for coefficients σ and b, a diffusion process $(X_t)_{t\geq 0}$ can be defined as the solution of the SDE

$$X_t = X_0 + \int_0^t \sigma(s, X_s) dW_s + \int_0^t b(s, X_s) ds, \tag{3.30}$$

and Theorem 5.1.5 [45] shows that for such a process with σ, b uniformly bounded and Lipschitz continuous it is possible to define the associated family of probability measures $\{P_{s,x} : (s,x) \in [0,\infty) \times \mathbb{R}\}$ on a measurable space $(\Omega, \mathcal{M})$ which serve as the law of X_s, consider also a filtration $(\mathcal{M}_t)_{t\geq 0}$. We define

$$P(s, x; t, \Gamma) := P_{s,x}(X_t \in \Gamma)$$

for $0 \leq s \leq t$, $x \in \mathbb{R}$ and $\Gamma \in \mathcal{B}(\mathbb{R})$, then $(s,x) \to P(s,x;t,\Gamma)$ is measurable on $[0,t] \times \mathbb{R}$ and is a transition probability function, so we have a diffusion process. For a function $f \in C_0^\infty$ we apply Itô's formula to (3.30) and we have

$$f(X_t) = f(X_0) + \int_0^t \sigma(s, X_s) \frac{\partial f}{\partial x} dW_s + \int_0^t \left(b(s, X_s) \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2(s, X_s) \frac{\partial^2 f}{\partial x^2} \right) ds. \tag{3.31}$$

For ease of notation we introduce the operator

$$L_t = \frac{1}{2} a(t, \cdot) \frac{\partial^2}{\partial x^2} + b(t, \cdot) \frac{\partial}{\partial x}, \tag{3.32}$$

rewrite and rearrange (3.31) to get

$$f(X_t) + \int_0^t L_s f(X_s) ds = f(X_0) + \int_0^t \sigma(s, X_s) \frac{\partial f}{\partial x} dW_s. \tag{3.33}$$

We know that the right-hand side of (3.33) is a martingale, since we have a constant and a stochastic integral, hence one must prove that

$$f(X_t) - \int_s^t L_u f(X_u) du \tag{3.34}$$

is a $P_{s,x}$-martingale for all $f \in C_0^\infty(\mathbb{R})$.

Then the question is if we can characterize the diffusion process X with coefficients σ and b by a probability measure $\mathbb{P}$ such that (3.34) is martingale under $\mathbb{P}$, this is termed *the martingale problem for* $(L_t)_{t\geq 0}$. This statement is indeed true—in some cases—and if a solution can be found it will be the pair $(X, \mathbb{P})$. The proof is beyond the scope of this work, and can be found in Chapter 6 of the aforementioned work by Stroock and Varadhan. However, this idea has percolated in the area of diffusion processes in such a way that a common approach on the construction of such processes is to create martingale formulations of different types. Additionally, the notion of solution to a martingale problem is equivalent to that of a weak solution to an SDE, which makes it a strong tool on the analysis of the dynamics of SDEs.

For our case, let us recall the family of SDEs we are concerned about

$$X_t = X_0 + \int_0^t b(s, X_s)ds + W_t \tag{3.35}$$

where $X_0 \sim \mu$. With the operator $\mathcal{L}$ from (3.21), see that if b is a function then $(X, \mathbb{P})$ is a solution to the classical Stroock-Varadhan martingale problem with respect to $\mathcal{L}$ if for every $f \in C_0^\infty$

$$f(X_t) - f(X_0) - \int_0^t \Big(\frac{1}{2}\partial_{xx} f(X_s) + \partial_x f(X_s) b(s, X_s)\Big)ds$$

is a local martingale.

Notice that we say that the initial condition is a random variable, we will sometimes assume that this is distributed as Dirac delta centred on[5] $x \in \mathbb{R}$—effectively having a real number as initial condition—but for the following discussion we can have it in this general form.

The domain $\mathcal{D}_\mathcal{L} \subseteq \mathcal{D}_\mathcal{L}^0$ proposed by Issoglio and Russo for their martingale formulation—see [27, Section 3]—is

$$\begin{aligned} \mathcal{D}_\mathcal{L} :=\{ & f \in C_T C^{(1+\beta)+} : \exists g \in C_T \bar{C}_c^{0+} \text{ such that} \\ & f \text{ is a mild solution of } \mathcal{L}f = g \text{ and } f(T) \in \bar{C}_c^{(1+\beta)+}\} \end{aligned} \tag{3.36}$$

where the operator $\mathcal{L}$ the defined by (3.21). Using $\mathcal{D}_\mathcal{L}$ we can define a martingale problem for SDE (3.35) with distributional drift as follows.

Definition 17. [[27, Definition 3.1]] Let X be a continuous process indexed by $t \in [0, T]$ and $\mathbb{P}$ a probability measure on some measurable space, we say that a couple $(X, \mathbb{P})$ is a *solution to the martingale problem with distributional drift b and initial condition μ* if and only if for every $f \in \mathcal{D}_\mathcal{L}$ the quantity

$$f(t, X_t) - f(0, X_0) - \int_0^t \mathcal{L}f(s, X_s)ds \tag{3.37}$$

is a local martingale under $\mathbb{P}$, and $X_0 \sim \mu$ under $\mathbb{P}$, where the domain $\mathcal{D}_\mathcal{L}$ is defined by (3.36) and the operator $\mathcal{L}$ is from Definition (3.21).

Issoglio and Russo [27] say that uniqueness holds for the martingale problem if for any two solutions $(X^1, \mathbb{P}^1)$ and $(X^2, \mathbb{P}^2)$ with $X^1 \sim \mu$ and $X^2 \sim \mu$, the law of X^1 under $\mathbb{P}^1$ is the same as the law of X^2 under $\mathbb{P}^2$. Moreover, in [27, Lemma 4.1] they also show that when $b \in C_t C^{0+}$ the notion of martingale problem with distributional drift, the classical martingale problem, and the weak solution of the SDE (3.44) are equivalent. And they conclude that existence and uniqueness of the solution to the martingale problem with distributional drift in [27, Theorem 4.5] and in the end saying that by setting a certain notion of mild solution to the SDE (3.35), so called there is an equivalence in the notion of

[5]For instance in Chapter 4.

the martingale problem with distributional drift and this solution.

Virtual solutions

The concept of virtual solutions to the class of SDEs (3.1) is introduced in [16]. This idea presents several advantages for us, one of them is that it allows to replace the integral $\int_0^t b(s, X_s)ds$ for something more manageable, but additionally for a classical SDE the virtual solution is also a strong solution. Let us mention that in order to define a virtual solution for our setting we cannot rely on Itô's formula as we explain for the classical setting, instead its definition is heavily based on a class PDEs discussed on Section 3.3.

Before diving into the technical details let us provide some motivation for it. Consider PDE (3.24), but instead of Assumptions 12 and 13, suppose that b has better regularity, in particular $b \in C_T C_b^1$. In this case, the solution to the PDE, v, is at least of class $C^{1,2}$, so we can apply Itô's formula to $v(t, X_t)$:

$$v(t, X_t) = v(0, X_0) + \int_0^t \Big(v_t + \frac{1}{2}v_{xx} + bv_x\Big)(s, X_s)ds + \int_0^t v_x(s, X_s)dW_s, \tag{3.38}$$

by PDE (3.24) we can rewrite the above as

$$v(t, X_t) = v(0, X_0) + \int_0^t (\lambda v - b)(s, X_s)ds + \int_0^t v_x(s, X_s)dW_s. \tag{3.39}$$

Rearranging this equation, the drift can be equivalently rewritten as

$$\int_0^t b(s, X_s)ds = v(0, X_0) - v(t, X_t) + \lambda \int_0^t v(s, X_s)ds + \int_0^t v_x(s, X_s)dW_s, \tag{3.40}$$

which we can plug into the SDE (3.1)—with the relaxed assumption of the drift—to get

$$X_t = X_0 + v(0, X_0) - v(t, X_t) + \lambda \int_0^t v(s, X_s)ds + \int_0^t (v_x(s, X_s) + 1)dW_s. \tag{3.41}$$

Note then that this is nothing but algebraic manipulation of the equation, and we can see that the classical solution can also be expressed as in (3.41). Not only that, but in the process we have successfully transformed our SDE by removing the irregularity of the drift. Since in our setting we have a highly irregular drift, we want to find a way to transform our SDE similarly.

We are now going to enforce once again our original assumptions on the problem and use the prototype of a virtual solution (3.41) to define it in our case. We say that a stochastic basis is a quintuple $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, W)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space, together with a filtration $\mathbb{F} = \{\mathcal{F}_t : 0 \le t \le T\}$ and a Brownian motion $W = (W_t)_{t\in[0,T]}$.

Definition 18 (Virtual solution [16, Definition 25])**.** Given $x \in \mathbb{R}$, a *virtual solution* to (3.1) is a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, W)$ and an $\mathbb{F}$-adapted process $X := (X_t)_{t\in[0,T]}$ such that

$$X_t = x + v(0, x) - v(t, X_t) + \lambda \int_0^t v(s, X_s)ds + \int_0^t (v_x(s, X_s) + 1)\, dW_s, \tag{3.42}$$

holds $\mathbb{P}$-almost surely for all $t \in [0,T]$. Where v is the unique mild solution to (3.24).

As noted before this construction jumps the hurdle of defining the ill-posed integral $\int_0^t b(s,X_s)ds$, and now we can go ahead and prove that such a solution is meaningful for our setting.

In order to construct an SDE that we can handle from the SDE with distributional drift (3.1) let us assume that $(X,\mathbb{F})$ is a virtual solution of said SDE with initial condition $X_0 = x$. Note that in (3.42) we can add $v(t,X_t)$ on both sides and recall the definition of $\phi(t,X_t)$, so we get

$$\phi(t,X_t) = X_t + v(t,X_t) = x + v(0,x) + \lambda \int_0^t v(s,X_s)ds + \int_0^t (v_x(s,X_s)+1)\,dW_s, \tag{3.43}$$

we set the process $Y_t := \phi(t,X_t)$. Since $\phi(t,\cdot)$ is invertible for all $t \in [0,T]$ it is clear that $X_t = \psi(t,Y_t) = \psi(t,\phi(t,X_t))$ and $Y_0 := x + v(0,x)$, we have a classical SDE for Y with well-behaved coefficients, i.e.

$$Y_t = Y_0 + \lambda \int_0^t v(s,\psi(s,Y_s))ds + \int_0^t (v_x(s,\psi(s,Y_s))+1)\,dW_s, \tag{3.44}$$

If Y_0 is a deterministic initial condition then (3.44) is a classical SDE whose drift and diffusion are bounded and measurable, and it admits a classical weak solution, c.f. [27, Remark 3.13] and [45, Theorem 10.1.2]. Conversely, if $(Y,\mathbb{F})$ is a solution of (3.44), then $(X,\mathbb{F})$ with $X_t = \psi(t,Y_t)$ is a virtual solution of (3.1) with distributional drift and initial condition $x = \psi(0,y)$, see [16, Theorem 28]. Beyond seeing that the virtual solution coincides with the classical solution for certain types of drift, one wishes to find results on existence and uniqueness of the virtual solution. This is done by Flandoli et al. [16] in two steps, first, it is proved that (3.44) has a unique weak solution, and then this is translated into a result for (3.42) or (3.1).

Proposition 19 ([16, Proposition 27])**.** *For every initial condition $y \in \mathbb{R}$ there exists a unique in law weak solution $(Y,\mathbb{F})$ to the SDE (3.44) with initial value $y = x + v(0,x)$.*

Then given the transformation ψ, it is possible to find a unique virtual solution to SDE (3.1).

Theorem 20 ([16, Theorem 28])**.** *For every $x \in \mathbb{R}$ there exists a unique in law virtual solution $(X,\mathbb{F})$ to the SDE (3.1) with initial value, $X_0 = x$. The solution is given by $X_t = \psi(t,Y_t)$ for $t \in [0,T]$, where $(Y,\mathbb{F})$ is the solution with initial value $y = x + v(0,x)$ given in Proposition 19.*

3.5 Link between the martingale problem and virtual solutions

In section 3.4 we discussed two types of solutions to SDE (3.1). Note that both notions of solutions share an auxiliary process, denoted respectively[6] by (3.44) which, on attaining a solution, will guarantee that we can find either a solution to the martingale problem with distributional drift or a virtual solution for X.

Here we want to emphasize the relationship between the solutions to the martingale problem with distributional drift and virtual solutions. Specifically because the solution to the martingale problem does not give us the dynamics of the solution, but the virtual solution formulation does, and we want to have information about the dynamics to create the strong error estimates we are after. So, notice that we claim that solving equation (3.44) will provide us with solutions to the martingale problem and virtual solutions which holds by noting that Proposition 19 gives us weak solutions $(Y, \mathbb{P})$. Additionally, in Section 3.4 it is discussed that $(Y, \mathbb{P})$ is a solution to the classical martingale problem. And since we know that weak solutions and solutions to the classical martingale problem are equivalent we see that the solution to the martingale problem with distributional drift possesses the dynamics of the virtual solution, this is proved in [27, Corollary 6.13]

Now that we have the link between the different notion of solutions we see that our one dimensional SDE (3.1) has strong unique solutions, encoded in the next result.

Lemma 21. *Let $b \in C_T C^{(-\beta)+}$. Then there exist a process $(Y_t)_{t\geq 0}$ with a Brownian motion (W_t) in the original probability space such that it satisfies (3.44), and the process is unique up to modifications. Similarly, there is a unique process $(X_t)_{t\geq 0}$ that satisfies (3.42)*

Proof. Recall that v and v_x are bounded, therefore they have linear growth uniformly in time, which implies existence of a strong solution Y to (3.44) by [4, Remark 9.6]. The uniqueness of the solution comes again from the boundedness of the coefficients of (3.44) and the usage of a result of Yamada and Watanabe—see [28, Proposition 5.2.13]. v_x is bounded which implies Lipschitz continuity of v, moreover since $|v_x + 1| \leq 3/2$, we can select a strictly increasing function $h(v) = \frac{3}{2}v^\alpha$ for any $\alpha \geq 1/2$ that satisfies the result mentioned, so we see that the diffusion coefficient is α-Hölder continuous therefore strong uniqueness hold for Y. Recall that $Y(t,x) = \phi(t,x)$, whose space inverse is $\psi(t,\cdot)$, therefore there is a one-to-one correspondence between Y and X. Thus, since strong uniqueness hold for Y it does for the virtual solution of (3.1) X. □

One potential pushback to the usage of virtual solution is the apparent dependence of the solution on the parameter λ. This problem, however, is sorted by Flandoli et al. in [16, Proposition 29], where they express the virtual solution as the limit of the solutions of a sequence of SDEs. They note that with a smooth drift b^N which converges to b in the appropriate space, the strong solution of the SDEs with drift b^N converge in law to the virtual solution $(X, \mathbb{P})$ of (3.1). Thus, since the solution X^N with drift b^N does not depend

[6] Despite the slightly different notation, those two SDEs are exactly the same, we can observe that by the definition $\phi(t,x) = u(t,x) + x$.

on λ neither does its limit X. Although they work in the setting of Bessel potential spaces, their result holds as well for the Hölder-Zygmund spaces we work, the reader is referred to [27, Theorem 4.3] for a proof of this claim.

Chapter 4

SDEs with irregular drift

In this chapter we present the material related to the manuscript [9], which is concerned with the numerical approximation of a one-dimensional SDE with unit diffusion and a drift which is a generalized function. In particular, the drift takes values on a Hölder-Zygmund space of negative order. We design an Euler-Maruyama scheme and prove its convergence, along with an upper bound for the strong L^1 error.

4.1 The SDE and some motivation

We are interested in an SDE with the form of (3.1), which is

$$X_t = x + \int_0^t b(s, X_s)ds + W_t. \tag{4.1}$$

where $x \in \mathbb{R}$, $(W_t)_{t\geq 0}$ is a standard one-dimensional Brownian motion, and $b(t, \cdot) \in C^{-\gamma}$ for some $\gamma > 0$. In particular, the space where the SDE takes values is encoded in the following standing assumption.

Assumption 22. Let $b \in C_T^{1/2} C^{(-\beta)+}$ for $\beta \in (0, 1/2)$, this is an object which is $\frac{1}{2}$-Hölder in time and takes values in the space of distributions $C^{-\beta'}$ for some β' such that $\beta' < \beta$.

Working with this SDE, the first step is to develop a meaning for the term $\int_0^t b(s, X_s)ds$. In this work we presented this task on Chapter 3, but it was first developed by Flandoli et al. [16] and De Angelis et al. [14], albeit in a different setting where the drift belongs to a negative Sobolev space. More directly related to our subject are two works by Issoglio and Russo [27, 25] which share the same setting we have and deal with this kind of equations in the multidimensional case.

Having a meaning for the drift of this SDE is enough for the theoretical considerations which are the basis of the present work, however for the numerical analysis we need some additional steps. Particularly, notice that we cannot approximate b directly through a numerical method because it is a distribution. Instead we need some approximation b^N that converges to b in an appropriate space as $N \to \infty$, and then we can approximate an SDE with drift b^N.

4.2 Intuition for the numerical scheme

The idea behind the numerical scheme we implement serves as a good motivation for the theoretical work. The numerical scheme is based on two subsequent approximations. The first one replaces the distributional drift with a sequence of functional drifts that converge to the distributional one, let us call the elements of that sequence b^N, so we have that as $N \to \infty$ it holds that $b^N \to b$ in $C_T C^{-\beta}$. In finding such a sequence for the drift we will have a sequence of regularized SDEs with solutions X^N. The second step is to solve numerically the approximating SDEs, since these have a smoother drift and can be simulated using an Euler-Maruyama scheme. We will then balance errors coming from the approximation of the drift and the discretization of time to maximize the rate of convergence.

We obtain the sequence b^N by a regularization process on the drift, this is done applying heat the semigroup (see Definition 8) with parameter $t = 1/N$ for a real number $N > 0$, we define the regularized drift

$$b^N = P_{\frac{1}{N}} b = p_{\frac{1}{N}} * b. \tag{4.2}$$

Using Schauder estimates (Lemma 9) we can see that $b^N(t, \cdot) \in C^\gamma$ for any $\gamma > 0$ and for all $t \in [0, T]$, we have $b^N(t, \cdot) \in C_b^1, t \in [0, T]$, hence, it is Lipschitz continuous in the spatial variable. Therefore, the SDE

$$X_t^N = x + \int_0^t b^N(s, X_s^N) ds + W_t \tag{4.3}$$

has a unique strong solution in the classical sense.

For this sequence of *regularized SDEs* we can venture an approximation by the usual methods, in particular we are interested in using the simplest of all, the Euler scheme. The Milstein scheme, which is arguably the second simplest of the iterative methods of solutions for SDEs, is irrelevant since we have a unit diffusion, so we are back to the Euler scheme. Alternatively, it would be fair to conjecture that our scheme is in fact a Milstein scheme, and we should aim for the convergence rate to be equal to one. However, the fact that we are making two subsequent approximations will add up to the error which will be reflected both in the theoretical and the numerical error. Additionally, looking for a $1/2$ rate of convergence is supported by the work of multiple other authors which use irregular, but still functional, drifts such as [13] where the authors find a rate of convergence of $1/2$. In practice it is rare to consider any schemes of order greater than 1 (Milstein scheme), so we content ourselves studying the Euler method for this class of SDEs.

Keeping the usual approach for the Euler scheme, for $m \in \mathbb{N}$ we take an equally spaced partition $t_k = t_k^m = kT/m$, $k = 0, \dots, m$, of the interval $[0, T]$. We define

$$k(t) = k^m(t) = \max\{k : t_k \leq t\}, \tag{4.4}$$

for all $t \in [0, T]$, and consider an Euler-Maruyama approximation of (X^N) with m time steps[1]

$$X_t^{N,m} = x + \int_0^t b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m})ds + W_t \tag{4.5}$$

for all $t \in [0,T]$. In order to verify that these numerical approximations converge we first obtain a bound on the strong error between the approximation $(X_t^{N,m})_{t\geq 0}$ and the process $(X_t^N)_{t\geq 0}$ with explicit dependence of constants on the properties of the drift b^N. This is needed in order to balance the smoothing via choice of N and the number of time steps m to optimize the convergence rate of the Euler-Maruyama approximation to the true solution $(X_t)_{t\geq 0}$.

By having these two sequential approximations we can split the actual problem that we want to solve, which is to find a bound for the error of the numerical approximation (4.5) against the real solution of the SDE (4.1), into two problems by using triangle inequality

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t^{N,m} - X_t|] \leq \mathbb{E}[|X_t^{N,m} - X_t^N|] + \mathbb{E}[|X_t^N - X_t|].$$

4.3 Numerical approximation of the regularized SDE

We first want to find a bound on L^1 between the numerical approximation and the regularized solution, i.e. $\mathbb{E}[|X_t^{N,m} - X_t^N|]$. For this purpose we will use a similar procedure to the way we found virtual solutions to SDEs in Chapter 3 but now for the process X^N, then we will find a bound for some auxiliary stochastic processes. Let us introduce the following PDE

$$\begin{cases} \phi_t^N + \frac{1}{2}\phi_{xx}^N + b^N\phi_x^N = \lambda(\phi^N - 1) \\ \phi^N(T) = 0, \end{cases} \tag{4.6}$$

which has a solution ϕ^N for which ϕ_x^N exists and is bounded, therefore ϕ^N is Lipschitz. Note that the left-hand side can be written with the operator

$$\mathcal{L}^N f := f_t + \frac{1}{2}f_{xx} + b^N f_x. \tag{4.7}$$

This operator will be used to shorten notation in some places. Similarly as in the case of the auxiliary functions ϕ and ψ in Chapter 3, we define the space inverse of ϕ^N as

$$\psi^N := (\phi^N)^{-1}(t,\cdot), \tag{4.8}$$

and then we obtain the following processes as a result of the transformation ϕ^N applied on the solution to the regularized SDE X^N and to the Euler-Maruyama approximation $X^{N,m}$ from (4.5)

$$Y_t^N := \phi^N(t, X_t^N) \quad \text{and} \quad Y_t^{N,m} := \phi^N(t, X_t^{N,m}).$$

By the definition of ψ^N we see that

$$X_t^N := \psi^N(t, Y_t^N) \quad \text{and} \quad X_t^{N,m} := \psi^N(t, Y_t^{N,m}),$$

[1]Alternatively one can think of a step size $\Delta t = T/m$.

and since ψ^N is 2-Lipschitz (c.f. 3.3) instead of looking for a bound of $\mathbb{E}[|X_t^{N,m} - X_t^N|]$ we will look for one of $\mathbb{E}[|Y_t^{N,m} - Y_t^N|]$. This is a convenient trade of tasks since we will see that $Y_t^{N,m}$ and Y_t^N are processes with smooth coefficients.

First of all, to be convinced that the processes $Y_t^{N,m}$ and Y_t^N are indeed more regular we will find their dynamics applying Itô's formula to the definition we have just provided, recalling that $\phi^N(t,x) = x + u^N(t,x)$. So, let us start with Y_t^N. Note that b^N is in C^1, and therefore $u^N \in C^{1,2}$—c.f. [27, Theorem A.3]—which makes it a classical solution of

$$\begin{cases} u_t^N + \frac{1}{2}u_{xx}^N + b^N u_x^N = \lambda u^N - b^N \\ u^N(T) = 0, \end{cases} \tag{4.9}$$

which we can note is the equivalent equation to (3.24) with $\mathcal{L}^N$ from (4.7) instead of $\mathcal{L}$.

Since $b^N \to b$ in $C_T C^{-\hat{\beta}}$ by Lemma 9 and Assumption 22, we can apply [25, Lemma 4.19]. This lemma and its proof provide key properties of the above system and its relation to (4.9) and (3.44). The solution of the regularised Kolmogorov equation (4.9) enjoys a bound $\|u_x^N\|_\infty < 1/2$ for λ large enough; the constant λ can be chosen independently of N (but it depends on the drift b of the original SDE), and from now on we fix such λ. We also have $u^N \to u$ and $u_x^N \to u_x$ uniformly on $[0,T] \times \mathbb{R}^d$.

In order to find the dynamics of the processes $(Y_t^N)_{t\geq 0}$ and $(Y_t^{N,m})_{t\geq 0}$ we can use similar arguments to those we used when motivating the use of virtual solutions in Section 3.4. Just that in this case, these procedure is indeed possible and not just a motivation as we have mentioned before in the same section. Start applying Itô formula on $u^N(t, X_t^N)$ which yields

$$u^N(t, X_t^N) = u^N(0,x) + \int_0^t \left(u_t^N + \frac{1}{2}u_{xx}^N + b^N u_x^N\right)(s, X_s^N)ds + \int_0^t u_x^N(s, X_s)dW_s.$$

Note that the drift of this diffusion is the left hand side of (4.9), so we can substitute it as

$$u^N(t, X_t^N) = u^N(0,x) + \int_0^t (\lambda u^N - b^N)(s, X_s^N)ds + \int_0^t u_x^N(s, X_s)dW_s.$$

Using the definition of ϕ^N we have

$$\begin{aligned} Y_t^N &= \phi^N(t, X_t^N) = X_t^N + u(t, X_t^N) \\ &= u^N(0,x) + \int_0^t (\lambda u^N - b^N)(s, X_s^N)ds + \int_0^t u_x^N(s, X_s^N)dW_s \\ &\quad + x + \int_0^t b^N(s, X_s^N)ds + \int_0^t dW_s \\ &= u^N(0,x) + x + \int_0^t \lambda u^N(s, X_s^N)ds + \int_0^t (u_x^N(s, X_s^N) + 1)dW_s \end{aligned}$$

so we can see that

$$Y_t^N = \phi^N(0,x) + \lambda \int_0^t u^N(s, X_s^N)ds + \int_0^t (u_x^N(s, X_s^N) + 1)dW_s, \tag{4.10}$$

and using the space inverse of ϕ we recover the dynamics as

$$Y_t^N = \phi^N(0,\psi(0,y)) + \lambda \int_0^t u^N(s,\psi(t,Y_s^N))ds + \int_0^t (u_x^N(s,\psi(t,X_s^N)) + 1)dW_s. \quad (4.11)$$

In order to derive the dynamics of the numerical approximation $Y_t^{N,m}$ we have to consider that although we are using the same functional drift b^N, the actual drift that is used is slightly different because of where it is evaluated at each point. Recall that for the numerical scheme we are using a discretization of the time interval and the drift is actually $b^N\big(t_{k(s)}, X_{t_{k(s)}}^{N,m}\big)$ for all $s \in [0,t]$. Applying Itô formula on $u^N\big(t, X_t^{N,m}\big)$ we have

$$\begin{aligned} u^N(t, X_t^{N,m}) &= u^N(0,x) \\ &\quad + \int_0^t [(u_t^N + \frac{1}{2}u_{xx}^N)(s, X_s^{N,m}) + b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m})u_x^N(s, X_s^{N,m})]ds \\ &\quad + \int_0^t u_x^N(s, X_s^{N,m})dW_s. \end{aligned}$$

This time we will again use the PDE (3.24) to substitute, albeit differently, namely

$$\begin{aligned} u^N(t, X_t^{N,m}) &= u^N(0,x) \\ &\quad + \int_0^t [(\lambda u^N - b^N - b^N u_x^N)(s, X_s^{N,m}) + b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m})u_x^N(s, X_s^{N,m})]ds \\ &\quad + \int_0^t u_x^N(s, X_s^{N,m})dW_s \\ &= u^N(0,x) \\ &\quad + \int_0^t (\lambda u^N - b^N)(s, X_s^{N,m})ds + \int_0^t u_x^N(s, X_s^{N,m})dW_s \\ &\quad + \int_0^t u_x^N(s, X_s^{N,m})(b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m}))ds. \end{aligned}$$

We denote the last term as

$$E_t^{1,m} = \int_0^t u_x^N(s, X_s^{N,m})(b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m}))ds. \quad (4.12)$$

We can now use the definition of ϕ^N and $X_t^{N,m}$ to get the dynamics of $Y_t^{N,m}$

$$
\begin{aligned}
Y_t^{N,m} &= \phi^N(t, X_t^{N,m}) = X_t^{N,m} + u^N(t, X_t^{N,m}) \\
&= x + \int_0^t b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m})ds + \int_0^t dW_s \\
&\quad + u^N(0,x) + \int_0^t (\lambda u^N - b^N)(s, X_s^{N,m})ds + \int_0^t u_x^N(s, X_s^{N,m})dW_s + E_t^{1,m} \\
&= x + u^N(0,x) + \int_0^t \lambda u^N(s, X_s^{N,m})ds + \int_0^t (u_x^N(s, X_s^{N,m}) + 1)dW_s \\
&\quad + \int_0^t [b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m})]ds + E_t^{1,m} \\
&= x + u^N(0,x) + \int_0^t \lambda u^N(s, X_s^{N,m})ds + \int_0^t (u_x^N(s, X_s^{N,m}) + 1)dW_s + E_t^{2,m} + E_t^{1,m},
\end{aligned}
$$

where the term $E_t^{2,m}$ is the process

$$
E_t^{2,m} = \int_0^t [b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m})]ds.
$$

We define the error term $E_t^m = E_t^{1,m} + E_t^{2,m}$, so we have that $Y_t^{N,m}$ will be

$$
Y_t^{N,m} = x + u^N(0,x) + \lambda \int_0^t u^N(s, X_s^{N,m})ds + \int_0^t (u_x^N(s, X_s^{N,m}) + 1)dW_s + E_t^m. \tag{4.13}
$$

For convenience let us write the difference $Y_t^{N,m} - Y_t^N$ separately so we can refer to it directly

$$
\begin{aligned}
Y_t^{N,m} - Y_t^N = \lambda &\int_0^t (u^N(s, X_s^{N,m}) - u^N(s, X_s^N))ds \\
+ &\int_0^t (u_x^N(s, X_s^{N,m}) - u_x^N(s, X_s^N))dW_s + E_t^m.
\end{aligned} \tag{4.14}
$$

Definition 23. Let ξ_s be a process of bounded variation, if $Z_s := \int_0^t \xi_s ds$ then

$$
|Z_s| = \int_0^t |\xi_s| ds. \tag{4.15}
$$

The process E_t^m defined above will agree with Definition 23 and we can find a bound for it in the following lemma.

Lemma 24. *Let E_t^m be*

$$
E_t^m = E_t^{1,m} + E_t^{2,m} = \int_0^t \left[b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m})\right]\left[u_x^N(s, X_s^{N,m}) + 1\right]ds \tag{4.16}
$$

Then the following bound holds

$$
\mathbb{E}|E_t^m| \leq C_1(N)m^{-1} + C_2(N)m^{-1/2}, \tag{4.17}
$$

where

$$C_1(N) = \|b^N\|_{L_T^\infty L^\infty}\big(9T + \frac{3}{4}\|b_x^N\|_{L_T^\infty L^\infty}T^2\big)$$
$$C_2(N) = T^{3/2}(\|b_x^N\|_{L_T^\infty L^\infty} + [b^N]_{1/2,L^\infty})$$

Proof. From the definition of E_t^m, Jensen's inequality and since (3.27) tells us that $|u_x| \leq 1/2$, we have

$$\mathbb{E}|E_t^m| \leq \frac{3}{2}\mathbb{E}\int_0^t \big|b^N(t_{k(s)}, X_{t_{k(s)}}^{N,m}) - b^N(s, X_s^{N,m})\big|ds. \tag{4.18}$$

We can break the integral into the integrals between the points of the discretized grid which gives

$$\begin{aligned}\mathbb{E}|E_t^m| \leq \frac{3}{2}\mathbb{E}\Big[\int_0^{t_1} &\big|b^N(t_0, X_0^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big] \\ &+ \frac{3}{2}\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_{t_k}^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big] \\ &+ \frac{3}{2}\mathbb{E}\Big[\int_{t_{k(t)}}^t \big|b^N(t_{k(t)}, X_{t_{k(t)}}^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big],\end{aligned} \tag{4.19}$$

Note that it is true that $0 \leq t - t_{k(s)} \leq T/m$ for all $t \in [0,T]$, so the first and last terms above are bounded each by $3\frac{T}{m}\|b^N\|_{L_t^\infty L^\infty}$. Further for the second term in (4.19) we add and subtract $b^N(t_k, X_s^{N,m})$ and then split by triangle inequality and to get

$$\begin{aligned}&\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_{t_k}^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big] \\ &\quad\leq \frac{3}{2}\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_{t_k}^{N,m}) - b^N(t_k, X_s^{N,m})\big|ds\Big] \\ &\quad\quad+ \frac{3}{2}\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_s^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big]\end{aligned} \tag{4.20}$$

For the first term in the right hand side of (4.20) use the mean value theorem and the linearity of expectation to get

$$\begin{aligned}&\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_{t_k}^{N,m}) - b^N(t_k, X_s^{N,m})\big|ds\Big] \\ &\quad= \mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \|b_x^N\|_{L_T^\infty L^\infty}\big|X_{t_k}^{N,m} - X_s^{N,m}\big|ds\Big] \\ &\quad\leq \|b_x^N\|_{L_T^\infty L^\infty}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \mathbb{E}\big|X_{t_k}^{N,m} - X_s^{N,m}\big|ds\Big].\end{aligned} \tag{4.21}$$

Then the integral term inside the sum can be estimated using the definition of the numerical approximation (4.5) as

$$\int_{t_k}^{t_{k+1}} \mathbb{E}|X_{t_k}^{N,m} - X_s^{N,m}|ds = \mathbb{E}\big|(s-t_k)b^N(t_k, X_{t_k}^{N,m}) + (W_s - W_{t_k})\big|, \tag{4.22}$$

since $t_{k+1} - t_k = T/m$ is small. We can then bound b^N by its sup norm as

$$\int_{t_k}^{t_{k+1}} \mathbb{E}\big|X_{t_k}^{N,m} - X_s^{N,m}\big|ds \leq \int_{t_k}^{t_{k+1}} \mathbb{E}\big|(s-t_k)\|b^N\|_{L_T^\infty L^\infty}\big|ds + \int_{t_k}^{t_{k+1}} \mathbb{E}\big|W_s - W_{t_k}\big|ds, \tag{4.23}$$

then the first term on the right hand side is deterministic so its expectation is itself and we can integrate with respect to s, for the second we use the properties of the Brownian motion and again integrate which yields.

$$\begin{aligned}\int_{t_k}^{t_{k+1}} \mathbb{E}\big|X_{t_k}^{N,m} - X_s^{N,m}\big|ds &\leq \frac{1}{2}\|b^N\|_{L_T^\infty L^\infty}(t_{k+1}-t_k)^2 + \frac{2}{3}\mathbb{E}[W_1](t_{k+1}-t_k)^{3/2} \\ &= \frac{1}{2}\|b^N\|_{L_T^\infty L^\infty}\Big(\frac{T}{m}\Big)^2 + \frac{2}{3}\Big(\frac{T}{m}\Big)^{3/2}.\end{aligned} \tag{4.24}$$

We plug (4.24) into (4.21) and note that there are at most m terms in the sum to get

$$\begin{aligned}&\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_{t_k}^{N,m}) - b^N(t_k, X_s^{N,m})\big|ds\Big] \\ &\leq \|b_x^N\|_{L_T^\infty L^\infty} m\Big[\frac{1}{2}\|b^N\|_{L_T^\infty L^\infty}\Big(\frac{T}{m}\Big)^2 + \frac{2}{3}\Big(\frac{T}{m}\Big)^{3/2}\Big] \\ &= \|b_x^N\|_{L_T^\infty L^\infty}\Big[\frac{1}{2}\|b^N\|_{L_T^\infty L^\infty}T^2m^{-1} + \frac{2}{3}T^{3/2}m^{-1/2}\Big]\end{aligned} \tag{4.25}$$

Then the second term of (4.20) by linearity of expectation and bounding the difference of b^N only with the sup norm in space is

$$\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \big|b^N(t_k, X_s^{N,m}) - b^N(s, X_s^{N,m})\big|ds\Big] \leq \sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} \mathbb{E}\|b^N(t_k) - b^N(s)\|_{L^\infty}ds, \tag{4.26}$$

using the 1/2-Hölder seminorm of b^N, integrating and noting that there are at most m terms in the sum, and that $t_{k+1} - tk = T/m$ gives

$$\begin{aligned}&\mathbb{E}\Big[\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} |b^N(t_k, X_s^{N,m}) - b^N(s, X_s^{N,m})|ds\Big] \\ &\leq \Big(\sum_{k=1}^{k(t)-1}\int_{t_k}^{t_{k+1}} (s-t_k)^{1/2}ds\Big)[b^N]_{1/2,L^\infty} \\ &\leq \frac{2}{3}m(t_{k+1}-t_k)^{3/2}[b^N]_{1/2,L^\infty} \\ &\leq \frac{2}{3}[b^N]_{1/2,L^\infty}T^{3/2}m^{-1/2}\end{aligned} \tag{4.27}$$

Putting together the bound for the first and last terms of (4.19) plus (4.25) and (4.27) we have the bound

$$\mathbb{E}|E_t^m| \le \|b^N\|_{L_T^\infty L^\infty}\big(9T + \frac{3}{4}\|b_x^N\|_{L_T^\infty L^\infty}T^2\big)m^{-1} + T^{3/2}\big(\|b_x^N\|_{L_T^\infty L^\infty} + [b^N]_{1/2,L^\infty}\big)m^{-1/2} \tag{4.28}$$

as required. □

We will need to bound the L^1 norm of a certain process, which will turn out to depend on bounding the local time at zero of a related process. To this end, we recall the definition of a local time of a continuous semi-martingale Z at time 0 as

$$L_t^0(Z) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t \mathbb{1}_{\{|Z|\le\epsilon\}}d\langle Z\rangle_s, \tag{4.29}$$

$\mathbb{P}$-a.s. for all $t \ge 0$. Then we have a bound for this local time established in [14].

Lemma 25 ([14, Lemma 5.1])**.** *For any $\epsilon \in (0,1)$ and any real-valued, continuous semi-martingale Z we have*

$$\begin{aligned}\mathbb{E}[L_t^0(Z)] \le 4\epsilon - 2\mathbb{E}\left[\int_0^t \left(\mathbb{1}_{\{Z_s\in(0,\epsilon)\}} + \mathbb{1}_{\{Z_s\ge\epsilon\}}e^{1-Z_s/\epsilon}\right)dZ_s\right] \\ + \frac{1}{\epsilon}\mathbb{E}\left[\int_0^t \mathbb{1}_{\{Z_s>\epsilon\}}e^{1-Z_s/\epsilon}d\langle Z\rangle_s\right]\end{aligned}. \tag{4.30}$$

Conveniently, we wish to estimate the absolute value of a difference of semi-martingales, so we can use one of the so called *Itô-Tanaka-Meyer* formulas (c.f. [28, Section 3.7] [42, Theorem 43.3]). This will coerce us to find a bound for the local time of said difference, so the following result is in order.

Lemma 26. *The following bound holds for $(Y^N)_{t\ge 0}$ and $(Y^{N,m})_{t\ge 0}$ defined by (4.11) and (4.13) respectively.*

$$\mathbb{E}[L_t^0(Y_s^N - Y_t^{N,m})] \le 4\lambda\mathbb{E}\Big[\int_0^t \big|Y_s^{N,m} - Y_s^N\big|ds\Big] + 2\mathbb{E}[|E_t^m|]. \tag{4.31}$$

Proof. Let us use Lemma 25 on $Y_t^{N,m} - Y_t^N$ which yields

$$\begin{aligned}&\mathbb{E}\big[L_t^0(Y^N - Y^{N,m})\big] \\ &\quad\le 4\epsilon - 2\mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m}\in(0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m}\ge\epsilon\}}e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}\Big)d(Y_s^N - Y_s^{N,m})\Big] \\ &\qquad + \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s^{N,m}>\epsilon\}}e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}d\langle Y^N - Y^{N,m}\rangle_s\Big]\end{aligned} \tag{4.32}$$

using (4.14). By that definition the second term in (4.32) becomes

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) d(Y_s^N - Y_s^{N,m})\Big] \\
&\quad = \mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) \\
&\qquad\qquad\qquad \lambda\big(u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big) ds\Big] \\
&\quad + \mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) \\
&\qquad\qquad\qquad \big(u_x^N(s,\psi(s,Y_s^{N,m})) - u_x^N(s,\psi(s,Y_s^N))\big) dW_s\Big] \\
&\quad + \mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) dE_t^m\Big],
\end{aligned} \tag{4.33}
$$

note that the second term has a stochastic integral and due to u_x being bounded its expectation vanishes, and we have

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) d(Y_s^N - Y_s^{N,m})\Big] \\
&\quad = \mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) \\
&\qquad\qquad\qquad \lambda\big(u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big) ds\Big] \\
&\quad + \mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) dE_t^m\Big].
\end{aligned} \tag{4.34}
$$

For the third term in (4.32), see that $Y_t^N - Y_t^{N,m}$ is an Itô process and therefore its quadratic variation is

$$
\langle Y^N - Y^{N,m}\rangle_t = \int_0^t \big[u_x^N(s,\psi(s,Y_s^{N,m})) - u_x^N(s,\psi(s,Y_s^N))\big]^2 ds. \tag{4.35}
$$

Then using the arguments above, (4.32) becomes

$$
\begin{aligned}
&\mathbb{E}[L_t^0(Y_s^N - Y_t^{N,m})] \\
&\leq 4\epsilon \\
&\quad - 2\lambda\mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) \\
&\qquad\qquad\qquad\qquad \big(u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big) ds\Big] \\
&\quad - 2\mathbb{E}\Big[\int_0^t \Big(\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon}\Big) dE_t^m\Big] \\
&\quad + \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s^{N,m} > \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \\
&\qquad\qquad\qquad\qquad (u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m})))^2 ds\Big],
\end{aligned} \tag{4.36}
$$

Note that if $Y_s^N - Y_s^{N,m} \geq \epsilon$, then

$$\mathbb{1}_{\{Y_s^N - Y_s^{N,m} \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s^{N,m} \geq \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \leq 2$$

and we have

$$\begin{aligned}
&\mathbb{E}[L_t^0(Y_s^N - Y_t^{N,m})] \\
&\leq 4\epsilon - 4\lambda\mathbb{E}\Big[\int_0^t \big(u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big)ds\Big] - 2\mathbb{E}[E_t^m] \\
&\quad + \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s^{N,m} > \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \\
&\qquad\qquad \big(u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\big)^2 ds\Big],
\end{aligned} \tag{4.37}$$

which is obviously bounded from above by the absolute value of the term on the right hand side, which using Jensen's inequality can be inside of the integrals and expectation; this gives

$$\begin{aligned}
&\mathbb{E}[L_t^0(Y_s^N - Y_t^{N,m})] \\
&\leq 4\epsilon + 4\lambda\mathbb{E}\Big[\int_0^t \big|u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big|ds\Big] + 2\mathbb{E}[|E_t^m|] \\
&\quad + \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s^{N,m} > \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \big|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\big|^2 ds\Big],
\end{aligned} \tag{4.38}$$

Note that the first term on the right-hand side above is such that $4\epsilon \to 0$ as $\epsilon \to 0$, and also we know how to bound the third term by lemma 24, we will come back to that, but for now we need estimates for the second and last terms. For the second term, note u^N is $\frac{1}{2}$-Lipschitz and ψ^N is 2-Lipschitz, so

$$\mathbb{E}\Big[\int_0^t \big|u^N(s,\psi(s,Y_s^{N,m})) - u^N(s,\psi(s,Y_s^N))\big|ds\Big] \leq \mathbb{E}\Big[\int_0^t \big|Y_s^{N,m} - Y_s^N\big|ds\Big]. \tag{4.39}$$

For the last term which involves the differences of u_x^N, note that u_x^N is γ'-Hölder, for $\gamma' < 1-\beta$ with constant $\|u^N\|_{C_T C^{\gamma'}}$. Chose $\zeta \in (0,1)$ such that $\zeta\gamma' > 1/2$, which since $\epsilon \in (0,1)$ implies that $\epsilon^\zeta > \epsilon$. We can split the term as follows

$$\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s^{N,m} > \epsilon\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \\
&\qquad\qquad \big|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\big|^2 ds\Big] \\
&= \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon < Y_s^N - Y_s^{N,m} < \epsilon^\zeta\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \\
&\qquad\qquad \big|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\big|^2 ds\Big] \\
&\quad + \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon^\zeta \leq Y_s^N - Y_s^{N,m}\}} e^{1-(Y_s^N - Y_s^{N,m})/\epsilon} \\
&\qquad\qquad \big|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\big|^2 ds\Big].
\end{aligned} \tag{4.40}$$

For the first term in (4.40) use the Hölder regularity of u_x^N to get the estimate

$$\left|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\right|^2 \leq \|u^N\|^2_{C_TC^{1+\gamma'}} \left|\psi^N(s,Y_s^N) - \psi^N(s,Y_s^{N,m})\right|^{2\gamma'}, \tag{4.41}$$

then since ψ^N is 2-Lipschitz

$$\left|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\right|^2 \leq 2^{2\gamma'}\|u^N\|^2_{C_TC^{1+\gamma'}} \left|Y_s^N - Y_s^{N,m}\right|^{2\gamma'}. \tag{4.42}$$

Therefore, the first integral in (4.40) is bounded as

$$\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon<Y_s^N-Y_s^{N,m}<\epsilon^\zeta\}} e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}\left|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\right|^2 ds \\
&\quad \leq \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t 2^{2\gamma'}\|u^N\|^2_{C_TC^{1+\gamma'}}\left|Y_s^N - Y_s^{N,m}\right|^{2\gamma'} ds\Big] \leq \frac{1}{\epsilon} t 2^{2\gamma'}\|u^N\|^2_{C_TC^{1+\gamma'}}\epsilon^{2\zeta\gamma'},
\end{aligned} \tag{4.43}$$

where the last inequality comes from the fact that $Y_s^N - Y_s < \epsilon^\zeta$ and then integrating the remaining terms. This term also goes to zero as $\epsilon \to 0$.

Finally, the last term in (4.40) using (4.42) is such that

$$\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon^\zeta\leq Y_s^N-Y_s^{N,m}\}} e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}\left|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\right|^2 ds\Big] \\
&\quad \leq \frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon^\zeta\leq Y_s^N-Y_s^{N,m}\}} e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}\big(2\|u_x^N\|_{L_T^\infty L^\infty}\big)^2 ds\Big],
\end{aligned} \tag{4.44}$$

note that the indicator function is bounded from above by 1 and the uniform norm of u_x^N is bounded by 1/2, so all factors in the integral are bounded by 1 except for the exponential, which is bounded by

$$e^{1-(Y_s^N-Y_s^{N,m})/\epsilon} \leq e^{1-\epsilon^{\zeta-1}}. \tag{4.45}$$

So, in sum the integral above is bounded by

$$\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\epsilon^\zeta\leq Y_s^N-Y_s^{N,m}\}} e^{1-(Y_s^N-Y_s^{N,m})/\epsilon}\left|u_x^N(s,\psi^N(s,Y_s^N)) - u_x^N(s,\psi^N(s,Y_s^{N,m}))\right|^2 ds\Big] \\
&\quad \leq \frac{1}{\epsilon} t e^{1-\epsilon^{\zeta-1}}
\end{aligned} \tag{4.46}$$

which goes to zero as $\epsilon \to 0$. After discarding the terms that vanish in (4.38), and applying the bound (4.39) we have

$$\mathbb{E}[L_t^0(Y_s^N - Y_t^{N,m})] \leq 4\lambda\mathbb{E}\Big[\int_0^t \left|Y_s^{N,m} - Y_s^N\right| ds\Big] + 2\mathbb{E}[|E_t^m|], \tag{4.47}$$

as required. □

We finally can put together the lemmas above and come up with an estimate of the error of the numerical approximation for the regularized solution.

Proposition 27. *Assume that* $b^N \in C_T^{1/2}L^\infty \cap L_T^\infty C_b^1$. *Then*

$$\sup_{0\le t\le T} \mathbb{E}[|X_t^{N,m} - X_t^N|] \le C_3(N)m^{-1} + C_4(N)m^{-1/2}, \tag{4.48}$$

where

$$\begin{aligned} C_3(N) &= c_1\|b^N\|_{L_T^\infty L^\infty} + c_2\|b^N\|_{L_T^\infty L^\infty}\|b_x^N\|_{L_T^\infty L^\infty} \\ C_4(N) &= c_3(\|b_x^N\|_{L_T^\infty L^\infty} + [b^N]_{1/2L^\infty}) \end{aligned} \tag{4.49}$$

and the constants are $c_1 = 18Te^{5\lambda T}, c_2 = \frac{3}{2}e^{5\lambda T}, c_3 = 2T^{3/2}e^{5\lambda T}$.

Proof. Recall that ψ^N is 2-Lipschitz, so we have

$$\mathbb{E}|X_t^{N,m} - X_t^N| = \mathbb{E}|\psi^N(t, Y_t^{N,m}) - \psi^N(s, X_t^N)| \le 2\mathbb{E}|Y_t^{N,m} - Y_t^N| \tag{4.50}$$

we apply the Itô-Tanaka formula to $|Y_t^{N,m}Y_t^N|$ which gives

$$\begin{aligned} |Y_t^{N,m} - Y_t^N| = |Y_0^{N,m} - Y_0^N| & \\ &+ \lambda\int_0^t \text{sign}\left(Y_s^{N,m} - Y_s^N\right)\left(u^N(s,\psi^N(s,Y_s^{N,m})) - u^N(s,\psi^N(s,Y_s^N))\right)ds \\ &+ \int_0^t \text{sign}\left(Y_s^{N,m} - Y_s^N\right)\left(u_x^N(s,\psi^N(s,Y_s^{N,m})) - u_x^N(s,\psi^N(s,Y_s^N))\right)dW_s \\ &+ \int_0^t \text{sign}\left(Y_s^{N,m} - Y_s^N\right)dE_s^m + L_t^0\left(Y^{N,m} - Y^N\right). \end{aligned} \tag{4.51}$$

The above is obviously bounded by absolute value of the terms, using triangle inequality and Jensen's inequality leaves the following

$$\begin{aligned} &|Y_t^{N,m} - Y_t^N| \\ &\quad \le \lambda\int_0^t \left|u^N(s,\psi^N(s,Y_s^{N,m})) - u^N(s,\psi^N(s,Y_s^N))\right|ds \\ &\quad + \int_0^t \left|u_x^N(s,\psi^N(s,Y_s^{N,m})) - u_x^N(s,\psi^N(s,Y_s^N))\right|dW_s \\ &\quad + |E_t^M| + L_t^0\left(Y^{N,m} - Y^N\right). \end{aligned} \tag{4.52}$$

Take expectation of (4.52), and note that the expectation of the stochastic integral is zero yields

$$\begin{aligned} \mathbb{E}|Y_t^{N,m} - Y_t^N| \le \mathbb{E}\Big[\lambda\int_0^t \left|u^N(s,\psi^N(s,Y_s^{N,m})) - u^N(s,\psi^N(s,Y_s^N))\right|ds\Big] & \\ + \mathbb{E}|E_t^M| + \mathbb{E}[L_t^0\left(Y^{N,m} - Y^N\right)], & \end{aligned} \tag{4.53}$$

then recall that u^N is 1/2-Lipschitz and ψ^N is 2-Lipschitz

$$\mathbb{E}|Y_t^{N,m} - Y_t^N| \le \lambda\mathbb{E}\Big[\int_0^t |Y_s^{N,m} - Y_s^N|ds\Big] + \mathbb{E}|E_t^M| + \mathbb{E}[L_t^0\left(Y^{N,m} - Y^N\right)], \tag{4.54}$$

Using lemma 26 on the local time term

$$\mathbb{E}|Y_t^{N,m} - Y_t^N| \leq 5\lambda \mathbb{E}\Big[\int_0^t \big|Y_s^{N,m} - Y_s^N\big| ds\Big] + 3\mathbb{E}[|E_t^m|], \tag{4.55}$$

and then by lemma 24

$$\begin{aligned}\mathbb{E}|Y_t^{N,m} - Y_t^N| &\leq 3\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big) \\ &\quad + 5\lambda \mathbb{E}\Big[\int_0^t 3\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big)e^{5\lambda(t-s)} ds\Big]\end{aligned} \tag{4.56}$$

where the constants $C_1(N), C_2(N) > 0$. By Gronwall's inequality we have that

$$\begin{aligned}&\mathbb{E}|Y_t^{N,m} - Y_t^N| \\ &\leq 5\lambda \mathbb{E}\Big[\int_0^t \big|Y_s^{N,m} - Y_s^N\big| ds\Big] + 3\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big) \\ &\leq 5\lambda\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big)\mathbb{E}\Big[\int_0^t e^{5\lambda(t-s)} ds\Big] + 3\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big) \\ &\leq \big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big)\big[e^{5\lambda t} - 1\big] + 3\big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big) \\ &= \big(C_1(N)m^{-1} + C_2(N)m^{-1/2}\big)e^{5\lambda t}.\end{aligned} \tag{4.57}$$

Rearranging the bound and plugging into (4.50) and taking the supremum over t gives the result. □

The constants $C_3(N)$ and $C_4(N)$ in the bound above can be further estimated to get a bound in terms of the original drift b and the smoothing parameter used to create regularized equations, or more precisely the parameter N. Let us first note what the actual constants are when we substitute $C_1(N)$ and $C_2(N)$ from Lemma 24

$$\begin{aligned}C_3(N) &= 18Te^{5\lambda T}\|b^N\|_{L_T^\infty L^\infty} + \frac{3}{2}Te^{5\lambda T}\|b^N\|_{L_T^\infty L^\infty}\|b_x^N\|_{L_T^\infty L^\infty} \\ C_4(N) &= 2e^{5\lambda T}T^{3/2}\|b_x^N\|_{L_T^\infty L^\infty} + 2e^{5\lambda T}T^{3/2}[b^N]_{1/2,L^\infty}.\end{aligned} \tag{4.58}$$

In the following corollary we deduce bounds for the terms including norms or seminorms of b^N or its derivative.

Corollary 28. *Under Assumption 22, the condition of Proposition 27 is satisfied and for any $\epsilon > 0$ we have*

$$C_3(N) \leq c_S^1 c_1 N^{\frac{\epsilon+\hat{\beta}}{2}}\|b\|_{C_T C^{-\hat{\beta}}} + c_S^1 c_2 N^{\frac{\epsilon+\hat{\beta}+1}{2}}\|b\|^2_{C_T C^{-\hat{\beta}}}$$

and

$$C_4(N) \leq c_S^1 c_3 N^{\frac{\epsilon+\hat{\beta}+1}{2}}\|b\|_{C_T C^{-\hat{\beta}}} + c_S^1 c_3 N^{\frac{\epsilon+\hat{\beta}}{2}}\|b\|_{C_T^{\frac{1}{2}} C^{-\hat{\beta}}},$$

where the constants c_1, c_2, c_3 are from Proposition 27 and c_S^1 is from Lemma 9.

Proof. We will show that $b^N \in C_T^{\frac{1}{2}}L^\infty \cap L_T^\infty C_b^1$ and find the appropriate bounds simultaneously by finding a bound for each of the terms $\|b^N\|_{L_T^\infty L^\infty}$, $\|b_x^N\|_{L_T^\infty L^\infty}$ and $[b^N]_{1/2 L^\infty}$ in

(4.48).

First we apply Lemma 9 with ϵ from the statement of this corollary, so we have $\zeta = \frac{\epsilon+\hat{\beta}}{2}$, and $\xi - 2\zeta = -\hat{\beta}$, which means $\xi = \epsilon$. This tells us that $b^N(t,\cdot) \in C^\epsilon$ for all $t \in [0,T]$ and an arbitrarily small ϵ, so $b^N(t,\cdot) \in L^\infty$. Then, note that the classical Hölder norm (3.6) implies

$$\|b^N(t,\cdot) - b^N(s,\cdot)\|_{L^\infty} = \|P_{\frac{1}{N}}(b(t,\cdot) - b(s,\cdot))\|_{L^\infty} \leq \|P_{\frac{1}{N}}(b(t,\cdot) - b(s,\cdot))\|_{C^\epsilon}$$

and by the first Schauder estimate (Lemma 9)

$$\|b^N(t,\cdot) - b^N(s,\cdot)\|_{L^\infty} \leq c_S^1 N^{\frac{\epsilon+\hat{\beta}}{2}} \|b(t,\cdot) - b(s,\cdot)\|_{C^{-\hat{\beta}}}.$$

Hence, by definition of the seminorm $[\cdot]_{1/2,L^\infty}$ in (3.8)

$$[b^N]_{\frac{1}{2},L^\infty} \leq c_S^1 N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T^{\frac{1}{2}} C^{-\hat{\beta}}}. \tag{4.59}$$

By the same arguments applied to $b^N(t,\cdot)$, we obtain $\|b^N(t,\cdot)\|_{L^\infty} \leq c_S^1 N^{\frac{\epsilon+\hat{\beta}}{2}} \|b(t,\cdot)\|_{C^{-\hat{\beta}}}$, which yields

$$\|b^N\|_{L_T^\infty L^\infty} \leq \|b^N\|_{C_T C^\epsilon} \leq c_S^1 N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}}. \tag{4.60}$$

By bounds (4.59) and (4.60) we conclude that $b^N \in C_T^{\frac{1}{2}} L^\infty$.

It remains to show that $b_x^N(t,\cdot)$ exists and is bounded uniformly in $t \in [0,T]$. The derivative $b_x^N(t,\cdot)$ is well defined for all $t \in [0,T]$ because $b^N \in C_T C^\gamma$ for all $\gamma > 0$. By the definition of the equivalent norm $\|\cdot\|_{C^\gamma}$ (3.6), Bernstein's inequality (Lemma 7) and the first Schauder estimate we have

$$\|b_x^N\|_{L_T^\infty L^\infty} \leq \|b_x^N\|_{C_T C^\epsilon} \leq c_S^1 \|b^N\|_{C_T C^{\epsilon+1}} \leq c_S^1 N^{\frac{\epsilon+\hat{\beta}+1}{2}} \|b\|_{C_T C^{-\hat{\beta}}}. \tag{4.61}$$

Inequalities (4.60) and (4.61) show that $b^N \in L_T^\infty C_b^1$.

Finally we insert the bounds derived above into (4.48) from Proposition 27 to get the result. $\square$

4.4 Convergence of the regularized solution to the real solution

Since we are approximating $(X)_{t\geq 0}$ through $(X^N)_{t\geq 0}$ we need to know if there is convergence between those two. The result which proves that requires of several auxiliary results and the definition of new objects to work with. This result is in Proposition 35 and with the goal to get there, we recall the auxiliary process Y^N, which is the weak solution of the SDE (4.11).

The main idea of the result we are after is to rewrite the solutions X^N and X in their equivalent virtual formulations and thus study the error between the auxiliary processes Y^N and Y defined in (4.11) and (3.44) respectively. This transformation has the advantage that the SDEs for Y^N and Y are classical with strong solutions, see Remark 21. After writing the difference $X^N - X$ in terms of $Y^N - Y$ we apply Itô's formula to $|Y^N - Y|$. We will control the stochastic and Lebesgue integrals by $u^N - u$ and $\psi^N - \psi$, which in turn are bounded by some function of $b^N - b$. The term involving the local time at 0 of $Y^N - Y$ is handled in Lemma 25.

In the remaining of this section, we fix $\beta' \in (\hat{\beta}, 1/2)$ and $\hat{\alpha} \in (0, 1-\hat{\beta})$ in anticipation of the statement of Proposition 35. Since $\beta' \in (\hat{\beta}, 1/2)$, we have $b \in C_T C^{-\beta'}$ by Assumption 22. Notice that the solutions u and u^N obtained when taking b as an element of $C_T^{1/2} C^{(-\beta)+}$ are the same as when viewing the same b as an element of $C_T C^{-\beta'}$. We obviously have $u, u^N \in C_T C^{(2-\beta')-}$ and $u, u^N \in C_T C^{1+\alpha}$ for any $\alpha < 1 - \hat{\beta}$.

Lemma 29. *There are ρ_0 and c such that for any $\rho \geq \rho_0$ and for any $\alpha < 1 - \beta'$ we have*

$$\|u - u^N\|^{(\rho)}_{C_T C^{1+\alpha}} \leq 2c \|b - b^N\|_{C_T C^{-\beta'}} (\|u\|^{(\rho)}_{C_T C^{1+\alpha}} - 1) \rho^{\frac{\alpha + \beta' - 1}{2}}.$$

Proof. The bound in the statement of the lemma forms the main part of the proof of [25, Lemma 4.17] in the special case when the terminal condition is zero and $g^N = b^N$, $g = b$. □

The parameter ρ and the exact dependence of the bound on it is not important for our arguments, so we may fix ρ that satisfies the conditions of the above lemma. However, for completeness of the presentation, we will mention the dependence on ρ_0 and ρ in results below.

Using Lemma 29 we can derive a uniform bound on the L^∞-norm of the difference $u(t) - u^N(t)$ and $u_x(t) - u_x^N(t)$.

Lemma 30. *For any $\rho \geq \rho_0$, where ρ_0 is from Lemma 29, there is $\kappa > 0$ such that for all $t \in [0, T]$*

$$\|u(t) - u^N(t)\|_{L^\infty} + \|u_x(t) - u_x^N(t)\|_{L^\infty} \leq \kappa \|b - b^N\|_{C_T C^{-\beta'}}.$$

The constant κ depends also on c from Lemma 29 and on T and u.

Proof. We recall that $u, u^N \in C_T C^{1+\alpha}$ and $u_x, u_x^N \in C_T C^{\alpha}$ for any $\alpha < 1 - \hat{\beta}$. Recall also that the norm in C^γ for $\gamma \in (0,1)$ is given by (3.6) and in C^γ for $\gamma \in (1,2)$ is given by (3.7). Using this together with Bernstein inequality we get for all $t \in [0, T]$ that

$$\|u(t) - u^N(t)\|_{L^\infty} + \|u_x(t) - u_x^N(t)\|_{L^\infty} \leq c' \|u(t) - u^N(t)\|_{C^{1+\alpha}} \leq c' \|u - u^N\|_{C_T C^{1+\alpha}},$$

for some $c' > 0$. We use (3.10) on the last term above, which gives

$$c' \|u - u^N\|_{C_T C^{1+\alpha}} \leq c' e^{\rho T} \|u - u^N\|^{(\rho)}_{C_T C^{1+\alpha}}.$$

Finally, by Lemma 29 the statement holds with constant κ given by

$$\kappa = 2cc' e^{\rho T} (\|u\|^{(\rho)}_{C_T C^{1+\alpha}} - 1) \rho^{\frac{\alpha + \beta' - 1}{2}}.$$

□

Next we derive a bound for the difference $\psi - \psi^N$, where we recall that the two functions are the space-inverses of ϕ and ϕ^N defined in (3.29) and (4.8).

Lemma 31. *Take κ from Lemma 30. We have*

$$\sup_{(t,y)\in[0,T]\times\mathbb{R}} |\psi(t,y) - \psi^N(t,y)| \le 2\kappa \|b - b^N\|_{C_T C^{-\beta'}}. \tag{4.62}$$

Proof. Let us first recall that $\|u_x\|_{L^\infty} \le \frac{1}{2}$, see the discussion at the beginning of this section. Hence u_x is $\frac{1}{2}$-Lipschitz. Using this we have for any $x, x' \in \mathbb{R}$

$$\begin{aligned} |\phi(t,x) - \phi(t,x')| &= |(x + u(t,x)) - (x' + u(t,x'))| \\ &\ge |x - x'| - |u(t,x) - u(t,x')| \\ &\ge |x - x'| - \frac{1}{2}|x - x'| \\ &\ge \frac{1}{2}|x - x'|. \end{aligned} \tag{4.63}$$

Insert $x = \psi(t,y)$ and $x' = \psi^N(t,y)$ for some $y \in \mathbb{R}$, to get the bound

$$|\phi(t,\psi(t,y)) - \phi(t,\psi^N(t,y))| \ge \frac{1}{2}|\psi(t,y) - \psi^N(t,y)|.$$

This implies the first inequality below

$$\begin{aligned} |\psi(t,y) - \psi^N(t,y)| &\le 2|\phi(t,\psi(t,y)) - \phi(t,\psi^N(t,y))| \\ &= 2|\phi^N(t,\psi^N(t,y)) - \phi(t,\psi^N(t,y))| \\ &= 2|u^N(t,\psi^N(t,y)) - u(t,\psi^N(t,y))|, \end{aligned} \tag{4.64}$$

where we used $\phi^N(t,\psi^N(t,y)) = y = \phi(t,\psi(t,y))$ and the definition of ϕ and ϕ^N. Thus

$$|\psi(t,y) - \psi^N(t,y)| \le 2\|u(t) - u^N(t)\|_{L^\infty} \le 2\kappa\|b - b^N\|_{C_T C^{-\beta'}},$$

having used Lemma 30 in the last inequality. ☐

Finally, we derive a bound for $|u^N(s,\psi^N(s,y)) - u(s,\psi(s,y'))|$ and for $|u_x^N(s,\psi^N(s,y)) - u_x(s,\psi(s,y'))|$.

Lemma 32. *For any $\beta' \in (\hat\beta, 1/2)$, any $\alpha < 1 - \hat\beta$ and any $y, y' \in \mathbb{R}$, the following bounds are satisfied*

$$|u^N(s,\psi^N(s,y')) - u(s,\psi(s,y))| \le 2\kappa\|b^N - b\|_{C_T C^{-\beta'}} + |y - y'|, \tag{4.65}$$

and

$$\begin{aligned} |u_x^N(s,\psi^N(s,y')) - u_x(s,\psi(s,y))| \le &\kappa\|b - b^N\|_{C_T C^{-\beta'}} + 2^\alpha\kappa^\alpha\|u\|_{C_T C^{1+\alpha}}\|b^N - b\|^\alpha_{C_T C^{-\beta'}} \\ &+ |u_x(s,\psi(s,y)) - u_x(s,\psi(s,y'))|, \end{aligned} \tag{4.66}$$

where κ is from Lemma 30.

Proof. We start with (4.65). We rewrite the left-hand side as

$$\begin{aligned} |u^N(s,\psi^N(s,y')) - u(s,\psi(s,y))| \le |u^N(s,\psi^N(s,y')) - u(s,\psi^N(s,y'))| \\ + |u(s,\psi^N(s,y')) - u(s,\psi(s,y'))| \\ + |u(s,\psi(s,y')) - u(s,\psi(s,y))|. \end{aligned} \tag{4.67}$$

Using Lemma 30, we bound the first term above by

$$|u^N(s,\psi^N(s,y')) - u(s,\psi^N(s,y'))| \le \|u^N(s) - u(s)\|_{L^\infty} \le \kappa \|b - b^N\|_{C_T C^{-\beta'}}.$$

The second term in (4.67) is bounded using the fact that $\|u_x\|_{L^\infty} \le \frac{1}{2}$ (see the discussion at the beginning of this section) and Lemma 31:

$$|u(s,\psi^N(s,y')) - u(s,\psi(s,y'))| \le \frac{1}{2}|\psi^N(s,y') - \psi(s,y')| \le \kappa \|b - b^N\|_{C_T C^{-\beta'}}.$$

For the third term in (4.67), we use again that $u(t,\cdot)$ is $\frac{1}{2}$-Lipschitz

$$|u(s,\psi(s,y')) - u(s,\psi(s,y))| \le \frac{1}{2}|\psi(s,y') - \psi(s,y)| \le |y - y'|,$$

where the last inequality follows from the fact that $\psi(t,\cdot)$ is 2-Lipschitz (c.f. (3.3)). Combining the above estimates proves (4.65).

Let us now prove (4.66). Similarly, as above we write

$$\begin{aligned} |u^N_x(s,\psi^N(s,y')) - u_x(s,\psi(s,y))| \le |u^N_x(s,\psi^N(s,y')) - u_x(s,\psi^N(s,y'))| \\ + |u_x(s,\psi^N(s,y')) - u_x(s,\psi(s,y'))| \\ + |u_x(s,\psi(s,y')) - u_x(s,\psi(s,y))|. \end{aligned} \tag{4.68}$$

The first term is bounded using Lemma 30 as follows

$$|u^N_x(s,\psi^N(s,y')) - u_x(s,\psi^N(s,y'))| \le \|u^N_x(s) - u_x(s)\|_{L^\infty} \le \kappa \|b - b^N\|_{C_T C^{-\beta'}}.$$

Since $b \in C_T C^{-\beta'}$ and $\alpha \in (1/2, 1 - \hat{\beta})$ then $u \in C_T C^{1+\alpha}$, so by Bernstein inequality (Lemma 7) we have $u_x(t) \in C_T C^\alpha$ with $\|u_x\|_{C_T C^\alpha} \le \|u\|_{C_T C^{1+\alpha}}$. Recalling the norm (3.6) in C^α, we conclude that $u_x(t)$ is α-Hölder continuous with constant $\|u\|_{C_T C^{1+\alpha}}$. This allows us to bound the second term in (4.68) as

$$|u_x(s,\psi^N(s,y')) - u_x(s,\psi(s,y'))| \le \|u\|_{C_T C^{1+\alpha}} |\psi^N(s,y') - \psi(s,y')|^\alpha$$

and by Lemma 31 we get

$$|u_x(s,\psi^N(s,y')) - u_x(s,\psi(s,y'))| \le \|u\|_{C_T C^{1+\alpha}} 2^\alpha \kappa^\alpha \|b - b^N\|^\alpha_{C_T C^{-\beta'}}.$$

This concludes the proof. □

Remark 33. Note that in our main result we find the strong error in L^1 instead of the most commonly used estimates in L^2, the reason for this is illustrated here. Let us apply Itô's formula to $(Y_t^N - Y_t)^2$, integrate from 0 to t and take expectation. Apply further Lemma 32 and recall that $\|u_x\|_\infty < 1/2$, ψ is 2-Lipschitz, and u_x is α-Hölder continuous for any $\alpha < 1 - \hat{\beta}$. We get

$$\begin{aligned}\mathbb{E}(Y_t^N - Y_t)^2 - (Y_0^N - Y_0)^2 \leq 2\mathbb{E}\int_0^t |Y_s^N - Y_s|^2 + 4\kappa\|b^N - b\|_{C_T C^{-\beta'}}\mathbb{E}\int_0^t |Y_s^N - Y_s| ds \\ + t(\kappa\|b^N - b\|_{C_T C^{-\beta'}} + 2^\alpha\kappa^\alpha\|u\|_{C_T C^{-\beta}}\|b^N - b\|^\alpha_{C_T C^{-\beta'}})^2 \\ + \mathbb{E}\int_0^t c^2 2^{2\alpha}|Y_s^N - Y_s|^{2\alpha} ds.\end{aligned}$$

Note that the second and last terms in the inequality have the difference $|Y_s^N - Y_s|$ in smaller powers than 2 (the latter since $\alpha < 1$), so they decrease slower than the term on the left-hand side and Gronwall inequality cannot be applied. The power 2α in the last term follows from the fact that u_x is α-Hölder continuous which is an inherent feature of the construction of the virtual solution to (4.1).

The next Proposition is a bound for the local time of $Y^N - Y$. This is a key bound in the proof of the L^1-convergence of X^N to X, due to the application of Itô's formula to $|Y^N - Y|$.

Proposition 34. *For any $\alpha \in (1/2, 1 - \hat{\beta})$ and N such that $\|b^N - b\|_{C_T C^{-\beta'}} < 1$ we have*

$$\mathbb{E}[L_t^0(Y^N - Y)] \leq A\,\mathbb{E}[\int_0^t |Y_s^N - Y_s| ds] + B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}} + o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}), \tag{4.69}$$

for some constants $A, B > 0$.

Proof. This proof follows the same steps as the proof of [14, Proposition 5.4] with differences coming from the spaces that b and b^N belong to. It is provided for the reader's convenience.

Recall that Y_t, Y_t^N are strong solutions of (3.44) and (4.11) respectively, see Remark 21. Their difference satisfies

$$\begin{aligned}Y_t^N - Y_t = (y_0^N - y_0) + \lambda\int_0^t (u^N(s, \psi^N(s, Y_s^N)) - u(s, \psi(s, Y_s))) ds \\ + \int_0^t (u_x^N(s, \psi^N(s, Y_s^N)) - u_x(s, \psi(s, Y_s))) dW_s.\end{aligned} \tag{4.70}$$

We apply Lemma 25 to $Y_t^N - Y_t$ for $\epsilon \in (0, 1)$ and since the expectation of the integral with respect to the Brownian motion (W_t) is zero thanks to the fact that u_x and u_x^N are bounded uniformly in N we get

$$\begin{aligned}\mathbb{E}[L_t^0(Y^N - Y)] \leq 4\epsilon \\ - 2\lambda\mathbb{E}\Bigg[\int_0^t (\mathbb{1}_{\{Y_s^N - Y_s \in (0,\epsilon)\}} + \mathbb{1}_{\{Y_s^N - Y_s \geq \epsilon\}} e^{1 - \frac{Y_s^N - Y_s}{\epsilon}})\end{aligned} \tag{4.71}$$

$$
(u^N(s,\psi^N(s,Y_s^N)) - u(s,\psi(s,Y_s)))ds\Bigg] \tag{4.72}
$$

$$
+\frac{1}{\epsilon}\mathbb{E}\Bigg[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} \tag{4.73}
$$

$$
(u_x^N(s,\psi^N(s,Y_s^N)) - u_x(s,\psi(s,Y_s)))^2 ds\Bigg], \tag{4.74}
$$

Notice that if $Y_s^N - Y_s \geq \epsilon$, then $e^{1-\frac{Y_s^N - Y_s}{\epsilon}} \leq 1$. Hence, (4.72) is bounded above by

$$
\begin{aligned}
4\lambda\mathbb{E}[\int_0^t \big(u(s,\psi(s,Y_s)) - u^N(s,\psi^N(s,Y_s^N))\big)ds]& \\
&\leq 4\lambda(2\kappa\|b^N - b\|_{C_TC^{-\beta'}}t + \mathbb{E}[\int_0^t |Y_s^N - Y^N|ds]),
\end{aligned}
$$

where the last inequality is by Lemma 32.

Now for (4.74), we use again the observation that $\mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} \leq 1$, the estimate (4.66) from Lemma 32 choosing $\alpha' \in (\alpha, 1-\hat{\beta})$, and the inequality

$$
(x_1 + x_2 + x_3)^2 \leq 3(x_1^2 + x_2^2 + x_3^2)
$$

to get the bound

$$
\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}\int_0^t (3\kappa^2\|b - b^N\|^2_{C_TC^{-\beta'}} + 3\cdot 2^{2\alpha'}\kappa^{2\alpha'}\|b^N - b\|^{2\alpha'}_{C_TC^{-\beta'}}\|u\|^2_{C_TC^{1+\alpha'}})ds \\
&\qquad + \frac{1}{\epsilon}\mathbb{E}\int_0^t 3\mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} |u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds \\
&\leq \frac{3t}{\epsilon}\|b^N - b\|_{C_TC^{-\beta'}}(\kappa^2\|b^N - b\|_{C_TC^{-\beta'}} + (2\kappa)^{2\alpha'}\|u\|^2_{C_TC^{1+\alpha'}}\|b^N - b\|^{2\alpha'-1}_{C_TC^{-\beta'}}) \\
&\qquad + \frac{3}{\epsilon}\mathbb{E}(\int_0^t \mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} |u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds).
\end{aligned}
$$

Putting above estimates together yields the following bound for the expectation of the local time:

$$
\begin{aligned}
\mathbb{E}[L_t^0(Y^N - Y)] \leq 4\epsilon + 4\lambda(2\kappa\|b^N - b\|_{C_TC^{-\beta}}t + \mathbb{E}[\int_0^t |Y_s^N - Y^N|ds])& \\
+ \frac{3t}{\epsilon}\|b^N - b\|_{C_TC^{-\beta}}& \\
(\kappa^2\|b^N - b\|_{C_TC^{-\beta}} + (2\kappa)^{2\alpha'}\|u\|^2_{C_TC^{1+\alpha'}}\|b^N - b\|^{2\alpha'-1}_{C_TC^{-\beta}})& \\
+ \frac{3}{\epsilon}\mathbb{E}\Bigg[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}}& \\
|u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds\Bigg]&.
\end{aligned} \tag{4.75}
$$

We take $\epsilon = \|b^N - b\|_{C_TC^{-\beta}} < 1$ and choose $\zeta \in (0,1)$ such that $\alpha'\zeta > 1/2$. Recall that $u_x(s,\cdot)$ is α'-Hölder continuous with the constant $\|u\|_{C_TC^{1+\alpha'}}$, as $\alpha' < 1 - \hat{\beta}$, (see Section

3.4) and $\psi(s,\cdot)$ is 2-Lipschitz, so

$$|u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))| \le \|u\|_{C_T C^{1+\alpha'}} 2^{\alpha'} |Y_s^N - Y_s|^{\alpha'}.$$

This bound and the observation that $\epsilon^\zeta > \epsilon$ yield

$$\begin{aligned}
&\frac{3}{\epsilon}\mathbb{E}[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s > \epsilon\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} |u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds] \\
&\le \frac{3}{\epsilon}\mathbb{E}[\int_0^t \mathbb{1}_{\{\epsilon < Y_s^N - Y_s \le \epsilon^\zeta\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} |u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds] \\
&+ \frac{3}{\epsilon}\mathbb{E}[\int_0^t \mathbb{1}_{\{Y_s^N - Y_s > \epsilon^\zeta\}} e^{1-\frac{Y_s^N - Y_s}{\epsilon}} |u_x(s,\psi(s,Y_s^N)) - u_x(s,\psi(s,Y_s))|^2 ds] \\
&\le \frac{3}{\epsilon} t \|u\|^2_{C_T C^{1+\alpha'}} 2^{2\alpha'} \epsilon^{2\alpha'\zeta} + \frac{3}{\epsilon} t e^{1-\epsilon^{\zeta-1}},
\end{aligned}$$

where in the last inequality we used that $\|u_x\|_{L^\infty} \le 1/2$. We insert this bound into (4.75) and recall that $\epsilon = \|b^N - b\|_{C_T C^{-\beta'}}$ to obtain

$$\begin{aligned}
&\mathbb{E}[L_t^0(Y^N - Y)] \\
&\le (4 + 8\lambda\kappa t + 3\kappa^2 t)\|b^N - b\|_{C_T C^{-\beta'}} + 4\lambda(\mathbb{E}[\int_0^t |Y_s^N - Y^N| ds]) \\
&\quad + 3t((2\kappa)^{2\alpha'} \|u\|^2_{C_T C^{1+\alpha'}} \|b^N - b\|^{2\alpha'-1}_{C_T C^{-\beta'}}) \\
&\quad + 3t\|u\|^2_{C_T C^{1+\alpha'}} 2^{2\alpha'} \|b^N - b\|^{2\alpha'\zeta - 1}_{C_T C^{-\beta'}} + \frac{3t}{\|b^N - b\|_{C_T C^{-\beta'}}} e^{1-\|b^N - b\|^{\zeta-1}_{C_T C^{-\beta'}}} \\
&\le 3T\|u\|^2_{C_T C^{1+\alpha'}} 2^{2\alpha'} \|b^N - b\|^{2\alpha'\zeta-1}_{C_T C^{-\beta'}} + 4\lambda(\mathbb{E}[\int_0^T |Y_s^N - Y^N| ds]) + o(\|b^N - b\|^{2\alpha'\zeta-1}_{C_T C^{-\beta'}}).
\end{aligned}$$

Taking $\zeta = \alpha/\alpha'$ completes the proof. □

Proposition 35. *Under Assumption 22, for any $\hat\alpha \in (1/2, 1-\hat\beta)$ and any $\beta \in (\hat\beta, 1/2)$, there is a constant c such that*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t^N - X_t|] \le c\|b^N - b\|^{2\hat\alpha - 1}_{C_T C^{-\beta'}}$$

for all N sufficiently large so that $\|b^N - b\|_{C_T C^{-\beta'}} < 1$.

Proof. For our arguments we fix ρ that satisfies conditions of Lemma 29 and denote by κ the constant from Lemma 30. By the definition of ψ, ψ^N we have

$$\begin{aligned}
|X_t^N - X_t| &\le |\psi^N(t,\phi^N(t,X_t^N)) - \psi(t,\phi(t,X_t))| \\
&= |\psi^N(t,Y_t^N) - \psi(t,Y_t)|,
\end{aligned} \tag{4.76}$$

then adding and subtracting, and using the triangle inequality we get

$$|X_t^N - X_t| \le |\psi^N(t,Y_t^N) - \psi(t,Y_t^N)| + |\psi(t,Y_t^N) - \psi(t,Y_t)|. \tag{4.77}$$

Where the first term is bounded by $2\kappa\|b^N - b\|_{C_T C^{-\beta'}}$ (see Lemma 31) and since ψ is

2-Lipschitz uniformly in $t \in [0,T]$ the second term is bounded by $2|Y^N - Y|$, therefore

$$|X^N - X| \leq 2\kappa \|b^N - b\|_{C_T C^{-\beta'}} + 2|Y^N - Y|. \tag{4.78}$$

By assumption the first term above goes to zero as $N \to \infty$, then we only need a bound for the second term. By Itô-Tanaka's formula

$$|Y^N - Y| = |y_0^N - y_0| + \frac{1}{2} L_t^0 (Y^N - Y) + \int_0^t \operatorname{sign}(Y^N - Y) d(Y^N - Y), \tag{4.79}$$

by taking expectation and using the definitions of Y^N, Y we have

$$\begin{aligned} \mathbb{E}|Y^N - Y| &= \mathbb{E}|y_0^N - y_0| + \mathbb{E}\frac{1}{2} L_t^0 (Y^N - Y) \\ &\quad + \lambda \mathbb{E} \int_0^t \operatorname{sign}(Y^N - Y)(u^N(s, \psi^N(s, Y_s^N)) - u(s, \psi(s, Y_s)))ds, \end{aligned} \tag{4.80}$$

then observe that the first term above is a constant, for the second we have a bound in Proposition 34 and for the third we use Lemma 32, and the fact that $\operatorname{sign}(x) \leq 1$ therefore

$$\begin{aligned} \mathbb{E}|Y^N - Y| \leq\ & |u^N(0,x) - u(0,x)| \\ & + \frac{1}{2}[o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}) + A\mathbb{E}(\int_0^t \|Y^N - Y\| ds) + B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}] \\ & + \mathbb{E}[\int_0^t (2\kappa_\rho \|b^N - b\|_{C_T C^{-\beta'}} + \|Y_s^N - Y^N\|) ds] \\ \leq\ & |u^N(0,x) - u(0,x)| \\ & + \frac{1}{2}[o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}) + A\mathbb{E}(\int_0^t \|Y^N - Y\| ds) + B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}] \\ & + \lambda 2\kappa_\rho \|b^N - b\|_{C_T C^{-\beta'}} t + \lambda \mathbb{E}(\int_0^t \|Y_s^N - Y^N\| ds). \end{aligned} \tag{4.81}$$

Note that the first and third terms above are controlled by $o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}})$, so after merging those terms and the two involving $\mathbb{E}(\int_0^t \|Y_s^N - Y^N\| ds)$ we get

$$\mathbb{E}|Y^N - Y| \leq B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}} + (A+\lambda)\mathbb{E}(\int_0^t \|Y_s^N - Y^N\| ds) + o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}). \tag{4.82}$$

From there, using Gronwall's lemma we get the following bound

$$\mathbb{E}|Y^N - Y| \leq B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}} T e^{(A+\lambda)T} + o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}). \tag{4.83}$$

Now we use (4.83) to bound (4.78), and as the *small-o* term controls the second term in (4.78) we obtain

$$\mathbb{E}[|X^N - X|] \leq B\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}} T e^{(A+\lambda)T} + o(\|b^N - b\|^{2\alpha-1}_{C_T C^{-\beta'}}) \tag{4.84}$$

□

Notice that Proposition 35 does not really depend on the shape of the drift, but is rather a comparison of generic drifts which we can also phrase as follows.

Corollary 36. *Under Assumption 22, for any $\hat{\alpha} \in (1/2, 1-\hat{\beta})$ and any $\mu^1, \mu^2 \in C_T C^{-\beta'}$ for $\beta' \in (\hat{\beta}, 1/2)$ and SDEs*

$$dY_t^1 = \mu^1(t, Y_t^1)dt + dW_t$$
$$dY_t^2 = \mu^2(t, Y_t^2)dt + dW_t$$

there is a constant c such that if $\|\mu^1 - \mu^2\|_{C_T C^{-\beta'}} < 1$, we have

$$\sup_{t\in[0,T]} \mathbb{E}[|Y_t^1 - Y_t^2|] \leq c\|\mu_1 - \mu_2\|_{C_T C^{-\beta'}}^{2\hat{\alpha}-1}.$$

4.5 The convergence rate

The approximation error of our numerical scheme comes from two sources: the time discretization error from Euler-Maruyama scheme, which depends on m, and the smoothing error coming from replacing b with b^N in the SDE, which depends on N. We will now show how to balance those two sources of errors and bound the resulting convergence rate. To this end, we parametrize N in terms of m and postulate that this parametrization is of the form $N(m) = m^\eta$ for some $\eta > 0$. We denote

$$\hat{X}^m := X^{N(m),m}$$

and consider the strong error denoted by

$$\Upsilon(m) = \sup_{0\leq t\leq T} \mathbb{E}[|\hat{X}_t^m - X_t|].$$

Take any $\epsilon \in (0, 1/2 - \hat{\beta})$, $\beta' \in (\hat{\beta}, 1/2)$ and $\hat{\alpha} \in (1/2, 1-\hat{\beta})$. By the triangle inequality, we have

$$\Upsilon(m) \leq \sup_{0\leq t\leq T} \mathbb{E}[|\hat{X}_t^m - X_t^{N(m)}|] + \sup_{0\leq t\leq T} \mathbb{E}[|X_t^{N(m)} - X_t|] \tag{4.85}$$

then we use Propositions 27 and 35 to get

$$\Upsilon(m) \leq C_3(N)m^{-1} + C_4(N)m^{-1/2} + c\|b^{N(m)} - b\|_{C_T C^{-\beta'}}^{2\hat{\alpha}-1} \tag{4.86}$$

then we use Corollary 28 on the first two terms which yields

$$\begin{aligned} \Upsilon(m) \leq c' \Big[& m^{\eta\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}} (1 + m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}} \|b\|_{C_T C^{-\hat{\beta}}}) m^{-1} \\ & + \Big(m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}} \|b\|_{C_T C^{-\hat{\beta}}} + m^{\eta\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T^{\frac{1}{2}} C^{-\hat{\beta}}} \Big) m^{-\frac{1}{2}} \\ & + c\|b^{N(m)} - b\|_{C_T C^{-\beta'}}^{2\hat{\alpha}-1} \Big] \end{aligned} \tag{4.87}$$

where c' is a constant big enough to encompass the constant of the results used. Then the following estimate from Lemma 9

$$\|b^{N(m)}(t,\cdot)-b(t,\cdot)\|_{C^{-\beta'}} \leq cN(m)^{-\frac{\beta'-\hat{\beta}}{2}}\|b(t,\cdot)\|_{C^{-\hat{\beta}}},$$

for $t\in[0,T]$ allows us to bound the last term and we have

$$\begin{aligned}\Upsilon(m) \leq c''\Big[& m^{\eta\frac{\epsilon+\hat{\beta}}{2}}\|b\|_{C_TC^{-\hat{\beta}}}(1+m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}}\|b\|_{C_TC^{-\hat{\beta}}})m^{-1}\\ &+\Big(m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}}\|b\|_{C_TC^{-\hat{\beta}}}+m^{\eta\frac{\epsilon+\hat{\beta}}{2}}\|b\|_{C_T^{\frac{1}{2}}C^{-\hat{\beta}}}\Big)m^{-\frac{1}{2}}\\ &+m^{-\eta\frac{\beta'-\hat{\beta}}{2}(2\hat{\alpha}-1)}\|b\|^{2\hat{\alpha}-1}_{C_TC^{-\hat{\beta}}}\Big],\end{aligned} \tag{4.88}$$

where c'' is a constant that absorbs all other constants.

Since all the norms appearing on the right hand side are finite, they can be absorbed by a constant c''' and we have

$$\Upsilon \leq c'''\Big[m^{\eta\frac{\epsilon+\hat{\beta}}{2}-1}+m^{\eta\frac{2\epsilon+2\hat{\beta}+1}{2}-1}+m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}-\frac{1}{2}}+m^{\eta\frac{\epsilon+\hat{\beta}}{2}-\frac{1}{2}}+m^{-\eta\frac{(\beta'-\hat{\beta})(2\hat{\alpha}-1)}{2}}\Big]. \tag{4.89}$$

Before proceeding further, we select the optimal β' and $\hat{\alpha}$ to maximise the decreasing rate of the last term in m. Recalling the constraints for β' and $\hat{\alpha}$, the product $(\beta'-\hat{\beta})(2\hat{\alpha}-1)$ is maximised for $\hat{\alpha}\approx 1-\hat{\beta}$ and $\beta'\approx 1/2$. We take $\hat{\alpha}=1-\hat{\beta}-\epsilon$ and $\beta'=1/2-\epsilon$ which yields the exponent $2(1/2-\hat{\beta}-\epsilon)^2$. The last term of (4.89) takes the form

$$m^{-\eta(\frac{1}{2}-\hat{\beta}-\epsilon)^2}.$$

The monotonicity of the remaining four terms in (4.89) depends on η. For the scheme to converge, we need to make sure that they all decrease which is guaranteed if $\eta\frac{2\epsilon+2\hat{\beta}+1}{2}-1<0$ and $\eta\frac{\epsilon+\hat{\beta}+1}{2}-\frac{1}{2}<0$. Therefore we have the constraint

$$0<\eta<\frac{1}{\epsilon+\hat{\beta}+1}. \tag{4.90}$$

At this point we have to find the optimal value of η. Here, the slowest decreasing term out of the first four terms of (4.89) is $m^{\eta\frac{\epsilon+\hat{\beta}+1}{2}-\frac{1}{2}}$. To balance Euler-Maruyama error measured by the first four terms and the error of approximating b with $b^{N(m)}$ in the last term we equate the rates (exponents)

$$\eta\frac{\epsilon+\hat{\beta}+1}{2}-\frac{1}{2}=-\eta(\frac{1}{2}-\hat{\beta}-\epsilon)^2. \tag{4.91}$$

This leads to

$$\eta=\frac{1}{\epsilon+\hat{\beta}+1+2(\frac{1}{2}-\hat{\beta}-\epsilon)^2}. \tag{4.92}$$

Inserting this expression for η into the right-hand side of (4.91) we obtain the rate of convergence of our scheme as

$$\Big(\frac{\epsilon+\hat\beta+1}{(\frac{1}{2}-\hat\beta-\epsilon)^2}+2\Big)^{-1}.$$

We summarise the above derivation in the following theorem.

Theorem 37. *Let Assumption 22 hold and fix $\epsilon\in(0,1/2-\hat\beta)$. By taking*

$$N(m)=m^{\frac{1}{\epsilon+\hat\beta+1+2(\frac{1}{2}-\hat\beta-\epsilon)^2}},$$

the strong error of our scheme is bounded as follows:

$$\sup_{0\le t\le T}\mathbb{E}[|X_t^{N(m),m}-X_t|]\le cm^{-\Big(\frac{\epsilon+\hat\beta+1}{(1/2-\hat\beta-\epsilon)^2}+2\Big)^{-1}}, \tag{4.93}$$

where constant c depends on ϵ and drift b.

Remark 38. Note that the bound stated in Proposition 35 holds for N large enough so that $\|b^{N(m)}-b\|_{C_TC^{-\beta}}<1$. Formally, the error for small $N(m)$, i.e., small m, could be incorporated into the constant c in (4.93) when the condition $\|b^{N(m)}-b\|_{C_TC^{-\beta}}<1$ is not satisfied. However, when doing numerical estimation of the convergence rate we have to pick m large enough so that $\|b^{N(m)}-b\|_{C_TC^{-\beta}}<1$, otherwise the numerical estimate of the rate could not be reliable.

Corollary 39. *Our construction allows us to achieve any strong convergence rate strictly smaller than*

$$\lim_{\epsilon\downarrow 0}\Big(\frac{\epsilon+\hat\beta+1}{(\frac{1}{2}-\hat\beta-\epsilon)^2}+2\Big)^{-1}=\Big(\frac{\hat\beta+1}{(\frac{1}{2}-\hat\beta)^2}+2\Big)^{-1}=:r(\hat\beta).$$

Remark 40. We rewrite $r(\hat\beta)=\frac{(\frac{1}{2}-\hat\beta)^2}{2(\frac{1}{2}-\hat\beta)^2+\hat\beta+1}$ from Corollary 39. Figure 4.4 in Section 4.6 displays this function over the range of $\hat\beta\in(0,\frac{1}{2})$. We take a closer look at the limits as $\hat\beta$ approaches 0 and $\frac{1}{2}$:

- $\lim_{\hat\beta\downarrow 0}r(\hat\beta)=\frac{1}{6}$. This corresponds to $b\in C_TC^{0+}$, which is comparable to the case of a measurable function $b\in C_TC^0$.

- $\lim_{\hat\beta\uparrow\frac{1}{2}}r(\hat\beta)=0$, so when the roughness of the drift approaches the boundary $\frac{1}{2}$, the scheme deteriorates. This is expected due to the nature of the estimate in Proposition 35 that we use.

A direct comparison of our rate with other rates found in the literature is only possible in the case of measurable drifts with Brownian noise, which has been treated in [7]. The rate obtained there is $\frac{1}{2}$, so our estimate is clearly not optimal. However, the technique they use is different than ours, and in particular they use stochastic sewing lemma due to [30] to drive up the rate, but it seems it is not straightforward to apply the techniques used in [7] to our setting.

Paper [19] considers time-homogeneous distributional drifts and, between others, covers drifts $b \in C^{-\beta}$ for some $\beta > 0$, but the driving noise is a fractional Brownian motion with Hurst index $H \in (0, \frac{1}{2})$. For $\beta \in (0, \frac{1}{2H} - 1)$, the rate of convergence is $\frac{1}{2(1+\beta)} - \epsilon$ for any $\epsilon > 0$. However, within their framework they do not study the case where $H = \frac{1}{2}$ of Brownian noise, in this case they are restricted to measurable drifts.

In [14], the authors consider SDE (4.1) with the drift $b \in C^{\kappa}_T H^{-\beta}_{(1-\beta)^{-1},q}$ for $\kappa \in (\frac{1}{2}, 1)$, $\beta \in (0, \frac{1}{4})$ and $q \in (4, \frac{1}{\beta})$. The space $H^{-\beta}_{(1-\beta)^{-1},q}$ is a fractional Sobolev space of negative regularity $-\beta$. Their β is closely related to our $\hat{\beta}$ as fractional Sobolev spaces $H^s_{p,q}$ are related, albeit different, to Besov spaces $B^s_{p,q}$, see e.g. [47]. In both cases the index s is a measure of smoothness of the elements of the space. When $\beta \downarrow 0$, the convergence rate of the numerical scheme in [14] tends to $\frac{1}{6}$, as in our case. However, when $\beta \to \frac{1}{4}$ the convergence rate in [14] vanishes, while we get $r(\frac{1}{4}) = \frac{1}{22}$.

Finally we would like to mention that in our numerical study in Section 4.6 we obtain a numerical estimate of the convergence rate equal to $\frac{1+(-\hat{\beta})}{2}$, which is the equivalent of the rate $\frac{1+\gamma}{2}$ found in [7] for positive γ, i.e. for γ-Hölder continuous drifts. This suggest that the rate from the latter paper could apply in our setting but it would require a different approach and is left as future work.

4.6 Implementation of numerical methods

In this section we describe an implementation of our numerical scheme and analyze the results obtained. Our implementation was done in the Python programming language and can be found in [10]. We recall that said numerical implementation proceeds in two steps. First we approximate the drift b with b^N as in (4.2). Then we apply classical Euler-Maruyama scheme (4.5) to approximate the SDE with drift b^N. For the numerical examples we will consider a time homogeneous drift, i.e. $b \in C^{(-\beta)+}$.

Since the drift b is a Schwartz distribution in $C^{(-\beta)+}$, producing a numerical approximation of it poses some challenges. Firstly, we cannot simply discretize b and then perform the convolution with the heat kernel to get b^N as in (4.2), see also (3.14), because the discretization of a distribution is not meaningful. Instead, we use that, formally, an element of $C^{(-\beta)+}$ can be obtained as the distributional derivative of a function h in $C^{(-\beta+1)+}$, and for a distribution $\varrho \in \mathcal{S}'$, the convolution and derivative commute as mentioned in [41, Section 5.3] and [25, Remark 2.5], that is,

$$(p_t * \varrho)' = p_t' * \varrho = p_t * \varrho'. \tag{4.94}$$

This fact is useful for efficient construction of regularised SDEs when $b = B_x$ for some function $B \in C^{1/2}_T C^{1-\beta}$ as we do in the numerical example studied in the present.

Indeed, since $b \in C^{-\beta+\epsilon}$ for some $\epsilon > 0$ and $-\beta \in (-1/2, 0)$, we have $\gamma := -\beta+\varepsilon+1 > 0$ and $h \in C^{\gamma}$ is a function. Using (4.94) allows us to discretise the function h first, then convolve it with the heat kernel, which has a smoothing effect, and finally take the derivative.

Without loss of generality we will pick a function h that is constant outside of a compact set, which implies that the distribution h' is supported on the same compact. For example one can choose $L > 0$ large enough so that the solution X_t of the SDE with the drift $b\mathbb{1}_{[-L,L]}$ replacing b will, with a large probability, stay within the interval $[-L, L]$ for all times $t \in [0, T]$, rendering numerically irrelevant the fact that the drift b has been cut outside the compact support $[-L, L]$.

The simplest example of a drift $b \in C^{\gamma-1}$, $\gamma \in (1/2, 1)$, would be given by the derivative of the locally γ-Hölder continuous function $|x|^\gamma$, smoothly cut outside the compact set $[-L, L]$. However, this function is only not differentiable at $x = 0$. Instead, for our numerical implementation we consider $h \in C^\gamma$, which is capable of fully encompassing the rough nature of the drift $b = h'$. This can be obtained if the function h has a 'rough' behaviour in (almost) each point of the interval $[-L, L]$. In this section, we take as h a transformation of a trajectory of a fractional Brownian motion (fBm)[2] $(B^H_x)_{x\geq 0}$. Indeed, it is known that $\mathbb{P}$-almost all paths of fBm B^H are locally γ-Hölder continuous for any $\gamma < H$, see [6, Section 1.6] but paths are almost nowhere differentiable, see [6, Section 1.7]. We construct a compactly supported function h as follows. We take a path $(B^H_x(\omega))_{x\in[0,2L]}$ with Hurst index $H = -\beta + 1 + 2\varepsilon \in (1/2, 1)$ for some small $\varepsilon > 0$, which is thus locally γ-Hölder continuous with $\gamma = -\beta + 1 + \varepsilon$. The constraint $H \in (1/2, 1)$ ensures that $\beta \in (0, 1/2)$, as needed in Assumption 22. We define function $g : \mathbb{R} \to \mathbb{R}$ as

$$g(x) = \big(B^H_x - \frac{B^H_{2L}}{2L} x\big)\mathbb{1}_{\{x\in[0,2L]\}}(x),$$

which ensures that $g(0) = g(2L) = 0$. This transformation, inspired by the so called Brownian bridge, ensures that g is continuous and keeps the same regularity as the fBm B^H. Finally, for convenience, we translate g so that it is supported in $[-L, L]$ rather than $[0, 2L]$. This is done for the sake of symmetry, since we choose to start our SDE from an initial condition close to 0. We choose L large enough so that the path (X_t) has a big chance to stay in the strip $[-L, L]$ up to our terminal time T. In summary, the function h is constructed as

$$h(x) = \Big(B^H_{x+L}(\omega) - \frac{B^H_{2L}(\omega)}{2L}(x - L)\Big)\mathbb{1}_{\{x\in[-L,L]\}}(x), \tag{4.95}$$

and we have $h \in C^\gamma$ with $\gamma = -\beta + \varepsilon + 1$, so that $b := h' \in C^{(-\beta)+}$, and both are supported on $[-L, L]$. By a slight abuse of notation we denote $B^H(x) = h(x)$ so that $b = (B^H)'$.

To compute b^N we apply the semigroup $P_{1/N}$ to b, as in (4.2), which is equivalent to a convolution with the heat kernel $p_{1/N}$. Since the derivative commutes with the convolution as recalled in (4.94) we get

$$\begin{aligned} b^N(x) &:= (P_{1/N}b)(x) = (p_{1/N} * (B^H)')(x) \\ &= (p'_{1/N} * B^H)(x) = \int_{-\infty}^{\infty} p'_{1/N}(y)B^H(x - y)dy. \end{aligned} \tag{4.96}$$

[2]A fractional Brownian motion $(B^H_x)_{x\geq 0}$ with Hurst parameter $H \in (0, 1)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a centered Gaussian stochastic process with covariance given by $\mathbb{E}(B^H_x B^H_y) = \frac{1}{2}(|x|^{2H} + |y|^{2H} - |x - y|^{2H})$. When $H = \frac{1}{2}$ we recover a Brownian motion.

As the derivative of the heat kernel is $p'_{1/N}(y) = -\frac{y}{1/N} p_{1/N}(y)$, we have

$$b^N(x) = -\int_{-\infty}^{\infty} B^H(x-y) \frac{y}{1/N} p_{1/N}(y) dy. \tag{4.97}$$

In order to approximate the above integral numerically, we create a uniform discretization of the interval $[-L, L]$ with $2M+1$ points, denoted by $\mathfrak{M} = \{x_{-M}, ..., x_M\}$, with the mesh size $\delta := \frac{L}{M}$. We sample the fBm $B^H(x)$ in $x \in \mathfrak{M}$ and extend to the real line as a piecewise constant function $\hat{B}^H$ as follows:

$$\hat{B}^H(z) = \sum_{j=-M}^{M-1} B^H(x_j) \mathbb{1}_{\{z \in [x_j - \frac{\delta}{2}, x_j + \frac{\delta}{2})\}}, \qquad z \in \mathbb{R}. \tag{4.98}$$

Inserting (4.98) into (4.97) and performing the change of variable $z = x_i - y$ we obtain a numerical approximation of b^N at any point $x_i \in \mathfrak{M}$:

$$\begin{aligned} b^N(x_i) &\approx -\int_{-\infty}^{\infty} \hat{B}^H(x_i - y) \frac{y}{1/N} p_{1/N}(y) dy \\ &= -\int_{-\infty}^{\infty} \sum_{j=-M}^{M} B^H(x_j) \mathbb{1}_{\{z \in [x_j - \frac{\delta}{2}, x_j + \frac{\delta}{2})\}} \frac{x_i - z}{1/N} p_{1/N}(x_i - z) dz \\ &= -\sum_{j=-M}^{M} B^H(x_j) \int_{x_j - \delta/2}^{x_j + \delta/2} \frac{x_i - z}{1/N} p_{1/N}(x_i - z) dz \\ &= -\sum_{j=-M}^{M} B^H(x_j) \int_{-\delta/2}^{\delta/2} \frac{x_i - x_j - z'}{1/N} p_{1/N}(x_i - x_j - z') dz \end{aligned} \tag{4.99}$$

where we performed a second change of variable $z' = z - x_j$ in the last equality. Let us denote the integral in the last line of (4.99) as

$$\mathcal{I}^N(y) := \int_{-\delta/2}^{\delta/2} \frac{y - z'}{1/N} p_{1/N}(y - z') dz', \tag{4.100}$$

so that equation (4.99) reads

$$b^N(x_i) \approx -\sum_{j=-M}^{M} B^H(x_j) \mathcal{I}^N(x_i - x_j) =: -(B^H * \mathcal{I}^N)(x_i), \tag{4.101}$$

where $*$ denotes the discrete convolution between vectors

$$(B^H(x_i))_{i=-M}^{M} \text{ and } (\mathcal{I}^N(x_i))_{i=-2M}^{2M}.$$

The latter vector needs to be formally defined for a δ-discretization of the interval $[-2L, 2L]$, however, in practice, the values of $\mathcal{I}^N(y)$ decrease quickly to 0 as $|y|$ increases (this is illustrated in Figure 4.1 for two values of $1/N$), so it is enough to use the values of $(\mathcal{I}^N(x_i))$ for $x_i \in \mathfrak{M}$. We will use the above approximation (4.101) in our numerical computations.

At this stage a few remarks regarding the size of $\mathfrak{M}$ are in order. The drift b^N is

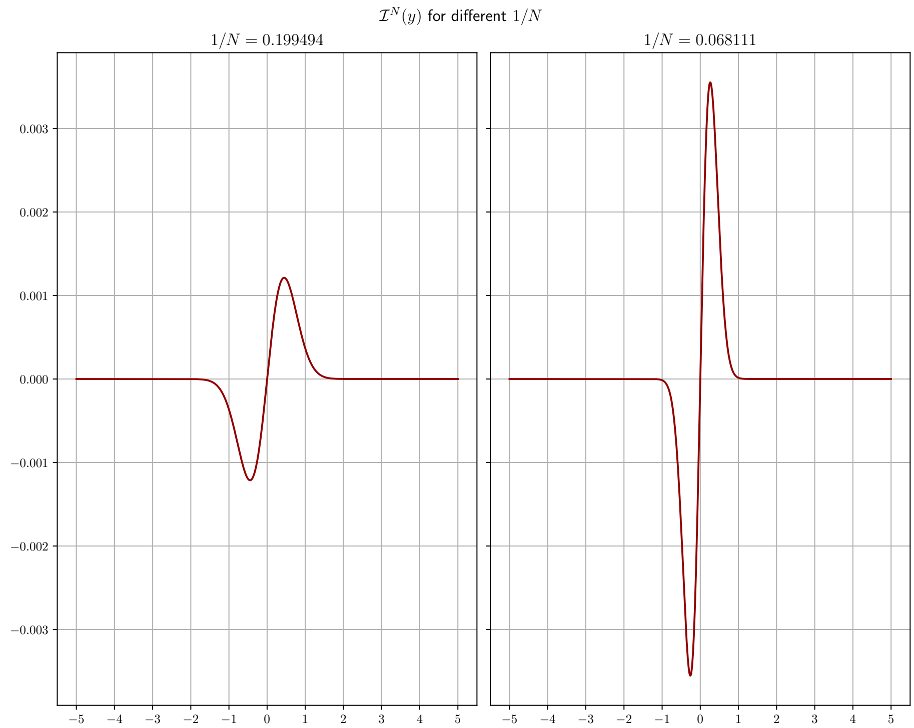

Figure 4.1: From left to right: $\mathcal{I}^N(y)$ for $y \in [-5,5]$ and $1/N = 0.199494, 0.068111$ respectively.

implemented with a discrete convolution given in equation (4.101), where one of the convoluting terms is $\mathcal{I}^N$, which is depicted in Figure 4.1. The peaks of $\mathcal{I}^N$ depend on the magnitude of $1/N$, indeed they become higher and thinner as the variance $1/N$ of the heat kernel $p_{1/N}$ decreases. We choose $N = N(m) = m^\eta$ for some $\eta > 0$ given in equation (4.92), so as the number of steps m in the Euler scheme increases, the variance $1/N(m) = m^{-\eta}$ decreases, leading to a vector $\mathcal{I}^{N(m)}$ of mostly zeros up to numerical precision. If the discretization $\mathfrak{M}$ is too coarse compared to $1/N(m)$, the vector $\mathcal{I}^{N(m)}$ does not retain enough information. To avoid this issue we have to make sure that there are sufficiently many points in $\mathfrak{M}$, i.e. that $2M$ is large enough compared to the size of the support $[-L, L]$. More specifically, we require that the distance δ between points in the grid is less than the standard deviation of the heat kernel $\sqrt{1/N(m)}$, i.e. $\delta < \sqrt{1/N(m)} = m^{-\eta/2}$ for every m considered in numerical computations, which leads to

$$2M > 2Lm^{\eta/2}. \tag{4.102}$$

As an example, let us consider a discretization of the interval $[-5,5]$ (hence $L = 5$) and use $m = 2^{12}$ points in the Euler-Maruyama scheme. We will need at least $2M + 1 > 2Lm^{\eta/2} + 1 = 10 \cdot 2^{6\eta} + 1$ points in $\mathfrak{M}$, where the value η depends on β and it is given explicitly in equation (4.92). The smallest admissible number of points $2M + 1$ of the discretization $\mathfrak{M}$ for varying β is given in Table 4.6.

We now summarise the procedure to compute numerically an approximation of the drift $b^{N(m)}$.

β	10^{-6}	$1/16$	$1/8$	$1/4$	$1/2$
$1/N(m) = m^{-\eta(\beta)}$	0.001288	0.001398	0.001489	0.001768	0.003906
$2M+1$	279	269	261	239	161

Table 4.1: Minimum amount of points $2M+1$ needed in the discretization $\mathfrak{M}$ when $m = 2^{12}$ and $L = 5$. This varies according to the variance $1/N$, which in turn is a function of β

.

1. Fix $\hat{\beta}, L$ and the number m of steps of the Euler scheme. Compute the smallest integer M that satisfies (4.102). Define the discretization $\mathfrak{M}$ and the corresponding mesh size δ.
2. Simulate a single path of fBm on the interval $[0, 2L]$ with mesh size δ and with a given Hurst parameter $H = -\beta + 1 + \epsilon$ for some small $\epsilon > 0$.
3. Transform the path of fBm into a bridge on $[-L, L]$ by applying the transformation (4.95), to get a vector $B^H(x_i)$, for $x_i \in \mathfrak{M}$.
4. Compute numerically the integral $\mathcal{I}^{N(m)}(x_i)$ for $x_i \in \mathfrak{M}$.
5. Perform the discrete convolution $-(B^H * \mathcal{I}^{N(m)})$ as in (4.101) to approximate numerically $b^{N(m)}(x_i)$ for $x_i \in \mathfrak{M}$.
6. Extend $b^{N(m)}(x)$ for all $x \in [-L, L]$ by linear interpolation.

One can note that the creation of a drift depends on exactly two parameters as opposed as in more classical examples where the drift is given as a function of $(t, x) \in [0, T] \times \mathbb{R}^d$. These parameters being the Hurst parameter H or more precisely for this discussion β, which will give us the regularity of the drift on x, and interestingly the number of time steps m that will be used in the numerical scheme by means of the heat kernel parameter $m^{-\eta}$.

Since the drift is created based on the simulation of a fractional Brownian motion, the one has to be aware of how to handle the random nature of the simulation for reproducibility purposes. This of course poses no challenge and it is just a matter of selecting a particular seed to generate the drift. Without selecting a seed for the simulation we will have a random drift with the selected regularity, which for our purposes is in fact beneficial since we can see how different drifts living in the same space will affect the rate of convergence. With this in mind, one has to also consider that having different drifts in different simulations effectively means that we are simulating different SDEs with solutions X^N, not only one SDE, which we will discuss in a moment.

For now, let us discuss the generation of a drift with a given regularity $-\beta$ but using a different number of time steps. We need to do this, because in the computation of the error we require to compare the "real solution" to the SDE with the approximation. Recall that for an SDE with a classical drift, this drift would be a given function which is fixed across different approximations, whereas we have that for different amount of time steps, the smoothing given by the heat kernel will change. In this sense, the numerical scheme we propose has two things that are improving by decreasing the length of the time steps, the solution of the SDE itself and the drift.

We can observe in figure 4.2 how the different levels of smoothing will change dramatically the behaviour of the drift. Primarily, recall that the drift is formally defined as the derivative of a fractional Brownian motion $b = B_x^H$, and note how the figure shows negative values of the drift when the slope of the fBm is roughly negative. This is observable with both of the drifts, but note how the one with a higher level of smoothing the oscillation is much more violent, which is to be expected since we are approximating the drift by $b^N = P_{1/N}b$ for $N = m^\eta$, so as N increases b^N gets closer to the distribution b, and it has to become more and more rough. Not only that, but also this roughness is indeed better capturing the behaviour of the derivative of the fBm. We know that a path of fBm is self-similar and not differentiable, so in principle its derivative has a negative and fractional differentiability order, which indicates us that any attempt to simulate the derivative has to reflect the fluctuations on the path.

Moreover, figure 4.3 shows a similar story with a different parameter of regularity on the drift—or different Hurst parameter—namely $-\beta \approx 1/2$. This is obviously a different fractional Brownian motion, since we are using a new H, however the features to emphasize of this plot are the massive change on the magnitude of the drifts and the frequency of the oscillations compared with figure 4.2. This is to show that our numerical implementation is able to simulate drifts with different regularities and reflect said regularity.

We know that the further we go in β the space $C^{-\beta}$ contains distributions with less regularity, and indeed even their approximations have to be rougher from the start. Which is why the values of N are different on each of the pictures—given by the dependency of η on β from (4.92)—even though both correspond to the same value of m.

Once we have a numerical approximation of the drift coefficient, the remaining step is to apply the standard Euler-Maruyama scheme using the numerical drift calculated and interpolating linearly so we indeed have a function we can evaluate at every point of the discretized time grid. To calculate the empirical rate of convergence of the numerical scheme we must have approximations with an increasing number of steps m, as well as a proxy of the real solution, since the real solution is unknown in a closed form. This real solution proxy is the most expensive run of the iterative scheme, so we compute it first and store it in memory in order to compare with the rest of approximations. The strong error of the scheme is calculated by Monte Carlo approximation of the L^1 norm of the difference between the approximations with less time steps and the proxy of the real solution at time T. Recall that as we discussed above, different amount of time steps will generate distinct drifts, and therefore in practice what is fixed for each run of the Euler-Maruyama scheme is the fractional Brownian motion, from which we create the drift using the corresponding smoothing parameters to each time step. The procedure to compute the strong error for a given fBm reads as follows:

1. Choose a 'large' m for the proxy of the real solution, and $m_i \ll m$, $i = 0, \ldots, I$, for the approximations and such that each m_i divides m. Choose a number of sample paths $Q \in \mathbb{N}$ sufficiently large for a Monte Carlo procedure.

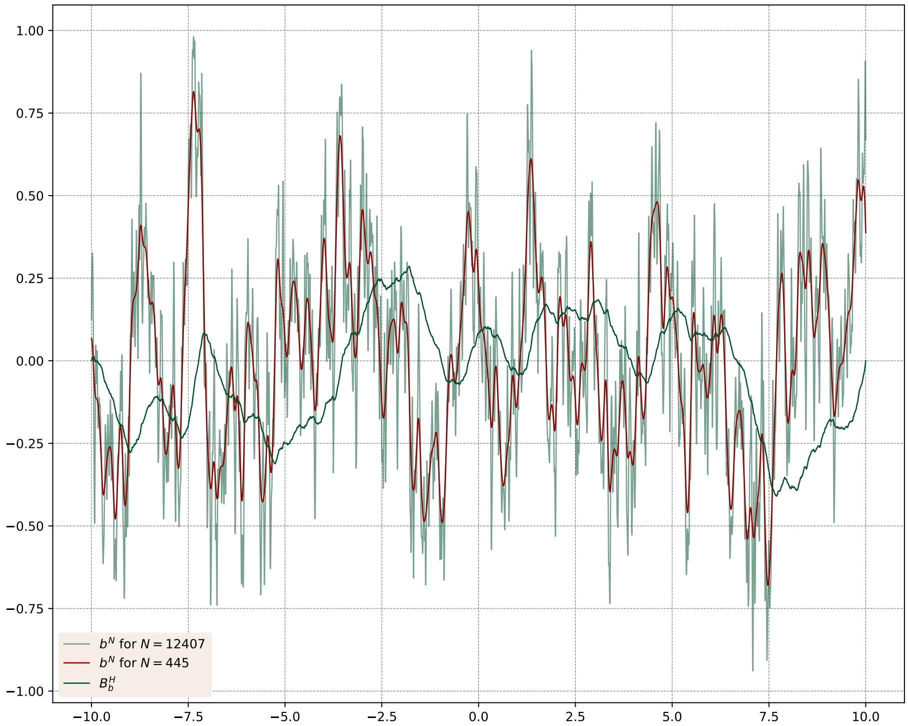

Figure 4.2: Comparison of and fBm with Hurst parameter $H = 1 - \epsilon$ for $\epsilon = 10^6$ and the drifts generated with 2^{11} and 2^{17} time steps which translates into $N = 445$ and $N = 12407$ respectively. The drift b has regularity $-\beta$ for $\beta = \epsilon$.

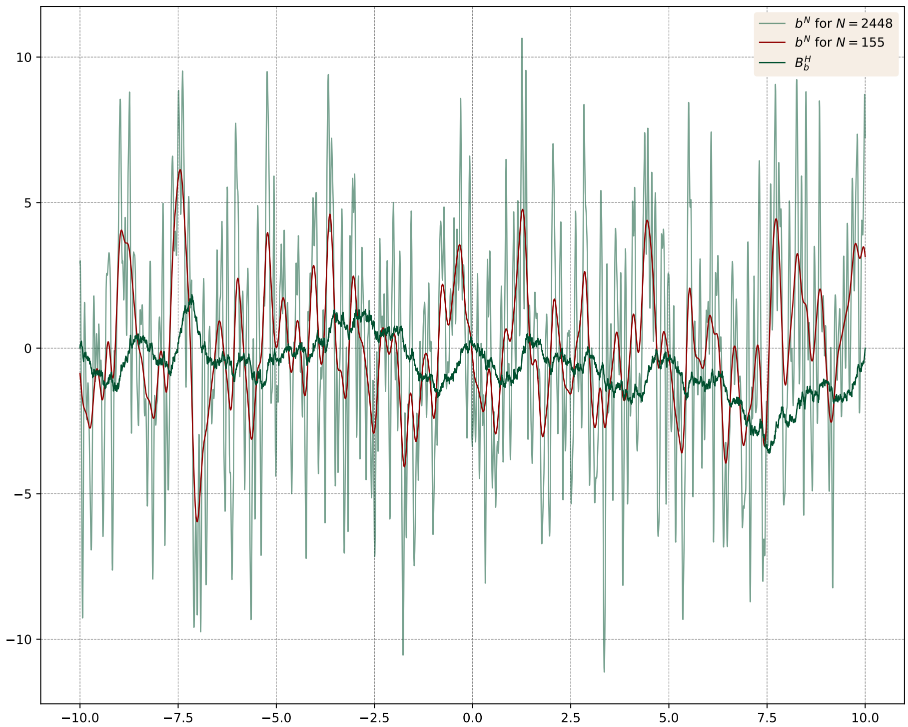

Figure 4.3: Comparison of and fBm with Hurst parameter $H = 1/2 + \epsilon$ for $\epsilon = 10^6$ and the drifts generated with 2^{11} and 2^{17} time steps which translates into $N = 155$ and $N = 2448$. The drift b has regularity $-\beta$ for $\beta = 1/2 - \epsilon$.

2. Denote by $\Delta t_i = T/m_i$ the time-step for the approximated solutions, where T is the terminal time.

3. Define $\mathfrak{M}$ with M points, where M and m satisfy (4.102). Generate one fBm path on the discrete grid $\mathfrak{M}$.

4. Run the Euler-Maruyama scheme for the proxy solution and for the approximated solutions up to time T, with the same Q Brownian motion paths.

5. Compute the strong error between the proxy solution corresponding to m and the approximations corresponding to m_i, by calculating a Monte Carlo average of the absolute differences between computed solutions at time T across the Q sample paths. Denote this approximations of the strong error by ϵ_i, $i = 0, \dots, I$.

6. As we expect $\epsilon_i \approx c{m_i}^{-r} = c(\Delta t_i)^r$, where r is the convergence rate, we compute the empirical rate r by performing a linear regression of $\log_{10}(\epsilon_i)$ on $\log_{10}(\Delta t_i)$, $i = 0, \dots, I$.

With this way to approximate the error the analysis of the rate of convergence is explored using two different approaches: the classical way comparing for a single drift the convergence depending on the number of time steps, and one comparing the rates of convergence across several levels of regularity of the drift. The first one has to obviously be done for different values of β, in fact the later analysis is not possible before doing the former first.

Finally, we performed a further experiment to better compare the empirical rate with the theoretical results. Since the drift of the SDE is obtained running a single path of a fBm, and clearly there is randomness there, we decided to run the algorithm for 50 different paths, for each value β, and then we computed the average of the empirical convergence rates as well as its 95% confidence interval. We compared this with the theoretical rate obtained in Theorem 37 and with the conjecture that the rate should be $1/2 - \beta/2$. The latter would be the natural extension of the results of [13] if they could be extended into the case of distributions, in particular with $-\beta \in (-1/2, 0)$ which is the case we treat here. We collected the results in Table 4.6 below and also plotted them in Figure 4.4. This experiment strongly suggests that our theoretical result is not optimal, and that the convergence rate indeed could be $1/2 - \beta/2$. Further studies are needed to prove or disprove this conjecture.

β	$\varepsilon = 10^{-6}$	$1/8$	$1/4$	$3/8$	$1/2 - \varepsilon$
Empirical rate	0.50814	0.44063	0.38163	0.32318	0.28275
$1/2 - \beta/2$	0.49999	0.43750	0.37500	0.31250	0.25000
Theoretical rate	0.16666	0.10000	0.04545	0.01111	0.00000

Table 4.2: Average of empirical convergence rates obtained using 50 Euler-Maruyama approximations with 10^4 sample paths for each β, the conjecture rate of $1/2 - \beta/2$ and the theoretical rate from Theorem 37. All values rounded to 5 decimal places. Same values are plotted in Figure 4.4.

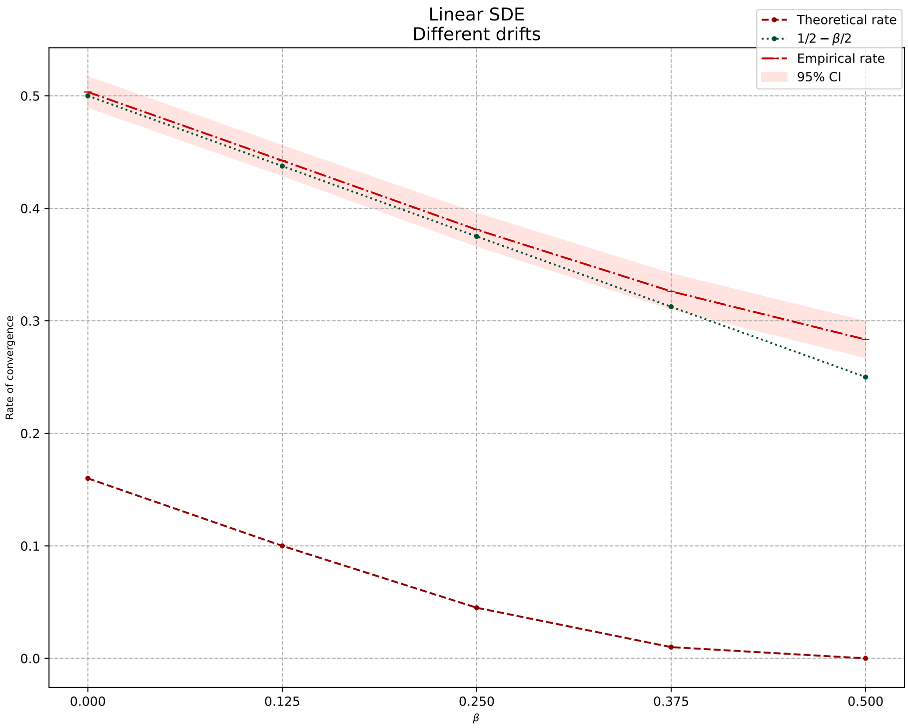

Figure 4.4: Average empirical convergence rates obtained from 40 Euler-Maruyama with drifts of the same regularity but generated for each run. Approximations have 10^4 sample paths for each β including 95% confidence interval for the average, the conjectured rate of $1/2-\beta/2$ and the theoretical rate obtained in Theorem 37.

4.7 Fokker-Planck equation and the law density of X

A generalization of equation (4.1) can be obtained if we admit a random initial condition, i.e.

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t b(s, \tilde{X}_s) ds + W_t, \tag{4.103}$$

where $\tilde{X}_0$ is a random variable and $\rho(t, \cdot)$ is the law density of $\tilde{X}_t$. Although throughout this chapter we have been considering (4.1), with a deterministic initial condition, we will argue that in the same setting we can arrive to (4.103) via a superposition argument—which we will develop below. We can associate Fokker-Plank PDE to (4.103)

$$\begin{cases} \partial_t \rho(t,x) = \frac{1}{2}\partial_x^2 \rho(t,x) - \partial_x[b\rho](t,x) \\ \rho(0,x) = \rho_0. \end{cases} \tag{4.104}$$

In smooth cases, as well as when b is a distribution of the class considered in this thesis, it is known that the density of the solution of X solves the Fokker-Planck PDE (4.104). Moreover, the latter has a unique solution—in a certain class—hence the unique solution to (4.104) will be the density of X. Also, see that (4.104) is a special case of a Fokker-Planck PDE studied in [26] associated to a McKean-Vlasov SDE, we will go into more details the McKean-Vlasov SDE and its related Fokker-Planck equation in Chapter 5. Note that in (4.104) the initial condition of the SDE cannot be deterministic, but instead it must be

$\tilde{X}_0 \sim \mu$ for some probability measure μ which admits a density because we cannot have the initial condition ρ_0 for the Fokker-Planck PDE to be a Dirac delta due to the regularity assumptions imposed in [26].

SDE (4.103) is akin to the main SDE from [27] where the authors describe in Theorem 4.5, in the setting of a martingale problem, that a superposition argument can be made to justify the usage of a more generic density to what we have—a Dirac delta. The following argument is almost identical to the one made in the proof of the aforementioned theorem.

Suppose that instead of the initial condition $\tilde{X}_0 = x$ that we have in (4.1) we say that $\tilde{X}_0 \sim \mu$ where μ is some probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and we define the push-forward measure $\nu(B) := \mu(\psi(0, B))$ of μ through ϕ, since we know $\phi = \psi^{-1}$, for any $B \in \mathcal{B}(\mathbb{R})$, then ν will be the initial condition for the process $\tilde{Y}_t = \phi(t, \tilde{X}_t)$. By Kolmogorov's extension theorem we have the canonical process $\tilde{Y}_t(\omega) = \omega(t)$ on the underlying probability space, so $\Omega = C_T$, and $\mathbb{P}^y$ be a law of $\tilde{Y}$ on C_T and such that $(\tilde{Y}, \mathbb{P}^y)$ is the weak/mild solution of (3.44) with initial condition $\tilde{Y}_0 = y$. The probability measure given by

$$\mathbb{P}(E) := \int \mathbb{P}^y(E)\nu(dy) \tag{4.105}$$

for $E \in \mathcal{B}(C_T)$ is well defined and if $\tilde{X}_t = \psi(t, \tilde{Y}_t)$, we can conjecture that $(\tilde{X}, \mathbb{P})$ is the solution to the martingale problem with singular drift b and initial condition μ. See that for any $E' \in \mathcal{B}(C_T)$ such that $E' = \{\omega : \omega_0 \in B\}$ for some $B \in \mathcal{B}(\mathbb{R})$ we have

$$\mathbb{P}^y(E') = \mathbb{P}^y(\tilde{Y}_0 \in B) = \mathbb{1}_B(y) \tag{4.106}$$

considering that $\mathbb{P}^y$-a.s. $Y_0 = y$. Let us set $B = \phi(0, A)$ for any $A \in \mathcal{B}(\mathbb{R})$ and we have

$$\mathbb{P}(\tilde{X}_0 \in A) = \mathbb{P}(\psi(0, Y_0) \in A) = \mathbb{P}(Y_0 \in \phi(0, A)) = \mathbb{P}(Y_0 \in B) \tag{4.107}$$

using the definition of $\mathbb{P}$ (4.105) and (4.106)

$$\begin{aligned} \mathbb{P}(\tilde{Y}_0 \in B) = \mathbb{P}(E') &= \int \mathbb{P}^y(E')\nu(dy) = \int 1_B(y)\nu(dy) \\ &= \int_B \nu(dy) = \nu(B) = \nu(\phi(0, A)) = \mu(\psi(0, \phi(0, A))) = \mu(A), \end{aligned} \tag{4.108}$$

where the penultimate equality comes from the definition of ν. Putting (4.107) and (4.108) together we have that $\mathbb{P}(X_0 \in A) = \mathbb{P}(E') = \mu(A)$. So that indeed X_0 is distributed as μ.

Given the discussion above, we are now in position to solve an SDE with a random initial condition in our setting such as (4.103), and to use the Fokker-Planck equation (4.104) to obtain the law density ρ of the solution. It is possible to obtain the law of the SDE by solving this PDE numerically so we can compare it with a density histogram of the solution to the SDE at terminal time $t = T$.

The drift of the SDE is created once as mentioned above and used both in the iterative Euler scheme that is used to solve the SDE numerically, but then it is also separately fed to the PDE solver. In Figures 4.5 and 4.6, where the initial condition for the SDE is a sample from a standard normal distribution, we can see how the PDE solution adjusts closely to the histogram, which is a sign of both the Euler scheme and the PDE solver working appropriately. This comparison is the normalized histogram of the solution to the SDE at time $t = T = 1$ against the solution of the Fokker-Planck equation at the same time for two different drifts, one which is close to the regime of measurable functions and one on the end of regularity $\beta \approx 1/2$.

Notice how for a more regular SDE the density looks "smoother", indicating the value at terminal time of the solution is distributed somewhat evenly as opposed to the case where the drift is less regular in figure 4.6—in this case with $\beta \approx 1/2$—where, although we do have most of the observations of X_T being around zero and indeed much more concentrated than in the previous picture, there are additional peaks on different locations.

We conclude this chapter by showing the results of performing a goodness of fit test,

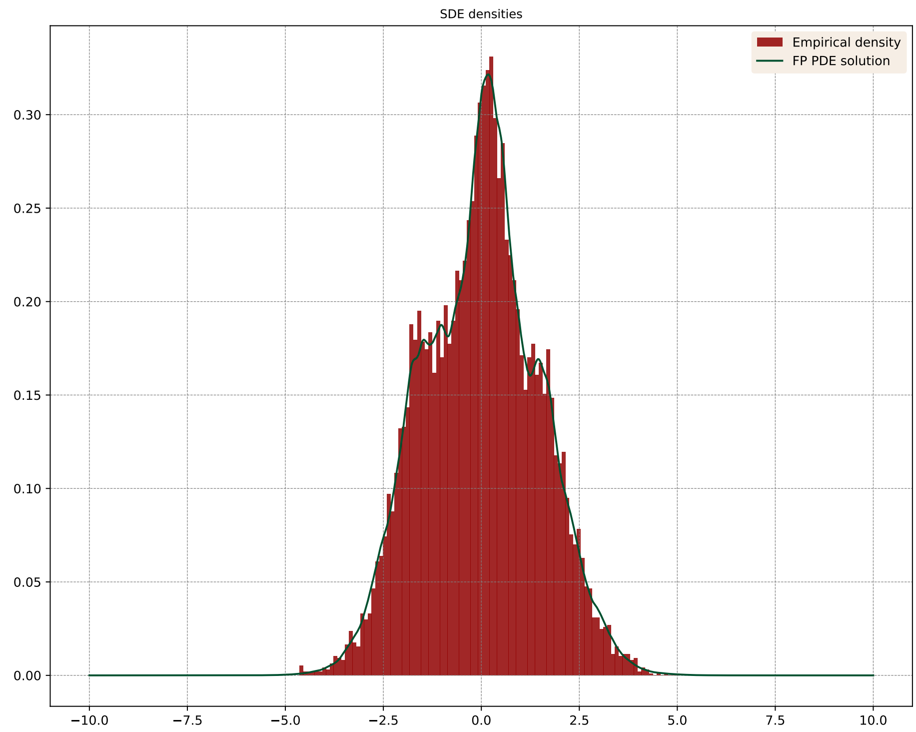

Figure 4.5: Law density of the SDE obtained by solving the associated Fokker-Planck PDE (green line) and the histogram of the empirical density (red bars) at time $t = T$ for an SDE with drift derived from a fBm with Hurst parameter $H = 1 - \epsilon$, or a drift in $C^{-\epsilon}$, for $\epsilon = 10^{-5}$.

using the Kolmogorov-Smirnov test for one sample, in order to compare the empirical density obtained with the solution to the Fokker-Planck equation. Recall that ρ^N is the law density of the solution, and it is stored as a numerical array, and that to use the Kolmogorov-Smirnov test we require the cumulative distribution function. We can achieve this simply by integrating numerically, and then we can use the array to make the comparison to obtain the results presented in Table 4.7.

In a few words, we know that the Kolmogorov-Smirnov test is considered passed when the K-S statistic is small and the p-value is large. This is the case for all our tests, and this acts as a sanity check for the generation of densities through the Fokker-Planck equation, but it additionally offers us a glimpse into a very interesting phenomenon in that occurs

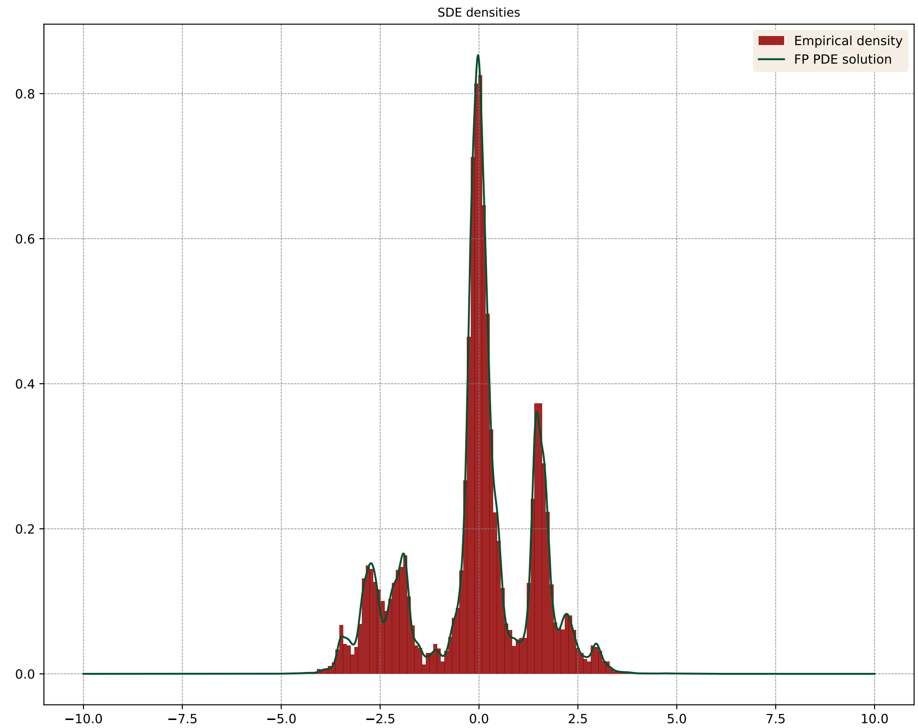

Figure 4.6: Law density of the SDE obtained by solving the associated Fokker-Planck PDE (green line) and the histogram of the empirical density (red bars) at time $t = T$ for an SDE with drift derived from a fBm with Hurst parameter $H = 1/2 + \epsilon$, or a drift in $C^{-1/2-\epsilon}$, for $\epsilon = 10^{-5}$.

Time steps	K-S statistic	p-value
2^{11}	0.006513891887619694	0.787279319179895
2^{9}	0.006246209815441578	0.8276968493111533
2^{7}	0.006513891890088164	0.7872793187950878

Table 4.3: Results of the Kolmogorov-Smirnov test for the goodness of fit of the PDE approximation of the law density with respect to the empirical density for SDEs with drift in the space $C^{-1/4}$ and time steps $2^{11}, 2^{9}, 2^{7}$.

with the increase of time steps. At first sight, one may be inclined to think that by increasing the amount of time steps and therefore being closer to the real solution—both for the SDE and the PDE—the fit of the distribution would be better and hence the K-S test statistic and the p-value would decrease and increase, respectively. However, in the example presented we see that this is not necessarily the case. For instance for 2^{11} time steps we have a slightly worse p-value than for 2^{9}, when one would potentially expect for the former to be slightly larger if anything. We explain this situation, with the fact that as we increase the number of time steps, the parameter N of the sequence b^N will change making both the SDE and the PDE more difficult to solve, hence explaining the difficulty for the goodness of fit to improve. Effectively, since both of the elements that are being compared are numerical approximations, they have an inherent error associated and this has the potential to be amplified if other sources of error converge.

Chapter 5

McKean SDEs with irregular drift

5.1 Introduction

In this chapter we are concerned with numerical schemes for McKean-Vlasov SDEs of the form

$$\begin{cases} X_t = X_0 + \int_0^t F(\rho(s, X_s)) b(s, X_s) ds + W_t \\ \rho(t, \cdot) \text{ is the law of} X_t, \end{cases} \tag{5.1}$$

for some initial condition X_0 which admits a density $\rho_0(t, \cdot)$, where W is a 1-dimensional Brownian motion, b is a Schwartz distribution, F is a suitable non-linear function. SDEs where the solution depends on the law of the unknown itself are called McKean-Vlasov, and for the case we are interested in the dependence is on the density of this law in a pointwise form like in (5.1). Such SDEs arise for example as limit of moderately interacting particle systems, as shown in [37, 36] for smooth coefficients. Here we are in a singular setting since the coefficient b is assumed to be a Schwartz distribution, in particular it is an element of a negative Besov space $C^{-\beta}$ with $\beta \in (0, \frac{1}{2})$, more details below. The theoretical well-posedness and other properties of solutions to (5.1) have been studied in [26] and here we are concerned in devising a numerical scheme and studying its convergence.

5.2 Problem statement

We start by making the following assumptions, which will be in force throughout the paper.

Assumption 41. Let $\beta \in (0, 1/2)$ and $b \in C_T^{1/2} C^{-\hat\beta}$ for some $0 < \hat\beta < \beta$.

Assumption 42. Let F be a function and $x \mapsto \tilde F(x) := xF(x)$ be such that

- F is bounded and differentiable, and $\partial_x F$ is Lipschitz and bounded,
- $\tilde F$ is differentiable, and $\partial_x \tilde F$ is Lipschitz and bounded.

Assumption 43. Let $\rho_0 \in C^{1+\nu}$ for some $\nu > 0$.

Remark 44. The convergence rate will be expressed in terms of β and the smaller this parameter is, the faster the numerical scheme will be shown to converge. The reader should think of β as being close to $\hat{\beta}$. For technical reasons, we require that $\beta - \hat{\beta} < \nu \wedge (1-\beta)$. Additionally, we remark that Assumption 41 is often written in an equivalent way: $b \in C_T^{1/2} C^{(-\beta)+}$, where $C_T^{1/2} C^{(-\beta)+} := \bigcup_{-\beta' > -\beta} C_T^{1/2} C^{-\beta'}$ recalling the notation introduced in Section 3.1.

In order to solve the McKean-Vlasov SDE (5.1), a key tool is the associated Fokker-Planck equation

$$\begin{cases} \partial_t \rho = \frac{1}{2}\Delta\rho - \partial_x[\tilde{F}(\rho)b], \\ \rho(0) = \rho_0. \end{cases} \tag{5.2}$$

We understand the Fokker-Planck equation (5.2) through its mild formulation—as we have been doing with many PDEs in this work—namely a function ρ is a solution to (5.2) if $\rho \in C_T C^\alpha$ for some $\alpha > \beta$ and it solves the integral equation

$$\rho(t) = P_t \rho_0 - \int_0^t P_{t-s} \left[\partial_x \left[\tilde{F}(\rho(s)) b(s)\right]\right] ds, \qquad t \geq 0. \tag{5.3}$$

We note that one can also study a weak formulation of the PDE (5.2). However, it was shown in [26, Proposition 3.2] that mild and weak formulations are equivalent, and as we do not need the properties of weak solutions, will not go deeper in their study for this work.

Proposition 45 ([26, Proposition 3.6])**.** *Under Assumptions 41, 42 and 43, PDE (5.2) has a unique mild solution in the space $C_T C^\alpha$ for any $\alpha \in (\beta, 1-\beta)$.*

The above result is proved in [26] under weaker assumptions, namely $\rho_0 \in C^\alpha$ and $b \in C_T C^{-\hat{\beta}}$ as there is no need for Hölder regularity of the drift b in time.

We are now in a position to formally define what we mean by a solution to (5.1) following [26, Definition 6.2].

Definition 46. A triple $(X, \rho, (\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}))$, in short written as $(X, \rho, \mathbb{P})$, is a *solution to McKean-Vlasov SDE* (5.1) if

- X is a stochastic process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$,
- $(X, \mathbb{P})$ is a solution to the *martingale problem* with distributional drift $B := F(\rho)b$ and $X_0 \sim \rho_0$ in the sense of Definition 17,
- $\rho \in C_T C^\beta$, and $\rho(t)$ is the density of X_t for all $t \in [0, T]$.

This problem admits *uniqueness* if for any two solutions $(X, \rho, \mathbb{P})$ and $(X', \rho', \mathbb{P}')$ the densities ρ, ρ' coincide as elements of $C_T C^\beta$ and the laws of $(X, \mathbb{P})$ and $(X', \mathbb{P}')$ are identical.

Theorem 47 ([26, Theorem 6.3 and Theorem 4.2])**.** *Let Assumptions 41, 42 and 43 hold. There is a unique solution to McKean-Vlasov SDE (5.1) and ρ is a mild solution to PDE (5.2).*

Our numerical approach to approximate the McKean-Vlasov SDE (5.1) is similar to the one we use in Chapter 4 but we will make a brief reminder in what follows for convenience of the reader with the necessary changes for this setting. We start by replacing b with

$$b^N := P_{\frac{1}{N}} b \tag{5.4}$$

and then applying Euler scheme to solve the resulting SDE. Since the density of the solution appears in the equation, and we have changed the drift b by b^N, we must start with the study of the solution to the following regularized Fokker-Planck equation

$$\begin{cases} \partial_t \rho^N = \frac{1}{2}\Delta\rho^N - \partial_x[\tilde{F}(\rho^N)b^N], \\ \rho^N(0) = \rho_0. \end{cases} \tag{5.5}$$

As before with equation (5.2), Proposition 45 guarantees the existence of a unique mild solution for the regularized PDE (5.5) which reads

$$\rho^N(t) = P_t\rho_0 - \int_0^t P_{t-s}\left[\partial_x\left[\tilde{F}(\rho^N(s))b^N(s)\right]\right] ds, \qquad t \in [0,T]. \tag{5.6}$$

Furthermore, this solution converges to ρ in $C_T C^\alpha$ as $N \to \infty$ for any $\alpha \in (\beta, 1-\beta)$ by [26, Proposition 4.4], more succinctly

$$\lim_{N\to\infty} \|\rho^N - \rho\|_{C_T C^\alpha} \to 0. \tag{5.7}$$

Denote the drift coefficient of the McKean-Vlasov SDE (5.1) by $B := F(\rho)b$ and its smooth approximation by $B^N := F(\rho^N)b^N$. Let $(X, \rho, \mathbb{P})$ be the solution to (5.1). On the same probability space, we consider the regularised McKean-Vlasov SDE:

$$X_t^N = X_0 + \int_0^t B^N(s, X_s^N)ds + W_t. \tag{5.8}$$

Importantly, note that X_0 is the initial value of the solution X and (W_t) is the Brownian motion driving X. Equation (5.8) has a unique strong solution X^N with density ρ_N.

We discretize the time interval $[0,T]$ into m equal subintervals and we set $t_i := i\frac{T}{m}$ for $i = 0, \dots, m$. The Euler-Maruyama approximation at time $t \in (t_i, t_{i+1}]$ of the SDE (5.8) will be

$$X_t^{N,m} = X_{t_i}^{N,m} + B^N(t_i, X_{t_i}^{N,m})(t - t_i) + W_t - W_{t_i}. \tag{5.9}$$

Our strategy is to identify how N depends on m so that we maximize our bound of the strong rate of convergence of $X^{N,m}$ to X. Once this is done, the Euler-Maruyama approximation will only depend on m and we denote it by $\hat{X}_t^m$ for simplicity. The main theoretical result of this work is formulated below and will be proven in Section 5.6.

Theorem 48. *Let Assumptions 41, 42 and 43 hold and* $0 < \epsilon < \frac{1}{2} - \beta$ *be fixed. Setting*

$$N(m) = m^{\frac{1}{\epsilon+\beta+1+2(\frac{1}{2}-\beta-\epsilon)^2}},$$

optimizes our bound for the strong rate of convergence and leads to the rate

$$\sup_{0\le t\le T} \mathbb{E}\left[\left|\hat{X}_t^m - X_t\right|\right] \le cm^{-\left(\frac{\epsilon+\beta+1}{(1/2-\beta-\epsilon)^2}+2\right)^{-1}}, \tag{5.10}$$

where constant c depends on ϵ and the drift b.

Notice that the convergence rate is the same as in the main result of Chapter 4, Theorem 37, which is not surprising considering that the drifts of equations (4.1) and (5.1) live in the same space of distributions. However, consider as well that our conclusion in the previous chapter is that this convergence rate is not optimal and has a significant chance of being improved by using different techniques to find it.

We briefly illustrate our strategy for the proof of the main result on convergence rate, which is the same basic idea we use in Chapter 4 for the computation of the error, i.e. we split the error into 2 terms

$$\sup_{t\in[0,T]} \mathbb{E}|\hat{X}_t^m - X_t| = \sup_{t\in[0,T]} \mathbb{E}|X_t^{N,m} - X_t| \le \sup_{t\in[0,T]} \mathbb{E}|X_t^{N,m} - X_t^N| + \sup_t \mathbb{E}|X_t^N - X_t|, \tag{5.11}$$

where the first one is the error of an Euler scheme applied to a smoothed SDE (called *Euler error* thereafter), while the second term is controlled with stability estimates on the SDE (called *stability error* thereafter). Clearly, since the drift B^N of the SDE (5.8) is smooth, a standard result on the convergence of Euler scheme could be applied to control the Euler error. The issue however, is that the constant in front of the rate depends on the parameters in an exponential way, in particuar it explodes as $N \to \infty$ as fast as $e^{\|B^N\|_\infty}$. Thus new estimates are needed, and these were establised for SDEs (not of McKean-Vlasov type) in [9]. These estimates involved the norms and seminorms $[B^N]_{\frac{1}{2},L^\infty}$, $\|B^N\|_{C_T L^\infty}$ and $\|B_x^N\|_{C_T L^\infty}$ The reader may notice that this decomposition of the error is not considering any sort of numerical approximation of the solution to the associated Fokker-Planck equation ρ^N, which may imply that we are able to get an exact solution for it. We do not prove that we can obtain an exact ρ^N as a solution in order to avoid a rather circuitous route of analytical results, and we are able to make use of the probabilistic nature of the problem in order to solve it. We trade this by proving that we are able to construct a numerical approximation $\hat{\rho}^N$ such that the convergence rate of the numerical scheme remains unchanged, this is presented in the last section of this chapter but for the next sections it may seem like we are considering ρ^N is obtained exactly.

5.3 Analytical results

In this section we explore a few results that lay slightly outside of the probabilistic aspects of the problem but are fundamental for the convergence results of the numerical methods that we are after.

First, note that we can see the function $\tilde{F}$ as an operator on C^α for $\alpha \in (0,1)$ and study some mapping properties. We also denote the operator as $\tilde{F}$, so that for $f \in C^\alpha$ we denote $\tilde{F}(f) := \tilde{F}(f(\cdot))$. The properties we need are given in the following result.

Lemma 49 ([23, Prop. 3.1])**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function such that φ_x is Lipschitz and bounded. For $\alpha \in (0,1)$ define the operator $\varphi : C^\alpha \to C^\alpha$ as $\varphi(f)(\cdot) := \varphi(f(\cdot))$ for any $f \in C^\alpha$. Then there is a constant $c_\varphi > 0$ such that for all $f, g \in C^\alpha$*

$$\|\varphi(f) - \varphi(g)\|_{C^\alpha} \le c_\varphi (1 + \|f\|^2_{C^\alpha} + \|g\|^2_{C^\alpha})^{1/2} \|f - g\|_{C^\alpha}$$

and

$$\|\varphi(f)\|_{C^\alpha} \le c_\varphi (1 + \|f\|_{C^\alpha}).$$

So, when we define the mild solution of the Fokker-Planck equation what we use is not the function $\tilde{F}$ but the operator.

Lemma 50. *For $f \in C_T C^\xi$ for some $\xi \in \mathbb{R}$ let us define*

$$g(t) := \int_0^t P_{t-s} f(s) ds. \tag{5.12}$$

Then $g \in C_T^{1/2} C^\zeta$ for any $\zeta < \xi + 1$. Furthermore, $\|g\|_{C_T^{1/2} C^\zeta} \le c_1 \|f\|_{C_T C^\xi}$, where

$$c_1 = c_1(T, \xi, \zeta) = c_S^1 \left(\frac{T^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}} + 2 + c_S^2 \frac{T^{1-\frac{1+\zeta-\xi}{2}}}{1 - \frac{1+\zeta-\xi}{2}} \right).$$

Proof. For clarity, we recall the definition of the norm to bound:

$$\|g\|_{C_T^{1/2} C^\zeta} = \|g\|_{C_T C^\zeta} + [g]_{1/2, C^\zeta} = \sup_{t \in [0,T]} \|g(t)\|_{C^\zeta} + \sup_{t \neq r \in [0,T]} \frac{\|g(t) - g(r)\|_{C^\zeta}}{|t - r|^{1/2}}. \tag{5.13}$$

We will estimate each term separately. For the first term in (5.13), using Jensen's inequality we have

$$\|g(t)\|_{C^\zeta} = \left\| \int_0^t P_{t-s} f(s) ds \right\|_{C^\zeta} \le \int_0^t \|P_{t-s} f(s)\|_{C^\zeta} ds.$$

Applying Lemma 9 (i) we obtain

$$\|g(t)\|_{C^\zeta} \le c_S^1 \int_0^t (t-s)^{-(\frac{\zeta-\xi}{2})} \|f(s)\|_{C^\xi} ds \le c_S^1 \frac{t^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}} \|f\|_{C_T C^\xi},$$

where c_S^1 is the constant from the Schauder estimate and in the last inequality we use $\zeta < \xi + 1$ and $\|f(s)\|_{C^\xi} \le \|f\|_{C_T C^\xi}$. We take the supremum over t and conclude

$$\|g\|_{C_T C^\zeta} \le c_S^1 \frac{T^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}} \|f\|_{C_T C^\xi}. \tag{5.14}$$

For the second term in (5.13), let us fix $0 \leq r < t \leq T$. We have

$$\begin{aligned} \|g(t) - g(r)\|_{C^\zeta} &= \left\| \int_0^t P_{t-s} f(s) ds - \int_0^r P_{r-s} f(s) ds \right\|_{C^\zeta} \\ &= \left\| \int_r^t P_{t-s} f(s) ds + \int_0^r \left[P_{t-s} f(s) - P_{r-s} f(s) \right] ds \right\|_{C^\zeta} \\ &\leq \int_r^t \|P_{t-s} f(s)\|_{C^\zeta} \, ds + \int_0^r \|P_{r-s} \left[P_{t-r} f(s) - f(s) \right]\|_{C^\zeta} \, ds, \end{aligned} \tag{5.15}$$

where the last inequality comes from the semigroup property and Jensen's inequality. For the first term on the right-hand side we apply Lemma 9(i) to get

$$\int_r^t \|P_{t-s} f(s)\|_{C^\zeta} \, ds \leq \int_r^t c_S^1 (t-s)^{-1/2} \|f(s)\|_{C^{\zeta-1}} \, ds.$$

Noting that $\xi > \zeta - 1$, taking the supremum over s and computing the integral we get

$$\int_r^t \|P_{t-s} f(s)\|_{C^\zeta} \, ds \leq \int_r^t c_S^1 (t-s)^{-1/2} \|f(s)\|_{C^\xi} \, ds \leq c_S^1 \, 2(t-r)^{1/2} \|f\|_{C_T C^\xi}. \tag{5.16}$$

For the second term in (5.15) we apply Lemma 9(i) and (ii) for some $\eta \in (0,1)$ to be specified later

$$\begin{aligned} \int_0^r \|P_{r-s} \left[P_{t-r} f(s) - f(s) \right]\|_{C^\zeta} \, ds &\leq \int_0^r c_S^1 (r-s)^{-\eta} \|P_{t-r} f(s) - f(s)\|_{C^{\zeta-2\eta}} \, ds \\ &\leq \int_0^r c_S^1 (r-s)^{-\eta} c_S^2 (t-r)^{1/2} \|f(s)\|_{C^{\zeta-2\eta+1}} \, ds \\ &= c_S^1 c_S^2 (t-r)^{1/2} \|f\|_{C_T C^{\zeta-2\eta+1}} \int_0^r (r-s)^{-\eta} ds \\ &= c_S^1 c_S^2 \frac{r^{1-\eta}}{1-\eta} (t-r)^{1/2} \|f\|_{C_T C^{\zeta-2\eta+1}}, \end{aligned}$$

where the last inequality comes from taking the supremum over s. Setting $\eta = \frac{1+\zeta-\xi}{2}$ so as $\zeta - 2\eta + 1 = \xi$ we obtain

$$\int_0^r \|P_{r-s} \left[P_{t-r} f(s) - f(s) \right]\|_{C^\zeta} \, ds \leq c_S^1 c_S^2 \frac{r^{1-\frac{1+\zeta-\xi}{2}}}{1-\frac{1+\zeta-\xi}{2}} (t-r)^{1/2} \|f\|_{C_T C^\xi}. \tag{5.17}$$

Stitching (5.16) and (5.17) together we get

$$\|g(t) - g(r)\|_{C^\zeta} \leq 2 c_S^1 (t-r)^{1/2} \|f\|_{C_T C^\xi} + c_S^1 c_S^2 \frac{r^{1-\frac{1+\zeta-\xi}{2}}}{1-\frac{1+\zeta-\xi}{2}} (t-r)^{1/2} \|f\|_{C_T C^\xi}.$$

Thus

$$[g]_{1/2, C^\zeta} \leq \left(2 c_S^1 + c_S^1 c_S^2 \frac{T^{1-\frac{1+\zeta-\xi}{2}}}{1-\frac{1+\zeta-\xi}{2}} \right) \|f\|_{C_T C^\xi}. \tag{5.18}$$

□

The result above allows us to increase the regularity in time at the cost of reducing the regularity in space and it will be used to find bounds involving the solution to the regularised Fokker-Planck equation ρ^N.

Lemma 51. *For $f \in L^\infty_T C^\xi$ for some $\xi \in \mathbb{R}$ define g by (5.12). Then $g \in L^\infty_T C^\zeta$ for any $\zeta < \xi + 2$, and $\|g\|_{L^\infty_T C^\zeta} \leq c_3 \|f\|_{L^\infty_T C^\xi}$, where*

$$c_3(T, \xi, \zeta) = c^1_S \frac{T^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}}.$$

Proof. We only need to derive a bound for $\|g(t)\|_{C^\zeta}$ which is uniform in $t \in [0, T]$. We have

$$\begin{aligned} \|g(t)\|_{C^\zeta} &= \left\| \int_0^t P_{t-s} f(s) ds \right\|_{C^\zeta} \leq \int_0^t \|P_{t-s} f(s)\|_{C^\zeta} \, ds \\ &\leq \int_0^t c^1_S (t-s)^{-\frac{\zeta-\xi}{2}} \|f\|_{C^\xi} ds = c^1_S \frac{t^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}} \|f\|_{C^\xi} \leq c^1_S \frac{T^{1-\frac{\zeta-\xi}{2}}}{1 - \frac{\zeta-\xi}{2}} \|f\|_{C^\xi}, \end{aligned}$$

where c^1_S is the constant from Lemma 9. □

5.4 Regularity properties of ρ^N

In this section, we derive estimates for 1/2-Hölder (in time) norm of ρ_N and for the spatial derivative of ρ_N. These will be needed in the following section to study the behaviour of the Euler scheme for the approximating SDE (X^N_t).

We start by citing known results. The bounds established in equations (4.60), (4.61) and (4.59) from Corollary 28 using Schauder estimates read in our setting as follows: for any $\epsilon > 0$ we have

$$[b^N]_{\frac{1}{2}, L^\infty} \leq c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C^{1/2}_T C^{-\hat{\beta}}} \tag{5.19}$$

$$\|b^N\|_{C_T L^\infty} \leq c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}} \tag{5.20}$$

$$\|b^N_x\|_{C_T L^\infty} \leq c_b N^{\frac{\epsilon+\hat{\beta}+1}{2}} \|b\|_{C_T C^{-\hat{\beta}}}, \tag{5.21}$$

where the constant c_b depends on ϵ, β and T. In what follows, ϵ will appear in the main result of this chapter and in many intermediate estimates. We do not know the limit of c_b as $\epsilon \to 0$, hence, we cannot set $\epsilon = 0$ in the above estimates. However, ϵ is meant to be arbitrarily small but fixed.

Below we will obtain bounds on ρ^N which will enable us to derive in the next section analogous estimates as in (5.19)-(5.21) but for $F(\rho^N)b^N$. By Proposition 45 we know that and ρ^N are continuous in time as functions with values in L^∞, but we also need estimates for the time-Hölder norm and for the L^∞ norm of the derivative in space with an explicit dependence on N.

Lemma 52. *Let us denote by $H^N = \tilde{F}(\rho^N)b^N$, and its derivative in space by H^N_x, then for $\eta \in (0, 1-\beta)$*

$$H^N \in C_T C^\eta \quad \text{and} \quad H^N_x \in C_T C^{\eta-1}. \tag{5.22}$$

Furthermore,

$$\|H_x^N\|_{L_T^\infty C^{\eta-1}} \le c_3 N^{\frac{\epsilon+\beta}{2}}. \tag{5.23}$$

where the constant c_3 *depends on* T*,* $\|\rho\|_{C^{1/2\vee\eta}}$*,* $\sup_N \|b^N\|_{C_T C^\beta}$ *and on* $\|b\|_{C_T C^{-\beta}}$*.*

Proof. First we show that for each $t \in [0,T]$ we have $H^N(t) \in C^\eta$ and we bound its norm uniformly in t. Recalling that $\|H^N(t)\|_{C^\eta} = \|H^N(t)\|_{L^\infty} + [H^N(t)]_\eta$, let us bound the second term as

$$[H^N(t)]_\eta = [(\tilde F(\rho^N)b^N)(t)]_\eta \le \|b^N(t)\|_{L^\infty}[\tilde F(\rho^N)(t)]_\eta + [b^N(t)]_\eta \|\tilde F(\rho^N)(t)\|_{L^\infty}. \tag{5.24}$$

Thus we have

$$\begin{aligned}
&\|(\tilde F(\rho^N)b^N)(t)\|_{C^\eta}\\
&\le \|(\tilde F(\rho^N))(t)\|_{L^\infty}\|b^N(t)\|_{L^\infty} + \|b^N(t)\|_{L^\infty}[\tilde F(\rho^N)(t)]_\eta + [b^N(t)]_\eta\|\tilde F(\rho^N)(t)\|_{L^\infty}\\
&= \|b^N(t)\|_{L^\infty}(\|(\tilde F(\rho^N))(t)\|_{L^\infty} + [\tilde F(\rho^N)(t)]_\eta) + [b^N(t)]_\eta\|\tilde F(\rho^N)(t)\|_{L^\infty}\\
&= \|b^N(t)\|_{L^\infty}\|(\tilde F(\rho^N))(t)\|_{C^\eta} + [b^N(t)]_\eta\|\tilde F(\rho^N)(t)\|_{L^\infty}\\
&\le \|b^N(t)\|_{L^\infty}\|(\tilde F(\rho^N))(t)\|_{C^\eta} + [b^N(t)]_\eta\|\tilde F(\rho^N)(t)\|_{C^\eta}\\
&= \|(\tilde F(\rho^N))(t)\|_{C^\eta}\|b^N(t)\|_{C^\eta}.
\end{aligned} \tag{5.25}$$

By Schauder estimates (Lemma 9) we have

$$\|b^N(t)\|_{C^\eta} \le c_S^1 N^{\frac{\hat\beta+\eta}{2}}\|b(t)\|_{C^{-\hat\beta}} \le c_S^1 N^{\frac{\hat\beta+\eta}{2}}\|b\|_{C_T C^{-\hat\beta}}.$$

We apply Lemma 49 on $\tilde F$ with $\alpha = \eta$ to obtain

$$\|\tilde F(\rho^N(t))\|_{C^\eta} \le c_\varphi(1 + \|\rho^N(t)\|_{C^\eta}).$$

By [26, Proposition 3.3] (choosing $\alpha = \frac12 \vee \eta$, and using that $b^N \in C_T C^{-\beta}$ and $\rho_0 \in C^{\frac12\vee\eta}$) there is an increasing function κ (depending also on $T, \|\rho_0\|_{C^{\frac12\vee\eta}}$) such that

$$\sup_{t\in[0,T]} \|\rho^N(t)\|_{C^\eta} = \|\rho^N\|_{C_T C^\eta} \le \|\rho^N\|_{C_T C^{\frac12\vee\eta}} \le \kappa\left(\|b^N\|_{C_T C^{-\beta}}\right) \le \kappa\left(\sup_N \|b^N\|_{C_T C^{-\beta}}\right),$$

and the latter is finite because $b^N \to b$ in $C_T C^{-\beta}$ as $N \to \infty$. Putting everything together in (5.25) and taking the sup over t we obtain

$$\begin{aligned}
\|H^N\|_{L_T^\infty C^\eta} &= \|\tilde F(\rho^N)b^N\|_{L_T^\infty C^\eta}\\
&\le c_\varphi\Big(1 + \kappa\big(\sup_N \|b^N\|_{C_T C^{-\beta}}\big)\Big)c_S^1 N^{\frac{\hat\beta+\eta}{2}}\|b\|_{C_T C^{-\hat\beta}}\\
&= C N^{\frac{\hat\beta+\eta}{2}},
\end{aligned}$$

where the constant C depends on the Schauder constant c_S^1, the constant c_φ, the final time T, and on the norms $\|\rho_0\|_{C^{\frac12\vee\eta}}, \sup_N \|b^N\|_{C_T C^{-\beta}}$ and $\|b\|_{C_T C^{-\hat\beta}}$, but crucially not on N. To obtain (5.23) it is now enough to use Bernstein inequality (Lemma 7) which ensures $\|H_x^N\|_{L_T^\infty C^{\eta-1}} \le c_B\|H^N\|_{L_T^\infty C^\eta}$. This gives (5.23) where $c_3 := Cc_B$.

Now let us prove the continuity in t of H^N. Take $t, s \in [0, T]$. We know that $H^N(s), H^N(t) \in C^\eta$, and with similar computations as in (5.25) we have

$$
\begin{aligned}
&\|H^N(t) - H^N(s)\|_{C^\eta} \\
&= \|(\tilde{F}(\rho^N)b^N)(t) - (\tilde{F}(\rho^N)b^N)(s)\|_{C^\eta} \\
&\leq \|((\tilde{F}(\rho^N))(t) - (\tilde{F}(\rho^N))(s))b^N(t)\|_{C^\eta} + \|(\tilde{F}(\rho^N))(s)(b^N(t) - b^N(s))\|_{C^\eta} \\
&\leq \|(\tilde{F}(\rho^N))(t) - (\tilde{F}(\rho^N))(s)\|_{C^\eta}\|b^N(t)\|_{C^\eta} + \|(\tilde{F}(\rho^N))(s)\|_{C^\eta}\|b^N(t) - b^N(s)\|_{C^\eta}.
\end{aligned}
\tag{5.26}
$$

The second term in (5.26) tends to zero as $t \to s$ since $b^N \in C_T C^\eta$ and $\|(\tilde{F}(\rho^N))(s)\|_{C^\eta} \leq \|\tilde{F}(\rho^N)\|_{L_T^\infty C^\eta} < \infty$ by the computations above. Let us consider now the first term in (5.26). For fixed t, s we know that $\rho^N(t)$ and $\rho^N(s)$ are elements of C^η so by Lemma 49 we have the bound

$$
\begin{aligned}
&\|(\tilde{F}(\rho^N))(t) - (\tilde{F}(\rho^N))(s)\|_{C^\eta} \\
&\leq c_\varphi\big(1 + \|\rho^N(t)\|^2_{C^\eta} + \|\rho^N(s)\|^2_{C^\eta}\big)^{1/2}\|\rho^N(t) - \rho^N(s)\|_{C^\eta} \\
&\leq c_\varphi\big(1 + \|\rho^N(t)\|^2_{C^{\eta\vee\frac{1}{2}}} + \|\rho^N(s)\|^2_{C^{\eta\vee\frac{1}{2}}}\big)^{1/2}\|\rho^N(t) - \rho^N(s)\|_{C^\eta} \\
&\leq c_\varphi(1 + 2\sup_N \|\rho^N\|_{C_T c^{\eta\vee\frac{1}{2}}})\|\rho^N(t) - \rho^N(s)\|_{C^\eta},
\end{aligned}
$$

where $\sup_N \|\rho^N\|_{C_T c^{\eta\vee\frac{1}{2}}}$ is finite since $\rho^N \to \rho$ in $C^{\eta\vee\frac{1}{2}}$ as $\hat{\beta} < \eta \vee \frac{1}{2} < 1 - \hat{\beta}$. Using that $\rho^N \in C_T C^\eta$, the right-hand side converges to 0 as $t \to s$. This proves that $H^N \in C_T C^\eta$. By Bernstein inequality, for every $s, t \in [0, T]$ we have

$$
\|H_x^N(t) - H_x^N(s)\|_{C^{\eta-1}} = \|\partial_x(H^N(t) - H^N(s))\|_{C^{\eta-1}} \leq c_B\|H^N(t) - H^N(s)\|_{C^\eta},
$$

therefore the continuity in time of the derivative H_x^N in $C^{\eta-1}$ holds by the continuity in time of H^N in C^η. □

Using Lemma 50, we will show that ρ^N is $\frac{1}{2}$-Hölder continuous on t.

Lemma 53. *Let ρ^N be a mild solution of the Fokker-Planck equation (5.5), then $\rho^N \in C_T^{1/2} L^\infty$. Furthermore the following bound holds for any $\eta \in (0, (1 - \beta) \wedge \nu)$*

$$
[\rho^N]_{1/2, L^\infty} \leq c_4 + c_5 N^{\frac{\epsilon+\beta}{2}}. \tag{5.27}
$$

Where

$$
\begin{aligned}
c_4 &= c_S\|\rho_0\|_{C^{1+\eta}} \\
c_5 &= c_2(T, \gamma, \nu)c_B c_b(1 + \kappa(T, \|\rho_0\|_{C^\alpha}, \|b\|_{C_T C^{-\beta}}))\|b\|_{C_T C^{-\beta}}.
\end{aligned}
$$

Proof. We observe that $\rho^N \in C_T C^\beta$ (see Proposition 45), so $\rho^N \in C_T L^\infty$. We only need to prove the Hölder continuity in time, i.e., we need to bound the seminorm $[\rho^N]_{\frac{1}{2}, L^\infty}$. Recall that by (5.6), $\rho^N(t) = P_t\rho_0 - \int_0^t P_{t-s}\partial_x H^N(t)ds$. Without any increase in complexity, we will provide a stronger bound: for any $\delta \in (0, \eta)$ we have

$$
\begin{aligned}
[\rho^N]_{\frac{1}{2}, C^{\eta-\delta}} &= \sup_{t \neq s} \frac{\|\rho^N(t) - \rho^N(s)\|_{C^{\eta-\delta}}}{|t-s|^{1/2}} \\
&= \sup_{t \neq s} \frac{\left\| P_t \rho_0 - P_s \rho_0 + \int_0^t P_{t-u} H_x^N(u) du - \int_0^s P_{s-u} H_x^N(u) du \right\|_{C^{\eta-\delta}}}{|t-s|^{1/2}} \\
&\leq \sup_{t \neq s} \frac{\|P_t \rho_0 - P_s \rho_0\|_{C^{\eta-\delta}}}{|t-s|^{1/2}} + \sup_{t \neq s} \frac{\left\| \int_0^t P_{t-u} H_x^N(u) du - \int_0^s P_{s-u} H_x^N(u) du \right\|_{C^{\eta-\delta}}}{|t-s|^{1/2}}.
\end{aligned}
\tag{5.28}
$$

For the first term, we use the Schauder estimates, Lemma 9(ii) with $\zeta = 1/2$:

$$
\|P_t \rho_0 - P_s \rho_0\|_{C^{\eta-\delta}} \leq c_S^2 |t-s|^{1/2} \|\rho_0\|_{C^{\eta-\delta+1}} \leq c_S^2 |t-s|^{1/2} \|\rho_0\|_{C^{1+\nu}} = c_5 |t-s|^{1/2},
$$

having used that $\eta - \delta + 1 < \eta + 1 < 1 + \nu$.

For the second term of (5.28), we start from noting that $H_x^N \in C_T C^{\eta-1}$ by Lemma 52. An application of Lemma 50 with $f = H_x^N$ and $\zeta = \eta - \delta$ yields

$$
t \mapsto \int_0^t P_{t-u} H_x^N(u) du \in C_T^{1/2} C^{\eta-\delta}
$$

with the norm

$$
\left\| \int_0^{\cdot} P_{\cdot-u} H_x^N(u) du \right\|_{C_T^{1/2} C^{\eta-\delta}} \leq c_1 \|H_x^N\|_{C_T C^{\eta-1}} \leq c_1 c_3 N^{\frac{\eta+\hat{\beta}}{2}},
$$

where the last inequality is by (5.23). It remains to note that $[\cdot]_{1/2, L^\infty} \leq \|\cdot\|_{C_T^{1/2} L^\infty} \leq \|\cdot\|_{C_T^{1/2} C^{\eta-\delta}}$. □

Lemma 54. *We have $\rho^N \in C_T^{1/2} C^\eta$ for any $\eta \in (0, (1-\beta) \wedge \nu)$ with*

$$
[\rho^N]_{1/2, C^\eta} \leq c_4 + c_5 N^{\frac{\eta+\hat{\beta}+\delta}{2}}
$$

where $\delta > 0$ is arbitrary but the constants c_4, c_5 from Lemma 53 depend on it.

Proof. Assume first that δ is such that $\eta + \delta < (1-\beta) \wedge \nu$. The estimate of the Hölder seminorm $[\rho^N]_{1/2, C^\eta}$ follows directly from the estimates in the proof of Lemma 53 upon taking $\eta = \eta + \delta$. If δ in the statement does not satisfy the condition $\eta + \delta < (1-\beta) \wedge \nu$ we obtain the result for any delta that satisfies it and the notice that the right-hand side of the estimate is increasing in δ. The finiteness of the Hölder seminorm together with the fact that $\rho^N \in C_T L^\infty$—as noted in the first line of the proof of the Lemma above—implies that $\rho^N \in C_T^{1/2} C^\eta$. □

Lemma 55. *We have $\rho_x^N \in L_T^\infty L^\infty$ for any $\eta \in (0, \nu)$*

$$
\|\rho_x^N\|_{L_T^\infty L^\infty} \leq c_7 + c_8 N^{\frac{\epsilon+\beta}{2}}. \tag{5.29}
$$

Where

$$c_7 = \|\rho_0\|_{C_T C^{\eta+1}}$$
$$c_8 = c_3(T,\gamma,\nu) c_B c_b (1+\kappa(T, \|\rho_0\|_{C^\alpha}, \|b\|_{C_T C^{-\beta}}))\|b\|_{C_T C^{-\beta}}.$$

Proof. Fix $\eta \in (0,\nu)$ and $\delta \in (0,\eta)$. Recall that by (5.6), $\rho^N(t) = P_t\rho_0 - \int_0^t P_{t-s}\partial_x H^N(t)ds$. We will show that $\rho^N \in L^\infty C^{\eta+1-\delta}$ by bounding the norm. We have

$$\|\rho^N\|_{L^\infty_T C^{\eta+1-\delta}} \leq \|P_\cdot \rho_0\|_{L^\infty_T C^{\eta+1-\delta}} + \left\| \int_0^\cdot P_{\cdot-s} H^N_x(s) ds \right\|_{L^\infty_T C^{\eta+1-\delta}}. \tag{5.30}$$

The continuity of the operator P_t on $C^{\eta+1-\delta}$ implies $\|P_t\rho_0\|_{C^{\eta+1-\delta}} \leq C\|\rho_0\|_{C^{\eta+1-\delta}}$ for some constant $C > 0$, so

$$\|P_\cdot \rho_0\|_{L^\infty_T C^{\eta+1-\delta}} \leq C\|\rho_0\|_{C^{\eta+1-\delta}} \leq C\|\rho_0\|_{C^{1+\nu}}, \tag{5.31}$$

where the right-hand side is bounded as $\eta - \delta \leq \nu$ and by Assumption 43. For the second term of (5.30), we apply Lemma 51 with $\zeta = \eta + 1 - \delta$ to obtain

$$\left\| \int_0^\cdot P_{\cdot-s} H^N_x(s) ds \right\|_{L^\infty_T C^{\eta+1-\delta}} \leq c_2 \|H^N_x\|_{L^\infty_T C^{\eta-1}},$$

where we used that $H^N_x \in C_T C^{\eta-1}$ by Lemma 52 with the bound $\|H^N_x\|_{C_T C^{\eta-1}} \leq c_3 N^{\frac{\eta+\hat\beta}{2}}$ and the constant c_3 depending on η and δ. Inserting the bounds into (5.30) gives

$$\|\rho^N\|_{L^\infty_T C^{\eta+1-\delta}} \leq C\|\rho_0\|_{C^{1+\nu}} + c_3 c_4 N^{\frac{\eta+\hat\beta}{2}}.$$

By the Bernstein inequality, Lemma 7, applied for each time t, we have

$$\|\rho^N_x\|_{L^\infty_T C^{\eta-\delta}} \leq c_B \|\rho^N\|_{L^\infty_T C^{\eta+1-\delta}}.$$

As $\eta - \delta > 0$, the bound $\|\rho^N_x\|_{L^\infty_T L^\infty} \leq \|\rho^N_x\|_{L^\infty_T C^{\eta-\delta}}$ holds. As we do not want the estimate in the lemma to depend on δ, we need to specify δ as a function of η, for example, $\delta = \eta/2$. The constants η and δ influence the constant c_3. $\square$

5.5 Three key bounds for B^N

The main aim of this section is to derive regularity results for the smoothed drift B^N in order to apply similar arguments to Chapter 4. More precisely, we are aiming at obtaining analogues of results in Proposition 27 and Corollary 28. In particular, we need to show that $B^N \in C_T^{1/2} L^\infty \cap L^\infty_T C^1_b$, where, $B^N = F(\rho^N)b^N$. In this section, we fix $\epsilon \in (0, (1-\beta)\wedge\nu)$ and consider the bounds (5.19)–(5.21), where the constant c_b depends on ϵ. We recall that ϵ is meant to be small but we cannot send it to 0 as the constant c_b in the aforementioned bounds may explode. We will also apply bounds from the previous section with $\eta = \epsilon$ and the constants there explode when $\eta \to 0$.

Lemma 56. *We have* $B^N \in C_T^{1/2}L^\infty$ *and*

$$\|B^N\|_{L_T^\infty L^\infty} \leq c_9 N^{\frac{\epsilon+\hat{\beta}}{2}},$$
$$[B^N]_{\frac{1}{2},L^\infty} \leq c_{10} N^{\frac{\epsilon+\hat{\beta}}{2}} + c_{11} N^{\epsilon+\hat{\beta}},$$

where

$$c_9 = \|F\|_\infty c_b \|b\|_{C_T C^{-\hat{\beta}}},$$
$$c_{10} = c_b \|b\|_{C_T C^{-\hat{\beta}}} \big(\|F\|_\infty + \|F_x\|_\infty c_5\big),$$
$$c_{11} = c_b \|b\|_{C_T C^{-\hat{\beta}}} \|F_x\|_\infty c_6.$$

Proof. Since F is bounded, we write

$$\|B^N\|_{L_T^\infty L^\infty} = \|F(\rho^N)b^N\|_{L_T^\infty L^\infty} \leq \|F\|_\infty \|b^N\|_{L_T^\infty L^\infty} \leq \|F\|_\infty c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}},$$

where (5.20) is used the last inequality. This proves the first bound in the lemma.

To show that $B^N \in C_T^{1/2}L^\infty$, it is sufficient to bound the following seminorm

$$[B^N]_{1/2,L^\infty} = \sup_{t\neq s} \frac{\|B^N(t,\cdot) - B^N(s,\cdot)\|_{L^\infty}}{|t-s|^{1/2}}.$$

We write

$$\begin{aligned}
&\|B^N(t,\cdot) - B^N(s,\cdot)\|_{L^\infty} \\
&= \|F(\rho^N(t,\cdot))b^N(t,\cdot) - F(\rho^N(t,\cdot))b^N(s,\cdot) + F(\rho^N(t,\cdot))b^N(s,\cdot) - F(\rho^N(s,\cdot))b^N(s,\cdot)\|_{L^\infty} \\
&\leq \|F(\rho^N(t,\cdot))[b^N(t,\cdot) - b^N(s,\cdot)]\|_{L^\infty} + \|b^N(s,\cdot)[F(\rho^N(t,\cdot)) - F(\rho^N(s,\cdot))]\|_{L^\infty}.
\end{aligned} \tag{5.32}$$

For the first term, the boundedness of F and (5.19) yield

$$\begin{aligned}
\|F(\rho^N(t,\cdot))[b^N(t,\cdot) - b^N(s,\cdot)]\|_{L^\infty} &\leq \|F\|_\infty \|b^N(t,\cdot) - b^N(s,\cdot)\|_{L^\infty} \\
&\leq \|F\|_\infty [b^N]_{1/2,L^\infty} |t-s|^{1/2} \\
&\leq \|F\|_\infty c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}} |t-s|^{1/2}.
\end{aligned}$$

In order to bound the second term in (5.32), we recall that the derivative F_x is bounded and b^N is bounded, so

$$\begin{aligned}
&\|b^N(s,\cdot)[F(\rho^N(t,\cdot)) - F(\rho^N(s,\cdot))]\|_{L^\infty} \\
&\leq \|b^N\|_{C_T L^\infty} \|F(\rho^N(t,\cdot)) - F(\rho^N(s,\cdot))\|_{L^\infty} \\
&\leq \|b^N\|_{C_T L^\infty} \|F_x\|_\infty \|\rho^N(t,\cdot) - \rho^N(s,\cdot)\|_{L^\infty}.
\end{aligned}$$

Inserting the estimate (5.20) for $\|b^N\|_{C_T L^\infty}$ and the bound from Lemma 53 with $\eta = \epsilon$ for $\|\rho^N(t,\cdot) - \rho^N(s,\cdot)\|_{L^\infty}$ gives

$$\|b^N(s,\cdot)[F(\rho^N(t,\cdot)) - F(\rho^N(s,\cdot))]\|_{L^\infty} \leq c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}} \|F_x\|_\infty (c_4 + c_5 N^{\frac{\epsilon+\hat{\beta}}{2}}) |t-s|^{1/2}.$$

Putting the above estimates into (5.32) and dividing by $|t-s|^{1/2}$ provides a bound for $[B^N]_{1/2,L^\infty}$ and completes the proof. □

Lemma 57. *We have $B^N \in L^\infty_T C^1_b$ and*

$$\|B^N_x\|_{L^\infty_T L^\infty} \le c_{12} N^{\frac{\epsilon+\hat\beta}{2}} + c_{13} N^{\epsilon+\hat\beta} + c_{14} N^{\frac{\epsilon+\hat\beta+1}{2}}, \tag{5.33}$$

where

$$c_{12} = \|F_x\|_\infty c_7 c_b \|b\|_{C_T C^{-\hat\beta}},$$

$$c_{13} = \|F_x\|_\infty c_8 c_b \|b\|_{C_T C^{-\hat\beta}},$$

$$c_{14} = \|F\|_\infty c_b \|b\|_{C_T C^{-\hat\beta}}.$$

Proof. For any $t \in [0,T]$, it is clear that $B^N(t)$ is differentiable in x because so are F by assumption, $\rho^N(t,\cdot)$ by Lemma 55 and $b^N(t,\cdot)$ by (5.21). To prove (5.33) we write

$$B^N_x(t,\cdot) = \frac{d}{dx}[F(\rho^N(t,\cdot))b^N(t,\cdot)] = F_x(\rho^N(t,\cdot))\rho^N_x(t,\cdot)b^N(t,\cdot) + F(\rho^N(t,\cdot))b^N_x(t,\cdot), \tag{5.34}$$

and bound its L^∞-norm as follows

$$\begin{aligned} \|B^N_x\|_{L^\infty_T L^\infty} &\le \|F_x(\rho^N)\rho^N_x b^N\|_{L^\infty_T L^\infty} + \|F(\rho^N) b^N_x\|_{L^\infty_T L^\infty} \\ &\le \|F_x(\rho^N)\|_{L^\infty_T L^\infty} \|\rho^N_x\|_{L^\infty_T L^\infty} \|b^N\|_{L^\infty_T L^\infty} + \|F(\rho^N)\|_{L^\infty_T L^\infty} \|b^N_x\|_{L^\infty_T L^\infty}. \end{aligned} \tag{5.35}$$

Recalling that F and F_x are bounded by Assumption 42 and using equations (5.20), (5.21) and Lemma 55 with $\eta = \epsilon$ we get

$$\begin{aligned} \|B^N_x\|_{L^\infty_T L^\infty} &\le \|F_x\|_\infty (c_7 + c_8 N^{\frac{\epsilon+\hat\beta}{2}}) c_b N^{\frac{\epsilon+\hat\beta}{2}} \|b\|_{C_T C^{-\hat\beta}} + \|F\|_\infty c_b N^{\frac{\epsilon+\hat\beta+1}{2}} \|b\|_{C_T C^{-\hat\beta}} \\ &\le \|F_x\|_\infty c_7 c_b N^{\frac{\epsilon+\hat\beta}{2}} \|b\|_{C_T C^{-\hat\beta}} + \|F_x\|_\infty c_8 c_b N^{\epsilon+\hat\beta} \|b\|_{C_T C^{-\hat\beta}} \\ &\quad + \|F\|_\infty c_b N^{\frac{\epsilon+\hat\beta+1}{2}} \|b\|_{C_T C^{-\hat\beta}}, \end{aligned}$$

which concludes the proof. □

5.6 Convergence rate with an exact solution to the Fokker-Planck PDE

In this section we finalize our analysis of the convergence rate for (5.1) assuming that we know ρ^N exactly, which solves the regularized Fokker-Planck equation (5.5). This assumption is relaxed in Section 5.7.

By the results above we can see that B^N belongs to the space $C^{\frac{1}{2}}_T L^\infty \cap L^\infty C^1_b$, therefore a modified version of Proposition 27—where the Euler approximation $X^{N,m}$ is for the solution of (5.8)—can be stated as follows.

Proposition 58. *Let $B^N \in C^{\frac{1}{2}}_T L^\infty \cap L^\infty C^1_b$, then*

$$\sup_t \mathbb{E}|X^{N,m} - X^N| \le d_1(N) m^{-1} + d_2(N) m^{-\frac{1}{2}}$$

where

$$d_1(N) = c_9 N^{\frac{\epsilon+\hat\beta}{2}} \left(1 + c_{12} N^{\frac{\epsilon+\hat\beta}{2}} + c_{13} N^{\epsilon+\hat\beta} + c_{14} N^{\frac{\epsilon+\hat\beta}{2}}\right)$$

$$d_2(N) = (c_{10} + c_{12}) N^{\frac{\epsilon+\hat\beta}{2}} + (c_{11} + c_{13}) N^{\epsilon+\hat\beta} + c_{14} N^{\frac{\epsilon+\hat\beta+1}{2}}.$$

Proof. See that by Lemmas 56 and 57 $B^N \in C_T^{\frac{1}{2}} L^\infty \cap L^\infty C_b^1$ holds, also Assumption 41 holds, so we can use Proposition 27 to get

$$\sup_t \mathbb{E}|X^{N,m} - X^N| \leq C_3(N) m^{-1} + C_3(N) m^{-\frac{1}{2}}$$

where

$$C_3(N) = \|B^N\|_{L_T^\infty L^\infty}(1 + \|B_x^N\|_{L_T^\infty L^\infty}),$$

and

$$C_4(N) = \|B_x^N\|_{L_T^\infty L^\infty} + [B^N]_{1/2, L^\infty}.$$

Also, by Lemmas 56 and 57 we can bound C_3 and C_4 by

$$C_3(N) \leq c_9 N^{\frac{\epsilon+\hat\beta}{2}} \left(1 + c_{12} N^{\frac{\epsilon+\hat\beta}{2}} + c_{13} N^{\epsilon+\hat\beta} c_{14} N^{\frac{\epsilon+\hat\beta+1}{2}}\right)$$

$$C_4(N) \leq (c_{10} + c_{12}) N^{\frac{\epsilon+\hat\beta}{2}} + c_{14} N^{\frac{\epsilon+\hat\beta+1}{2}} + (c_{11} + c_{13}) N^{\epsilon+\hat\beta}.$$

□

For the problem of finding a bound for the difference $\sup_t \mathbb{E}|X_t^N - X_t|$ let us see that Corollary 36 holds for any drift $\mu \in C_T^{1/2} C^{(-\beta)+}$, in particular we have assumed that B and B^N satisfy that condition, so we can use said result to bound the difference above

$$\sup_t \mathbb{E}|X_t^N - X_t| \leq c\|B^N - B\|_{C_T C^{-\hat\beta}}^{2\hat\alpha - 1} \tag{5.36}$$

for any $\hat\alpha \in (1/2, 1-\hat\beta)$ and $\hat\beta \in (\beta, 1/2)$ such that $\|b^N - b\|_{C_T C^{-\hat\beta}} < 1$. Note that we also use constants $\alpha \in (\beta, 1-\beta)$ (see Proposition 45) and β (from Assumption 41). The flexibility in the choice of $\hat\beta$ will be used in the following proposition when we estimate the difference $\|B^N - B\|_{C_T C^{-\hat\beta}}$ applying Schauder estimates.

Proposition 59. *For any $\hat\alpha \in (1/2, 1-\hat\beta)$, and any $\gamma \in (\hat\beta, 1/2)$, we have*

$$\sup_t \mathbb{E}|X_t^N - X_t| \leq c_{15} \left(c_{16} N^{-\frac{\gamma-\beta}{2}}\right)^{2\hat\alpha - 1}, \tag{5.37}$$

for N sufficiently large so that $\|B^N - B\|_{C_T C^{-\gamma}} < 1$, where the constants $c_{15}, c_{16} > 0$ do not depend on $\hat\alpha$ and $\hat\beta$.

Proof. We will use the estimate (5.36) and bound its right-hand side. Recall the definitions $B^N := F(\rho^N) b^N$ and $B := F(\rho) b$. Let $\hat\gamma \in (\gamma, \infty) \cap (\beta, 1-\beta)$, then adding and subtracting

$F(\rho^N)b$ and using the triangle inequality we get

$$\begin{aligned}\|B^N - B\|_{C_T C^{-\gamma}} &= \|F(\rho^N)b^N - F(\rho)b\|_{C_T C^{-\gamma}} \\ &= \|F(\rho^N)b^N - F(\rho^N)b + F(\rho^N)b - F(\rho)b\|_{C_T C^{-\gamma}} \\ &\leq \|F(\rho^N)b^N - F(\rho^N)b\|_{C_T C^{-\gamma}} + \|F(\rho^N)b - F(\rho)b\|_{C_T C^{-\gamma}} \\ &\leq \|F(\rho^N)(b^N - b)\|_{C_T C^{-\gamma}} + \|b(F(\rho^N) - F(\rho))\|_{C_T C^{-\gamma}}.\end{aligned}$$

Since we have the product of functions and distributions, by the norm property of the point-wise product (3.20) where we must have $\gamma < \hat{\gamma}$, we have the bound

$$\|B^N - B\|_{C_T C^{-\gamma}} \leq \|F(\rho^N)\|_{C_T C^{\hat{\gamma}}} \|b^N - b\|_{C_T C^{-\gamma}} + \|b\|_{C_T C^{-\gamma}} \|F(\rho^N) - F(\rho)\|_{C_T C^{\hat{\gamma}}}.$$

Let us split the right-hand side of the inequality. The term $\|F(\rho^N)\|_{C_T C^{\hat{\gamma}}}$ is bounded as follows:

$$\|F(\rho^N)\|_{C_T C^{\hat{\gamma}}} \leq c_\varphi(1 + \|\rho^N\|_{C_T C^{\hat{\gamma}}}) \leq c_\varphi(1 + \sup_N \|\rho^N\|_{C_T C^{\hat{\gamma}}}) \leq c_\varphi(1 + \|\rho\|_{C_T C^{\hat{\gamma}}}),$$

where the first inequality is by the linear growth for the operator F (see Lemma 49) and the third inequality uses the convergence of $\rho^N \to \rho$ in $C_T C^{\hat{\gamma}}$, which implies the boundedness of $\|\rho^N\|_{C_T C^{\hat{\gamma}}}$ in $N \in \mathbb{N}$ for the choice of $\hat{\gamma} \in (\beta, 1 - \beta)$.

The term $\|F(\rho^N) - F(\rho)\|_{C_T C^{\hat{\gamma}}}$ is bounded by the first part in Lemma 49

$$\|F(\rho^N) - F(\rho)\|_{C_T C^{\hat{\gamma}}} \leq c(1 + \|\rho^N\|^2_{C_T C^{\hat{\gamma}}} + \|\rho\|^2_{C_T C^{\hat{\gamma}}})^{1/2} \|\rho^N - \rho\|_{C_T C^{\hat{\gamma}}}.$$

Then since $\gamma \in (\hat{\beta}, 1/2)$ we use [26, Proposition 4.4] to get

$$\begin{aligned}\|\rho^N - \rho\|_{C_T C^{\hat{\gamma}}} &= \sup_{t \in [0,T]} \|\rho^N(t) - \rho(t)\|_{C^{\hat{\gamma}}} \\ &\leq l_{\hat{\gamma}}\big(\|v_0\|_{C^{\hat{\gamma}}}, \|b\|_{C_T C^{-\gamma}} \vee \|b^N\|_{C_T C^{-\gamma}}\big) \sup_{t \in [0,T]} \|b^N(t, \cdot) - b(t, \cdot)\|_{C^{-\beta}} \\ &= l_{\hat{\gamma}}\big(\|v_0\|_{C^{\hat{\gamma}}}, \|b\|_{C_T C^{-\gamma}} \vee \|b^N\|_{C_T C^{-\gamma}}\big) \|b^N - b\|_{C_T C^{-\gamma}},\end{aligned}$$

where $l_{\hat{\gamma}}$ comes from the aforementioned proposition, and is increasing in its second variable. This allows us to write

$$l_{\hat{\gamma}}\big(\|v_0\|_{C^{\hat{\gamma}}}, \|b\|_{C_T C^{-\gamma}} \vee \|b^N\|_{C_T C^{-\gamma}}\big) \leq l_{\hat{\gamma}}\big(\|v_0\|_{C^{\hat{\gamma}}}, \|b\|_{C_T C^{-\gamma}}\big) =: c_l,$$

since b^N converges to b in $C_T C^{-\beta}$ so the norm $\|b^N\|_{C_T C^{-\gamma}}$ is uniformly bounded in N.

Combining the above estimates gives

$$\begin{aligned}&\|B^N - B\|_{C_T C^{-\beta}} \\ &\leq (c_\varphi(1 + \|\rho\|_{C_T C^{\hat{\gamma}}}) + c(1 + \|\rho^N\|^2_{C_T C^{\hat{\gamma}}} + \|\rho\|^2_{C_T C^{\hat{\gamma}}})^{1/2} c_l \|b\|_{C_T C^{\gamma}}) \|b^N - b\|_{C_T C^{-\beta}}.\end{aligned}$$

Finally, by Schauder estimates 9 (ii) we have a bound in terms only of $\|b\|_{C_T C^{-\gamma}}$ for the last term above

$$\|b^N - b\|_{C_T C^{-\beta}} \leq c_S^2 \left(\frac{1}{N}\right)^{\frac{\gamma-\beta}{2}} \|b\|_{C_T C^{-\gamma}},$$

then we can put our estimates together to get

$$\begin{aligned}&\|B^N - B\|_{C_T C^{-\beta}} \\ &\leq (c_\varphi(1 + \|\rho\|_{C_T C^{\hat\gamma}}) + c(1 + \|\rho^N\|^2_{C_T C^{\hat\gamma}} + \|\rho\|^2_{C_T C^{\hat\gamma}})^{1/2} c_l \|b\|_{C_T C^{-\hat\gamma}}) c \left(\frac{1}{N}\right)^{\frac{\gamma-\beta}{2}} \|b\|_{C_T C^{-\gamma}}.\end{aligned}$$

Note that only the constant that accompanies $\left(\frac{1}{N}\right)^{\frac{\gamma-\beta}{2}} \|b\|_{C_T C^{-\gamma}}$ is dependent of $\hat\gamma$, so we chose $\hat\gamma = 1/2$. □

The final result to bound the strong error is then obtained by merging Propositions 58 and 59.

Theorem 60. *Assume that the conditions for Propositions 58 and 59 hold, then the numerical error is bounded by*

$$\mathbb{E}|X_t^{N,m} - X_t| \leq c_{17} m^{-1} + c_{18} m^{-\frac{1}{2}} + c_{15}(c_{16} N^{-\frac{\hat\beta-\beta}{2}})^{2\hat\alpha-1}, \tag{5.38}$$

where

$$c_{17} = c_8(F) N^{\frac{\epsilon+\beta}{2}} \left(1 + c_9 N^{\frac{\epsilon+\beta}{2}} + c_{10} N^{\epsilon+\beta}\right),$$

$$c_{18} = (c_9 + c_{11}) N^{\frac{\epsilon+\beta}{2}} + c_{12} N^{\frac{\epsilon+\beta+1}{2}} + (c_{13} + c_{10}) N^{\epsilon+\beta},$$

and the constant c_{16} comes from Proposition 59 and is independent of the choice of $\hat\alpha$ and $\hat\beta$.

Proof. The proof follows immediately by the triangle inequality and the usage of Propositions 58 and 59. □

The technical proposition above expresses the bound for the approximation error in terms of the smoothing parameter N and the number of time steps $1/m$, plus several auxiliary parameters ϵ, α, γ. We want to choose those auxiliary parameters in such a way that we can determine the dependency of N on m and finally express the convergence rate only in terms of the time steps m and the regularity β, thus we choose

$$\epsilon := \beta - \hat\beta,$$

which appears in the estimates in the previous sections starting from (5.19)–(5.21). We note that ϵ must be smaller than $\nu \wedge (1 - \beta)$, see the beginning of Section 5.5, which is guaranteed by the discussion in Remark 44.

Inserting the above ϵ into the formulas for $d_1(N)$ and $d_2(N)$ in Theorem 58

$$d_1(N) = c_9 N^{\frac{\beta}{2}} \left(1 + c_{12} N^{\frac{\beta}{2}} + c_{13} N^{\beta} + c_{14} N^{\frac{\beta+1}{2}}\right),$$
$$d_2(N) = (c_{10} + c_{12}) N^{\frac{\beta}{2}} + (c_{11} + c_{13}) N^{\beta} + c_{14} N^{\frac{\beta+1}{2}}.$$

The fastest growing terms in $d_1(N)$ and $d_2(N)$ are, respectively, $c_9c_{14}N^{\beta+\frac{1}{2}}$ and $c_{14}N^{\frac{\beta}{2}+\frac{1}{2}}$.

Now, we proceed to determine the relationship between N and m to balance the error of the Euler-Maruyama scheme for X^N and the error of approximating X with X^N. We postulate that $N = m^\omega$ for some $\omega > 0$ to be determined in order to maximize the rate. The asymptotic behaviour is guided by the fastest growing terms in (5.38), i.e.

$$\begin{aligned} & c_9c_{14}N^{\beta+\frac{1}{2}}m^{-1} + c_{14}N^{\frac{\beta+1}{2}}m^{-\frac{1}{2}} \\ &= c_9c_{14}m^{\omega(\beta+\frac{1}{2})}m^{-1} + c_{14}m^{\omega\frac{\beta+1}{2}}m^{-\frac{1}{2}} \\ &= c_9c_{14}m^{\omega(\beta+\frac{1}{2})-1} + c_{14}m^{\omega\frac{\beta+1}{2}-\frac{1}{2}} \end{aligned} \tag{5.39}$$

Next we turn our attention to the third addend in the rate (5.38). We select $\alpha \in (1/2, 1-\hat{\beta})$ and $\gamma \in (\hat{\beta}, 1/2)$ according to Theorem 60 so to maximize it, that is, to maximize

$$N^{-\frac{(\gamma-\beta)(2\alpha-1)}{2}}.$$

This is achieved by picking the largest possible α and γ, that is $\alpha = 1 - \hat{\beta} - \lambda$ and $\gamma = 1/2 - \lambda$, for some small $\lambda > 0$. Hence we have

$$\begin{aligned} N^{-\frac{(\gamma-\beta)(2\alpha-1)}{2}} &= N^{-\frac{(\frac{1}{2}-\lambda-\beta)(2-2\beta-2\lambda-1)}{2}} = N^{-\frac{(\frac{1}{2}-\lambda-\beta)(1-2\beta-2\lambda)}{2}} \\ &= m^{\frac{-\omega(\frac{1}{2}-\lambda-\beta)(1-2\beta-2\lambda)}{2}} = m^{-\omega(\frac{1}{2}-\lambda-\beta)^2}. \end{aligned}$$

Finally, let us find the best choice for $\omega > 0$. First of all, in order to guarantee convergence we must have that the exponents in (5.39) are negative, that is

$$\omega(\beta + \frac{1}{2}) - 1 < 0, \qquad \text{and} \qquad \omega\frac{\beta+1}{2} - \frac{1}{2} < 0$$

which simplifies to $\omega < \frac{1}{\beta+\frac{1}{2}}$ and $\omega < \frac{1}{\beta+1}$, thus the only requirement we need to enforce is $0 < \omega < \frac{1}{\beta+1}$. Under this constraint it is clear that asymptotic behaviour of (5.39) is governed by

$$m^{\omega\frac{\beta+1}{2}-\frac{1}{2}},$$

indeed, this is the leading term as long as $\omega < 1/\beta$. Therefore, the asymptotic behaviour of (5.38) is controlled by the sum

$$C_1 m^{\omega\frac{\beta+1}{2}-\frac{1}{2}} + C_2 m^{-\omega(\frac{1}{2}-\beta-\lambda)}$$

for some constants $C_1, C_2 > 0$. As the first term is increasing in ω while the second is decreasing, this expression decreases the fastest as $m \to \infty$ if the two exponents are equal:

$$\omega\frac{\beta+1}{2} - \frac{1}{2} = -\omega\left(\frac{1}{2} - \beta - \lambda\right)^2$$

which yields $\omega = \frac{1}{(1+\beta)+2(\frac{1}{2}-\beta-\lambda)^2}$. We collect this result in the following theorem.

Theorem 61. *Let $N(m) = m^\omega$. Then the optimal rate of convergence derived from*

Theorem 60 is obtained by choosing

$$\omega = \frac{1}{(1+\beta) + 2(\frac{1}{2} - \beta - \lambda)^2}$$

for any $\lambda \in (0, 1/2 - \beta)$*. This yields the convergence rate*

$$\sup_{t \in [0,T]} \mathbb{E}|X_t^{N(m),m} - X_t| \leq cm^{-\frac{1}{2} + \frac{(1+\beta)/2}{(1+\beta)+2(\frac{1}{2}-\beta-\lambda)^2}} \tag{5.40}$$

when m *is sufficiently large so that* $\|B^{N(m)} - B\|_{C_T C^{-\gamma}} < 1$ *and* $c > 0$ *independent of* N*.*

5.7 Numerical approximation of ρ^N

So far, we have operated under the assumption that the solution ρ^N to the regularized Fokker-Planck equation can be obtained exactly. Although this has allowed us to focus on the probabilistic aspects of the numerical scheme and derive a bound for the approximation error, in practice we must also approximate the function ρ^N numerically, therefore this approximation will incur in an error that we have to take into account in the convergence rate. In this section we will show how the approximation $\hat{\rho}^N$ of ρ^N can be constructed so that we preserve the bound from Theorem 61. In this theorem, m stands for the cost of computing one trajectory of the approximate solution of SDE (5.1) The cost of generating multiple trajectories is an appropriate multiplicity of the cost of generating one trajectory. However, the calculation of ρ^N is done only once, so the cost of it is independent from the number of trajectories of (X_t) that are subsequently simulated and it cannot be incorporated into the bound stated in terms of m.

We denote by $\hat{\rho}^N$ a numerical approximation of ρ^N, to which we associate the solution of SDE (5.8) $\hat{X}^N$ when we change the drift from B^N into $\hat{B}^N(x) = F(\hat{\rho}^N(x))b^N(x)$, then $\hat{X}^{N,m}$ is the Euler-Maruyama approximation of $\hat{X}^N$. The following results state the conditions which need to be satisfied by the solution $\hat{\rho}^N$ so that the rate (5.40) holds with a mere change in the constants. With this we will conclude that the derivation of Theorem 61 is solely based on the exponents of m and N.

In contrast with the decomposition of the error for the numerical scheme from (5.11), in this case we will have the error decomposed in three terms, namely

$$\mathbb{E}|\hat{X}_t^{N,m} - X_t| \leq \mathbb{E}|\hat{X}_t^{N,m} - \hat{X}_t^N| + \mathbb{E}|\hat{X}_t^N - X_t^N| + \mathbb{E}|X_t^N - X_t|, \tag{5.41}$$

so now our task is to find bounds for those three terms. For the first term we need to find an analogous bound as we did for $\mathbb{E}|X_t^{N,m} - X_t^N|$ but replacing ρ^N by $\hat{\rho}^N$, the second term is new, and the last term is bounded by Proposition 59. The final result of this section is condensed in the following theorem, and its proof is the subsequent results.

Theorem 62. *Let us assume that for* $\hat{\gamma} \in (\gamma, \infty) \cap (\beta, 1-\beta)$ *and* $\hat{\alpha} \in (1/2, 1-\beta)$*, that the solution to the regularized Fokker-Planck equation is such that* $\rho^N \in C_T^{1/2} L^\infty$ *and* $\hat{\rho}^N$ *is* *its numerical approximation, and the following holds for some $C > 0$*

$$\|\hat{\rho}^N\|_{C_T C^{\hat{\gamma}}} \le C\|\rho^N\|_{C_T C^{\hat{\gamma}}},$$

$$\|\hat{\rho}^N - \rho^N\|_{C_T C^{\hat{\gamma}}} \le CN^{-\frac{\gamma-\beta}{2}},$$

$$\|\hat{\rho}_x^N\|_{L_T^\infty L^\infty} \le C\|\rho_x^N\|_{L_T^\infty L^\infty},$$

$$[\hat{\rho}^N]_{1/2,L^\infty} \le C[\rho^N]_{1/2,L^\infty}.$$

Then

$$\begin{aligned}\sup_{0\le t\le T} \mathbb{E}|\hat{X}_t^{N,m} - X_t| \le{}& \hat{C}(d_1(N)m^{-1} + d_2(N)c_{18}m^{-\frac{1}{2}}) \\ &+ c_{15}(c_{16}N^{-\frac{\hat{\beta}-\beta}{2}})^{2\hat{\alpha}-1} \\ &+ \Big(c_\varphi(1 + C\sup_N \|\rho^N\|^2_{C_T C^{\hat{\gamma}}})^{1/2} CN^{-\frac{\gamma-\beta}{2}}\Big)^{2\alpha-1}\end{aligned} \tag{5.42}$$

where the constants d_1 and d_2 come from Proposition 58, c_{15} and c_{16} from Proposition 59, c_{18} from Theorem 60, and c_φ from Lemma 49.

Proof. The proof follows directly from using Lemmas 65, 66, and Proposition 59 on equation (5.41). □

Note the similarity between Theorem 60 and the result above, where besides the terms with m we have terms with explicit dependence on N. As we discussed after that that result, the second two terms in (5.42) will not play an important role in the convergence rate since they decrease way too quickly and again the optimization of the dependence of N on the number of steps m is going to be m^ω for ω as in Theorem 61. Leading us to the same convergence rate when we consider a numerical approximation of the Fokker-Planck equation under the assumptions of Theorem 62.

For the first term in (5.41) we need to find a bound for the quantities $\|\hat{B}^N\|_{L_T^\infty L^\infty}$, $[\hat{B}^N]_{1/2,L^\infty}$, and $\|\hat{B}_x^N\|_{L_T^\infty L^\infty}$ which will be given in the following lemmas that are analogous to Lemmas 56 and 57 for B^N.

Lemma 63. *Let $\hat{\rho}^N$ be an element of $C_T^{1/2}L^\infty$, and assume that there exists a constant $C > 0$ such that $[\hat{\rho}^N]_{1/2,L^\infty} \le C[\rho^N]_{1/2,L^\infty}$. Then we have $\hat{B}^N \in C_T^{1/2}L^\infty$ and*

$$\|\hat{B}^N\|_{L_T^\infty L^\infty} \le c_9 N^{\frac{\epsilon+\hat{\beta}}{2}},$$

$$[\hat{B}^N]_{1/2,L^\infty} \le c_{10}N^{\frac{\epsilon+\hat{\beta}}{2}} + Cc_{11}N^{\epsilon+\hat{\beta}},$$

where c_{10}, c_{11} and c_{12} are from Lemma 56.

Proof. Note that F is bounded, so

$$\|\hat{B}^N\|_{L_T^\infty L^\infty} = \|F(\hat{\rho}^N)b^N\|_{L_T^\infty L^\infty} \le \|F\|_\infty \|b^N\|_{L_T^\infty L^\infty}$$

and by (5.20)

$$\|\hat{B}^N\|_{L_T^\infty L^\infty} \leq \|F\|_\infty c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{-\hat{\beta}}}$$

which gives the first estimate. We will show that $\hat{B}^N \in C_{1/2,L^\infty}$ by bounding the seminorm

$$[\hat{B}^N]_{1/2,L^\infty} = \sup_{t,s\in[0,T],t\neq s} \frac{\|\hat{B}^N(t) - \hat{B}^N(s)\|_{L^\infty}}{|t-s|^{1/2}}.$$

which will also prove the second estimate. We write the numerator of the right hand side as

$$\begin{aligned} \|\hat{B}^N(t,\cdot) - \hat{B}^N(s,\cdot)\|_{L^\infty} = \|F(\hat{\rho}^N(t,\cdot))[b^N(t,\cdot) - b^N(s,\cdot)]\|_{L^\infty} \\ + \|b^N(s,\cdot)[F(\hat{\rho}^N(t,\cdot) - F(\hat{\rho}^N(s,\cdot)]\|_{L^\infty} \end{aligned} \tag{5.43}$$

Note that the first term is the same as in Lemma 56 and by the same argument on the boundedness of F, (5.19) and (5.20) we can bound it as

$$\begin{aligned} \|F(\hat{\rho}^N(t,\cdot)[b^N(t,\cdot) - b^N(s,\cdot)]\|_{L^\infty} &\leq \|F\|_\infty \|b^N(t,\cdot) - b^N(s,\cdot)\|_{L^\infty} \\ &\leq \|F\|_\infty [b^N]_{1/2,L^\infty} |t-s|^{1/2} \\ &\leq \|F\|_\infty c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b^N\|_{C_T C^{-\hat{\beta}}} |t-s|^{1/2} \end{aligned}$$

For the second term we recall that F_x and b^N are bounded we have

$$\begin{aligned} \|b^N(s,\cdot)[F(\hat{\rho}^N(t,\cdot)) - F(\hat{\rho}^N(s,\cdot))]\|_{L^\infty} &\leq \|b^N\|_{C_T L^\infty} \|F_x\|_\infty \|\hat{\rho}^N(t,\cdot) - \hat{\rho}^N(s,\cdot)\|_{L^\infty} \\ &\leq \|b^N\|_{C_T L^\infty} \|F_x\|_\infty [\hat{\rho}^N]_{1/2,L^\infty} |t-s|^{1/2} \\ &\leq \|b^N\|_{C_T L^\infty} \|F_x\|_\infty C[\rho^N]_{1/2,L^\infty} |t-s|^{1/2} \end{aligned}$$

where the last inequality comes from the assumed bound in $[\hat{\rho}^N]_{1/2,L^\infty}$. Furthermore, by (5.20) we get

$$\|b^N(s,\cdot)[F(\hat{\rho}^N(t,\cdot)) - F(\hat{\rho}^N(s,\cdot))]\|_{L^\infty} \leq c_b N^{\frac{\epsilon+\hat{\beta}}{2}} \|b\|_{C_T C^{\hat{\beta}}} \|F_x\|_\infty C[\rho^N]_{1/2,L^\infty} |t-s|^{1/2}.$$

Multiplying both sides of (5.43) by $|t-s|^{1/2}$ and considering the bounds we derived, gives the bound required for $[\hat{B}^N]_{1/2,L^\infty}$ and therefore $\hat{B}^N \in C_T^{1/2} L^\infty$. □

Lemma 64. *Let ρ^N be an element of $C_T^{1/2} L^\infty$ and assume there exists a constant $C > 0$ such that $\|\hat{\rho}_x^N\|_{L_T^\infty L^\infty} \leq C \|\rho_x^N\|_{L_T^\infty L^\infty}$, we have $\hat{B}^N \in L_T^\infty C_b^1$ and*

$$\|\hat{B}^N\|_{L_T^\infty L^\infty} \leq C c_{12} N^{\frac{\epsilon+\hat{\beta}}{2}} + C c_{13} N^{\epsilon+\hat{\beta}} + c_{14} N^{\frac{\epsilon+\hat{\beta}+1}{2}} \tag{5.44}$$

where the constants c_{12}, c_{13}, c_{14} come from Lemma 57.

Proof. By the same arguments as for B^N (c.f. Lemma 57) and the assumed bound on ρ_x^N, $\hat{B}^N$ is differentiable. We write

$$\hat{B}_x^N(t,\cdot) = \frac{d}{dx}[F(\hat{\rho}^N(t,\cdot))b^N(t,\cdot)] = F_x(\hat{\rho}^N(t,\cdot))\hat{\rho}_x^N(t,\cdot)b^N(t,\cdot) + F(\hat{\rho}^N(t,\cdot))b_x^N(t,\cdot), \tag{5.45}$$

so we can calculate the bound on L^∞ from the term above as

$$\begin{aligned}\|\hat{B}^N_x(t,\cdot)\|_{L^\infty_T L^\infty} \leq \|F_x(\hat{\rho}^N(t,\cdot))\|_{L^\infty_T L^\infty} C\|\rho^N_x(t,\cdot)\|_{L^\infty_T L^\infty}\|b^N(t,\cdot)\|_{L^\infty_T L^\infty} \\ + \|F(\hat{\rho}^N(t,\cdot))\|_{L^\infty_T L^\infty}\|b^N_x(t,\cdot)\|_{L^\infty_T L^\infty}.\end{aligned} \tag{5.46}$$

Using the fact that F and F_x are bounded, and Schauder estimates on the norms of b^N and b^N_x we get the desired result. □

So we are finally ready to give the bound to $\mathbb{E}|\hat{X}^{N,m}_t - X^N_t|$ in the following result.

Lemma 65. *Given the assumptions from Lemmas 63 and 64 we have the bound*

$$\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^{N,m}_t - X^N_t| \leq \hat{C}d_1(N)m^{-1} + \hat{C}d_2(N)m^{-1/2} \tag{5.47}$$

where $\hat{C} > 0$, and the constants d_1 and d_2 come from Proposition 58.

Proof. Using [9, Proposition 4] we have that

$$\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^{N,m}_t - X^N_t| \leq \hat{d}_1(N)m^{-1} + \hat{d}_2(N)m^{-1/2}$$

with

$$\begin{aligned}\hat{d}_1(N) &= \|\hat{B}^N\|_{L^\infty_T L^\infty}(1 + \|\hat{B}^N_x\|_{L^\infty_T L^\infty}) \\ \hat{d}_2(N) &= \|\hat{B}^N_x\|_{L^\infty_T L^\infty} + [\hat{B}^N]_{1/2,L^\infty}.\end{aligned}$$

So, it suffices to prove that there is a constant $\hat{C} > 0$ such that $\hat{d}_1 \leq \hat{C}d_1$ and $\hat{d}_2 \leq \hat{C}d_2$ to have the bound we require. To acomplish this, note that by Lemmas 63 and 64 we have

$$\|\hat{B}^N\|_{L^\infty_T L^\infty} \leq c_9 N^{\frac{\epsilon+\hat{\beta}}{2}} \tag{5.48}$$

$$[\hat{B}^N]_{1/2,L^\infty} \leq c_{10}N^{\frac{\epsilon+\hat{\beta}}{2}} + Cc_{11}N^{\epsilon+\hat{\beta}} \tag{5.49}$$

$$\|\hat{B}^N_x\|_{L^\infty_T L^\infty} \leq Cc_{12}N^{\frac{\epsilon+\hat{\beta}}{2}} + Cc_{13}N^{\epsilon+\hat{\beta}} + c_{14}N^{\frac{\epsilon+\hat{\beta}+1}{2}}. \tag{5.50}$$

Then we can see that

$$\begin{aligned}\hat{d}_1(N) &= \|\hat{B}^N\|_{L^\infty_T L^\infty}(1 + \|\hat{B}^N_x\|_{L^\infty_T L^\infty}) \\ &\leq c_9 N^{\frac{\epsilon+\hat{\beta}}{2}}(1 + Cc_{12}N^{\frac{\epsilon+\hat{\beta}}{2}} + Cc_{13}N^{\epsilon+\hat{\beta}} + c_{14}N^{\frac{\epsilon+\hat{\beta}+1}{2}})\end{aligned}$$

and since

$$\begin{aligned}\hat{d}_2(N) &= \|\hat{B}^N_x\|_{L^\infty_T L^\infty} + [\hat{B}^N]_{1/2,L^\infty} \\ &\leq (Cc_{12} + c_{10})N^{\frac{\epsilon+\hat{\beta}}{2}} + (Cc_{13} + Cc_{11})N^{\epsilon+\hat{\beta}} + c_{14}N^{\frac{\epsilon+\hat{\beta}+1}{2}}\end{aligned}$$

And since

$$\begin{aligned}d_1(N) &= c_9 N^{\frac{\epsilon+\hat{\beta}}{2}}\left(1 + c_{12}N^{\frac{\epsilon+\hat{\beta}}{2}} + c_{13}N^{\epsilon+\hat{\beta}} + c_{14}N^{\frac{\epsilon+\hat{\beta}}{2}}\right) \\ d_2(N) &= (c_{10} + c_{12})N^{\frac{\epsilon+\hat{\beta}}{2}} + (c_{11} + c_{13})N^{\epsilon+\hat{\beta}} + c_{14}N^{\frac{\epsilon+\hat{\beta}+1}{2}},\end{aligned}$$

we can find $\hat{C}$ such that both equations above are bounded by $\hat{C}d_1$ and $\hat{C}d_2$ respectively. Which completes the proof. □

The second term of (5.41) can be bound using Proposition 35 as we explain in the following lemma.

Lemma 66. *Let us assume that for $\hat{\gamma} \in (\gamma, \infty) \cap (\beta, 1-\beta)$ the following holds*

$$\|\hat{\rho}^N\|_{C_T C^{\hat{\gamma}}} \leq C\|\rho^N\|_{C_T C^{\hat{\gamma}}}, \tag{5.51}$$

$$\|\hat{\rho}^N - \rho^N\|_{C_T C^{\hat{\gamma}}} \leq CN^{-\frac{\gamma-\beta}{2}}, \tag{5.52}$$

for some $C > 0$. Then

$$\sup_{t\in[0,T]} \mathbb{E}|\hat{X}_t^N - X_t^N| \leq \left(c_\varphi(1 + C\sup_N \|\rho^N\|^2_{C_T C^{\hat{\gamma}}})^{1/2} CN^{-\frac{\gamma-\beta}{2}}\right)^{2\alpha-1} \tag{5.53}$$

Proof. Consider X^N is the solution to (5.8) and $\hat{X}_t^N$ its numerical approximation where $\hat{B}^N$ replaces B^N in (5.8). Note that both $\hat{B}^N, B^N \in C_T C^{-\gamma}$, so we can use Proposition 35 which yields

$$\sup_{t\in[0,T]} \mathbb{E}|\hat{X}_t^N - X_t^N| \leq c\|\hat{B}^N - B^N\|^{2\alpha-1}_{C_T C^{-\gamma}}. \tag{5.54}$$

Following similar arguments to the proof of Proposition 59 we see that

$$\begin{aligned}\|\hat{B}^N - B^N\|^{2\alpha-1}_{C_T C^{-\gamma}} &= \|F(\hat{\rho}^N)b^N - F(\rho^N)b^N\|^{2\alpha-1}_{C_T C^{-\gamma}} \\ &\leq \|b^N\|_{C_T C^{-\gamma}}\|F(\hat{\rho}^N) - F(\rho^N)\|^{2\alpha-1}_{C_T C^{\hat{\gamma}}},\end{aligned} \tag{5.55}$$

where the last inequality comes from an application of (3.20) since $b^N \in C_T C^{-\gamma}$ and $F \in C_T C^{\hat{\gamma}}$ with $\hat{\gamma} > \gamma$. Then we use Lemma 49 to find a bound to $\|F(\hat{\rho}^N) - F(\rho^N)\|_{C_T C^{\hat{\gamma}}}$ as follows

$$\|F(\hat{\rho}^N) - F(\rho^N)\|_{C_T C^{\hat{\gamma}}} \leq c_\varphi(1 + \sup_N \|\hat{\rho}^N\|^2_{C_T C^{\hat{\gamma}}})^{1/2}\|\hat{\rho}^N - \rho^N\|_{C_T C^{\hat{\gamma}}} \tag{5.56}$$

where we can bound $\|\hat{\rho}^N\|^2_{C_T C^{\hat{\gamma}}}$ by (5.51) and $\|\hat{\rho}^N - \rho^N\|_{C_T C^{\hat{\gamma}}}$ by (5.52) which gives

$$\|F(\hat{\rho}^N) - F(\rho^N)\|_{C_T C^{\hat{\gamma}}} \leq c_\varphi(1 + C\sup_N \|\rho^N\|^2_{C_T C^{\hat{\gamma}}})^{1/2} CN^{-\frac{\gamma-\beta}{2}}. \tag{5.57}$$

Finally, stitching together (5.54), (5.55) and (5.57) we get the desired result. □

5.8 Implementation of numerical methods

This section has several goals. The first one is to demonstrate that we can implement the proposed numerical methods. Additionally, it acts as a sanity check, since this is the only way to simulate the SDEs we study and to verify that the regularization process that we apply to the drift captures the characteristics of the original distributional drift. The second goal is to study the effect of the function F on the performance of the numerical scheme and on the behaviour of the solution. Finally, by using the Kolmogorov-Smirnov test we can study, albeit in a more exploratory way, the goodness of fit of the density obtained by solving the PDE compared to the empirical density of the Euler approximation of the solution to the SDE.

Implementation of the Euler-Maruyama scheme

We approach the numerical methods for McKean-Vlasov equations similarly as for the linear case presented in Chapter 4. With the added twist that we must simulate the law density of the solution, i.e. ρ^N. In [26], the authors prove that ρ^N is the solution to the associated Fokker-Planck PDE of (5.8), which we recall here for the convenience of the reader

$$\begin{cases} \partial_t \rho^N = \frac{1}{2}\Delta\rho^N - \partial_x[\tilde{F}(\rho^N)b^N], \\ \rho(0) = \rho_0. \end{cases} \tag{5.58}$$

Note that this PDE is one-dimensional, so it can be solved numerically using standard methods. As we have discussed in Section 5.7, we can find a sufficiently good numerical approximation of ρ^N which can be used in the Euler scheme without affecting its convergence rate.

In order to ease our way to solve the PDE we searched for a solver implementation that could be of use for us, finding out that an appropriate tool was the Python package `py-pde` [50] which was designed to solve general (nonlinear) PDEs. Additionally, this implementation proved to be efficient and satisfied the requirements of pre-computation of $\hat{\rho}^N$ we desired.

Similarly to Chapter 4, we have to take into consideration the coarseness of the approximations; but now, additionally, in the solution of the PDE. Namely, we need to produce an approximation of $\hat{\rho}^N$ that is granular enough for us to use it and not lose the characteristics of this function, meaning that the amount of points in the discretization of the time-space domain of the PDE cannot be too small. However, we cannot spend too much time on that, and fortunately once the PDE is solved on a relatively coarse grid, we can use interpolation methods on the coarser solution of $\hat{\rho}^N$ to get a smoother one which we can plug in the Euler solver for the SDE.

We refer to the implementation of numerical methods for the linear SDE (4.1) as the basis of what we aim to implement for (5.1). We recommend the reader to see Section 4.6 for more details, so that in the present section we can focus on the additional steps we need in the McKean-Vlasov case.

Since we do not make any assumptions a priory about the characteristics of the initial distribution of the SDE, this is the probability distribution of X_0, we have the freedom to choose one that is convenient. Recall that we cannot use as initial condition a Dirac delta as it is the case for the linear SDE. However, we could use any other variation of an appropriate density; for simplicity, we stick to the density of the standard normal distribution for this chapter.

We apply the Euler-Maruyama scheme using the drift $B^N = F(\hat{\rho}^N)b^N$. The spatial discretization of b^N follows the approach in Section 4.6, while $\hat{\rho}^N$ is obtained by solving the Fokker-Planck PDE on a predefined time-space grid. To balance accuracy and computational feasibility, both b^N and $\hat{\rho}^N$ are precomputed, stored as numerical arrays, and then used as functions through linear interpolation—e.g., using NumPy interfaces. This allows us to solve the PDE with a limited number of time and space points—e.g., 2^8 time steps and 10 spatial points for testing—with the understanding that the solution can be accurately approximated between these points via interpolation. Consequently, the SDE can be solved with a finer discretization—e.g., 2^{14} time steps. When computing convergence rates of the Euler-Maruyama scheme, we selectively increase the resolution—e.g., 10^4 sample paths for the SDE vs. 4×10^3 points in space for the PDE, where interpolation helps compensate for the coarser discretization of the PDE.

Apart for the particular interpolation method that we have to use for $F(\hat{\rho}^N)$ which is a two dimensional array, the implementation of the Euler scheme is based on the implementation used for the linear SDE. So, we solve the SDE and store only the terminal time, for the sake of efficiency, since it is what we will use to calculate the error of the scheme. Again, we make a costly computation to get the finest approximation to the SDE, which we treat as the real solution and compare it to each of the coarser approximations obtained with less time steps. We compare the values to calculate the strong error and then compute a linear regression using the log base 10 of the number of time steps as the explanatory variable and the error as the response, so that the slope obtained in the linear regression is the convergence rate. This process is done for the different values of β that we consider.

One detail to keep in mind is that using the fBm with some Hurst parameter H as the base of the creation of the distribution approximation b^N makes the drift inherently random and effectively each time we generate a new fBm we are using a different drift which we claim lives in the space $C^{-\beta}$ with $-\beta < H - 1$. However, it is important to notice that this technique is not without flaws, since the approximation procedure of the distributional drift uses the scale parameter $1/N$ and this can have a 'regularization' effect which affects how volatile the approximation b^N of the distribution is. We can observe this by comparing Figures 5.1 and 5.2. In Figure 5.1 we see an example where we ran the Euler-Maruyama scheme 40 times, with 10,000 sample paths and a fixed drift for each of the values of $\beta \approx 0, \beta \approx 1/2$ and $\beta = 0.125, 0.25, 0.375$. Fixing a drift and running the Euler scheme means that we use a single trajectory of fBm for each β, transform it into our drift, and use different Brownian motion drivers in each of the runs. In here we can observe that the 95% confidence interval is barely visible, meaning that the rates obtained for the same drift are not far apart considering different noises. On the other hand, when we do not fix the drift as in Figure 5.2, we notice that within the same values of β we must have some variation by the random nature of our technique for drift generation. Of course this is to be expected, since what is really being used to compute the error in are effectively different SDEs by virtue of having different drifts. Nonetheless, this confidence interval is still small, which indicates that our ways to generate drifts is correct, and it keeps them in the space we specify.

The usage of fractional Brownian motion as a base for our drift gives us a powerful

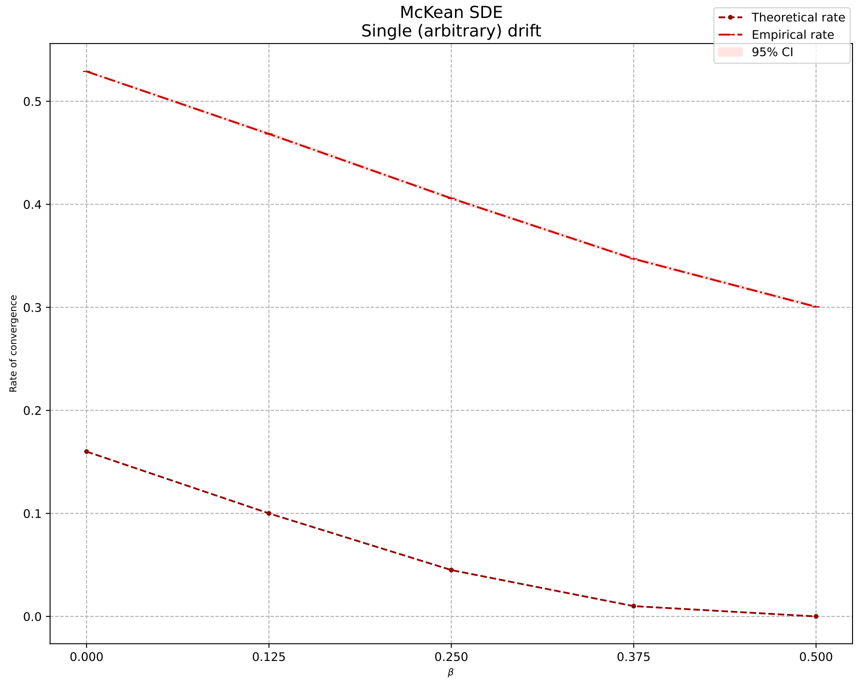

Figure 5.1: Example of convergence rate obtained for an SDE with a singular fixed drift $F(\hat{\rho}^N)b^N$ with $X_0 \sim \mathcal{N}(0,1)$, hence we have $\rho_0(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$, and $F(x) = \sin(x)$. The error is computed over 40 runs, each with 10,000 samples, of the Euler-Maruyama scheme with different Brownian motion drivers. The empirical rate is compared with the theoretical rate and a 95% confidence interval is drawn around the empirical rate.

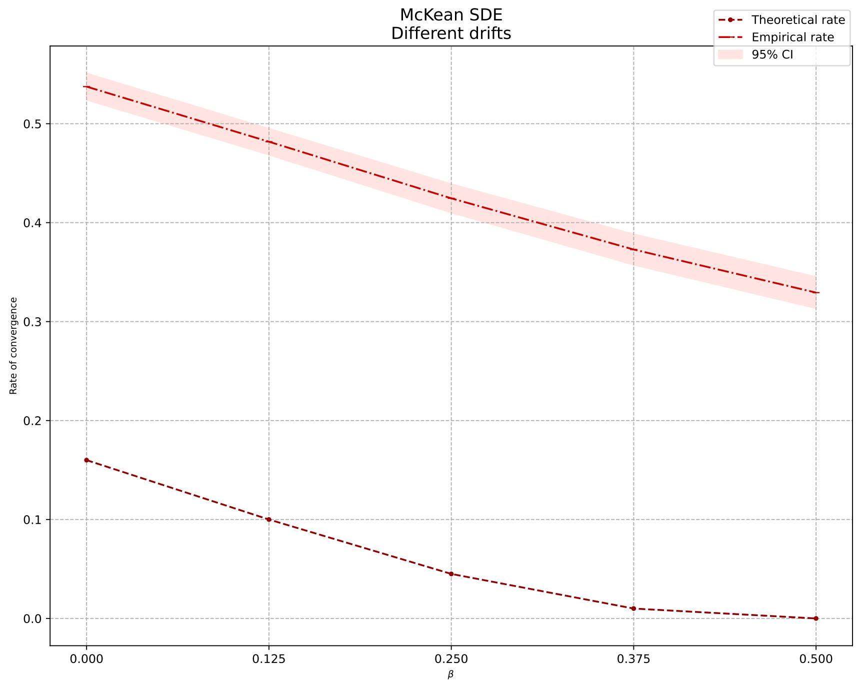

Figure 5.2: Example of convergence rate obtained for an SDE with multiple drifts $F(\hat{\rho}^N)b^N$ with $X_0 \sim \mathcal{N}(0,1)$, hence we have $\rho_0(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$, and $F(x) = \sin(x)$. The error is computed over 40 runs, each with 10,000 samples, of the Euler-Maruyama scheme with different Brownian motion drivers and 40 different drifts with the same regularity. The empirical rate is compared with the theoretical rate and a 95% confidence interval is drawn around the empirical rate.

method to study drifts with different regularities, and the randomness involved in the generation of a path of a stochastic process allows us to study infinitely many drifts with the same regularity. This allows us to explore in depth the behaviour of our algorithm. However, a natural question would be to use a deterministic Hölder continuous function, such as the Weierstrass function. Additional to the possibility to generate many different drifts, the usage of fBm is preferred in this work because the smoothing procedure using the convolution with the heat kernel seemed to overly smoothen the Weierstrass function, creating an object which was not a good representation of a distributional drift. Let us recall that the parameter of the heat kernel, $1/N$, is dependent on the time steps of the Euler scheme m. We noticed that the current design of our numerical scheme, would require a number of time steps that is too large in practice, so that we would have to think about a modification of the Euler scheme that can account for this problem. For this reason, we decided to leave this task for the future if it is indeed interesting to study.

Effect of $F(\hat{\rho}^N)$ on the distribution approximation b^N

We can see that the usage of different functions F will make the drift $B^N = F(\hat{\rho}^N)b^N$ of the McKean SDE behave different from the approximation b^N of the distribution b. Although the distribution b is approximated in an interval $[-L, L]$ for some $L > 0$, being rough all along the interval we can observe in Figures 5.3–5.4 than the product with $F(\hat{\rho}^N)$ will force the rough behaviour of the drift to be contained in a smaller interval. Importantly, the roughness of the drift is not changed since this is governed by b^N. In the aforementioned figures, we can see how outside of a certain interval the behaviour of the drift is the constant zero, which makes for very easy SDEs to solve. This is why we can observe that convergence rates for McKean SDEs as observed in Figures 5.1–5.2 are slightly higher than for SDEs without the factor $F(\hat{\rho}^N)$ in their drifts like the ones in chapter 4 even though their drifts have the same regularity, see Figure 4.4.

Notice that we do not impose any restriction on the function F besides the boundedness of itself and its derivative, and the theoretical convergence rate obtained are independent of it, however we can see the variety of law densities we can obtain with different functions. It is particularly relevant the fact that our convergence rate is a lower bound of the speed, hence it is the worst case scenario or the slowest our numerical method will converge. We know that for an equation such as (5.1), if the coefficients are smooth, the convergence rate of the Euler-Maruyama scheme is 1, just as for the Milstein scheme. The taming effect that $F(\hat{\rho}^N)$ has on b^N is a reduction of the area where the roughness of b^N affects. Notice that the usage of different functions $F(x)$ will not only make a change on the evolution of the solution to the SDE, but this will also be reflected on the solution to the PDE. If the distribution b—or rather its approximation b^N—looked ‘uniformly rough’ across an interval we selected for the empirical support $[-L, L]$ of the drift, once we multiply by the law density obtained by solving the Fokker-Planck PDE there will be an accentuation around the ‘mean’ of the density and a dimming on the edges as we can see in Figure 5.3. Here we can see that indeed for $F(x) = \sin(x)$, the original empirical domain of the drift changes from $[-10, 10]$ into roughly $[-5, 5]$ at terminal time, and although this was an

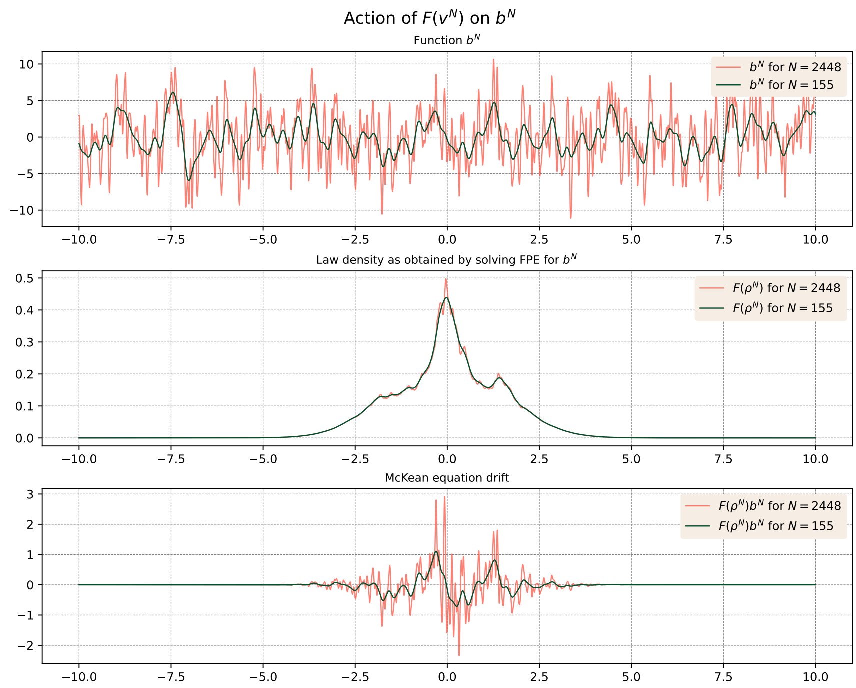

Figure 5.3: Effect of the product $F(\hat{\rho}^N)$ with b^N for the function $F(x) = \sin(x)$. Distribution approximations b^N with smoothing parameter $1/N$ for $N = 2448, 155$ and regularity $-\beta = -0.49$. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in x.

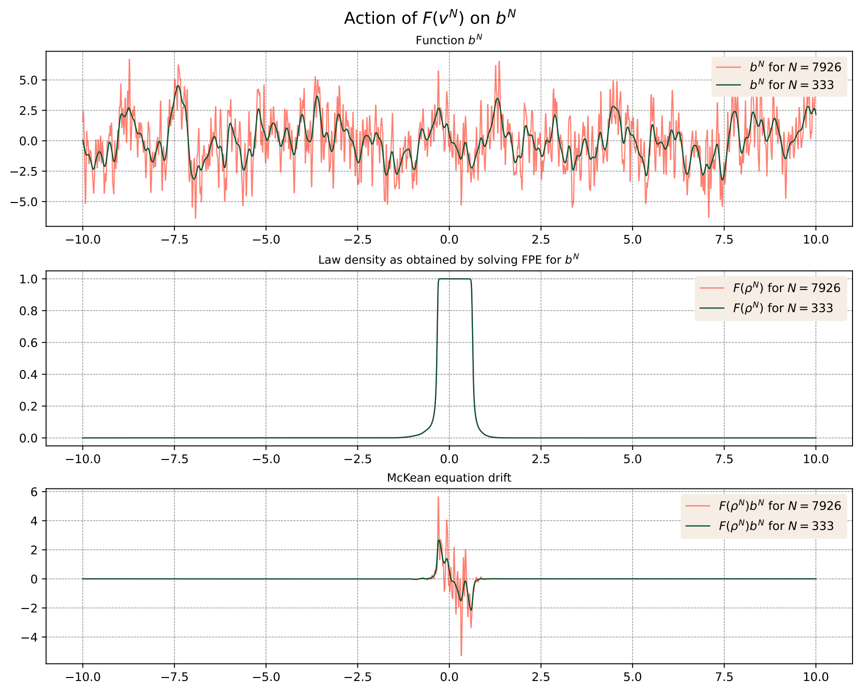

Figure 5.4: Effect of the product $F(\hat{\rho}^N)$ with b^N for the function $F(x) = 1/(1 + \exp(-100(x - 0.2)))$. Distribution approximations b^N with smoothing parameter $1/N$ for $N = 7926, 333$ and regularity $-\beta = -0.25$. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in x.

initial cause for concern since it could mean that, by the roughness of the drift, the values could fall outside the area where the drift exhibits its roughness, and we would not see the effects of the roughness on the solution of the SDE. This concern, however, is unfounded since we know that ρ is the law density of X, which means that it states the probability of the values of X. With this in mind we know that the values of X will never escape from the region where $\rho > 0$. Additionally, in Figure 5.4 we use continuous approximation of a step function to exaggerate the effect of F, this further reduces the area of effect of the drift. It can be seen on Figures 5.5–5.7 that even at terminal time, the values of the SDE will be still within range of the definition of the drift, and it is not necessary to have a drift defined in the entire $[-10, 10]$ interval. Additionally, the taming effect gives a variety on the drifts that we study and in fact allows us to see that within objects that have the same level of regularity, we can indeed obtain different results on evaluating the drifts and the convergence rates.

Goodness of fit of empirical density

We turn our attention to the fit that is attained by the solution to the Fokker-Planck equation at terminal time, $\hat{\rho}_T^N$, compared to the empirical density of X_T^N. In each example, the solution to the McKean-Vlasov equation is obtained through an Euler-Maruyama scheme with 2^{11} time steps, and 10^4 sample paths, whereas the Fokker-Planck equation is obtained with 2^{11} time steps. The number of points in the x variable is only 4×10^3 for the sake of economy in the computations. This compromise does not have a big impact in the generation of the law density with the PDE solver and on the other hand it allows us to compute the solution to the PDE and to the SDE in a time that in practice is usable.

We observe the comparison between $\hat{\rho}_T^N$ and the density histogram of X_T^N as we see in Figures 5.5–5.7. In said figures we are able to explore the effect of F on the law density of the solution on the same instance of the approximation b^N to the distribution b, this is for each value of β we generated a drift with regularity $-\beta$ and multiplied by $F(\hat{\rho}^N)$.

Time steps	K-S statistic	p-value
2^{11}	0.006495	0.790144
2^{9}	0.006261	0.825401
2^{7}	0.006127	0.844669

Table 5.1: Results of the Kolmogorov-Smirnov test for the goodness of fit of the PDE approximation of the law density with respect to the empirical density for McKean-Vlasov SDEs with drift in the space $C^{-0.25}$ and time steps $2^{11}, 2^9, 2^7$ and $F(x) = \cos(x)$.

As we can see from Table 5.8, the Kolmogorov-Smirnov test does not allow rejecting the hypothesis that the sample of the solution has a density given by the solution of the Fokker-Planck PDE. As we have mentioned it, we use this as a sanity check but at the same time it does not come as a surprise, merely by the observation of Figures 5.5–5.7, where we can indeed attest a close fit of the histogram—red bars—and the solution of the Fokker-Planck PDE—green line. One can notice that the p-value in the Table 5.8 seems to decrease slightly as we increase the number of time steps, which may at first seem to

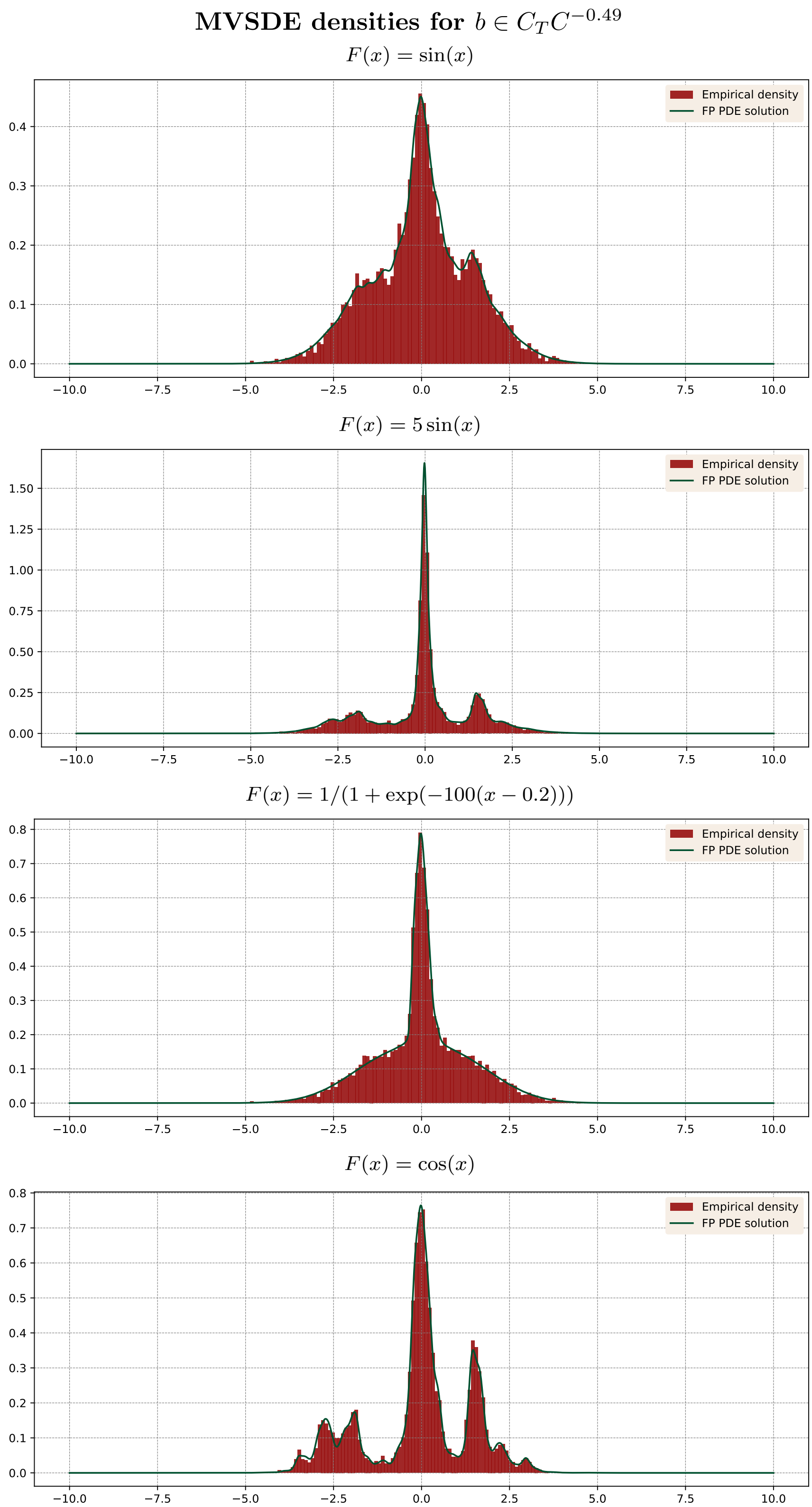

Figure 5.5: Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.49}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$.

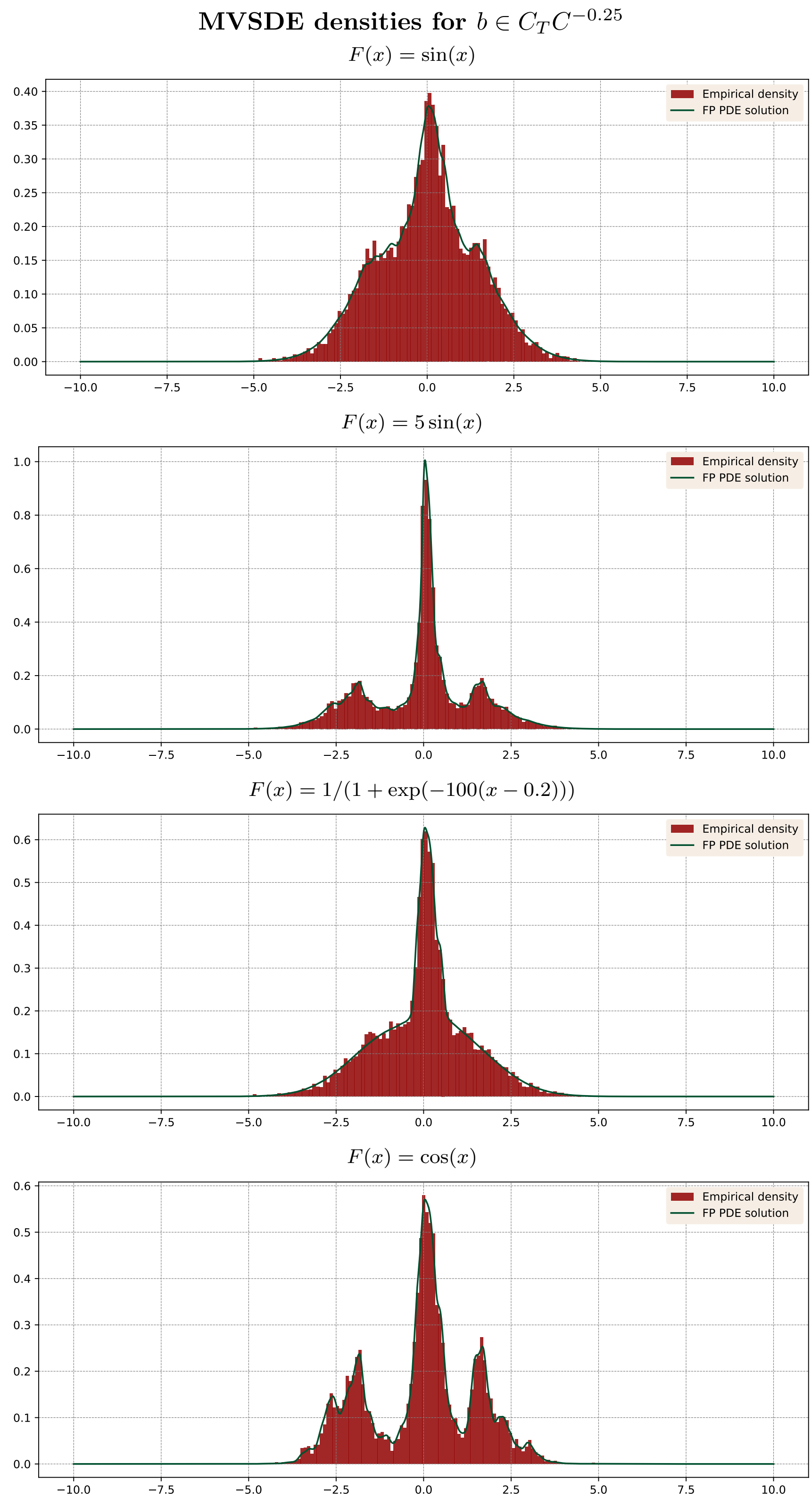

Figure 5.6: Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.25}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$.

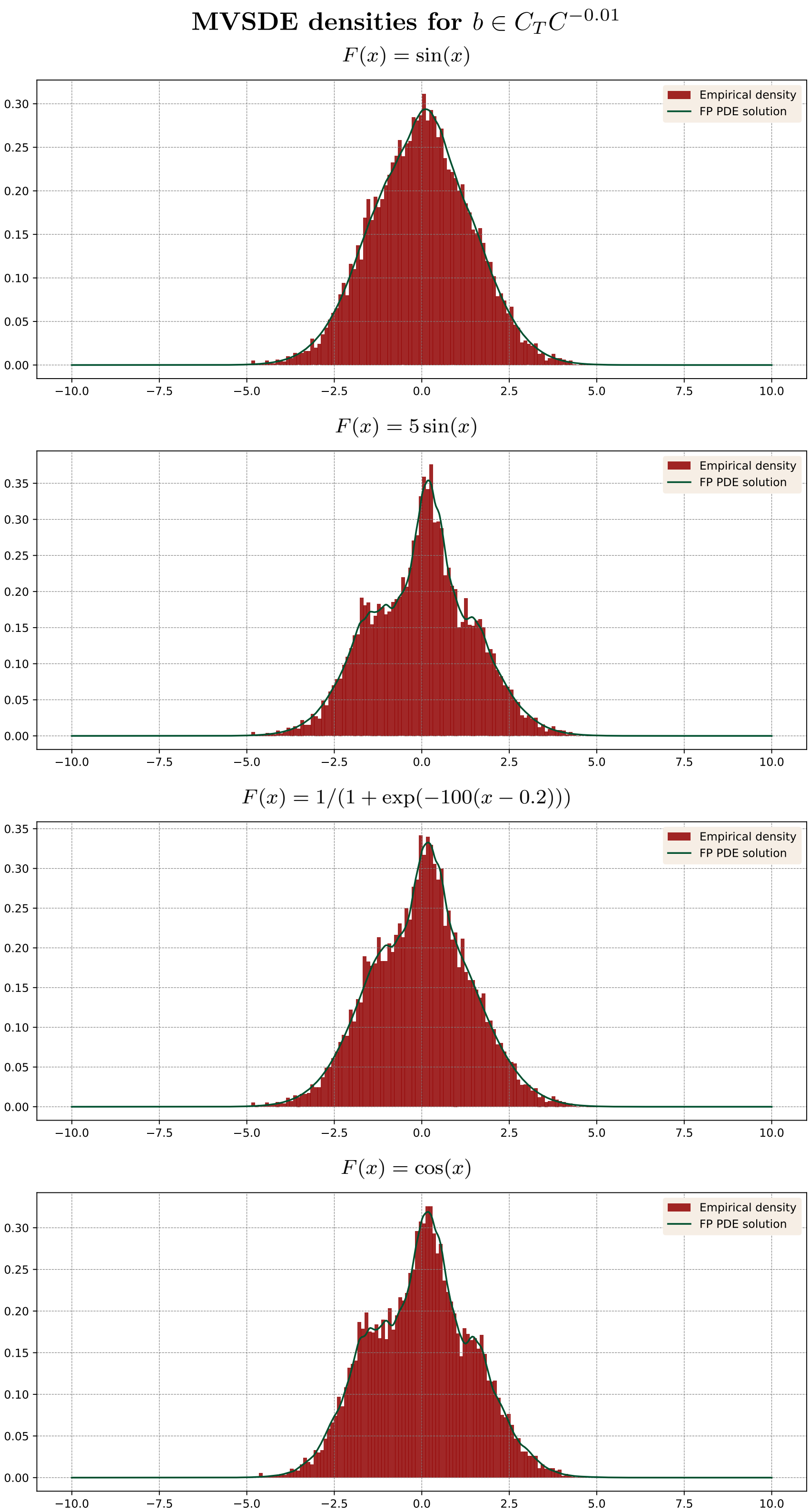

Figure 5.7: Representation of $\hat{\rho}^N$ for $b^N \in C_T C^{-0.01}$ with different values of F. From top to bottom $F(x) = \sin(x)$, $F(x) = 5\sin(x)$, $F(x) = 1/(1+\exp(-100(x-0.2)))$, $F(x) = \cos(x)$. SDE solved with 2^{11} time steps and 10^4 sample paths. PDE solved on a grid with 2^{11} time steps and 4×10^3 points in $x \in [-10, 10]$.

contradict the fact that both approximation of the SDE and the PDE must be better as we increase the time steps. However, we consider that the amount of time steps m on the iterative methods, by the dependence of the regularization parameter to create the drift B^N, will make the equations more difficult to solve each time we increase the time steps, hence there may also be a need to increase the resolution in space of the grid, which we keep fixed in our implementations. What this also indicates, is that in order to solve the PDE and for it to make sense with respect to the SDE density, we do not need to go very far in the sequence of drifts B^N, and the fit of the density obtained with the PDE is acceptable even with a lower amount of time steps. This means, that N does not increase in a way that we cannot control.

Bibliography

[1] Lukas Anzeletti, Alexandre Richard, and Etienne Tanré. "Regularisation by Fractional Noise for One-Dimensional Differential Equations with Distributional Drift". In: *Electronic Journal of Probability* 28 (2023), pp. 1–49.

[2] Siva Athreya, Oleg Butkovsky, and Leonid Mytnik. "Strong Existence and Uniqueness for Stable Stochastic Differential Equations with Distributional Drift". In: *The Annals of Probability* 48.1 (2020).

[3] Hajer Bahouri, Jean-Yves Chemin, and Raphaël Danchin. *Fourier Analysis and Nonlinear Partial Differential Equations*. 2011.

[4] Paolo Baldi. *Stochastic Calculus*. 2017.

[5] Richard F. Bass and Zhen-Qing Chen. "Brownian Motion with Singular Drift". In: *The Annals of Probability* 31.2 (2003), pp. 791–817.

[6] Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. 2008.

[7] Oleg Butkovsky, Konstantinos Dareiotis, and Mátê Gerencsér. "Approximation of SDEs: A Stochastic Sewing Approach". In: *Probability Theory and Related Fields* 181 (2021), pp. 975–1034.

[8] Giuseppe Cannizzaro and Khalil Chouk. "Multidimensional SDEs with Singular Drift and Universal Construction of the Polymer Measure with White Noise Potential". In: *The Annals of Probability* 46 (2018), pp. 1710–1763.

[9] Luis Mario Chaparro Jáquez, Elena Issoglio, and Jan Palczewski. *Convergence Rate of Numerical Scheme for SDEs with a Distributional Drift in Besov Space*. 2023. arXiv: 2309.11396.

[10] Luis Mario Chaparro Jáquez, Elena Issoglio, and Jan Palczewski. *Implementation of the Numerical Methods from "Convergence Rate of Numerical Solutions to SDEs with Distributional Drifts in Besov Spaces"*. Version v1.01. 2023. URL: https://doi.org/10.5281/zenodo.8239606.

[11] Paul-Éric Chaudru de Raynal and Stéphane Menozzi. "On Multidimensional Stable-Driven Stochastic Differential Equations with Besov Drift". In: *Electronic Journal of Probability* 27 (2022), pp. 1–52.

[12] Zhen-Qing Chen, Xicheng Zhang, and Guohuan Zhao. *Well-Posedness of Supercritical SDE Driven by Lévy Processes with Irregular Drifts*. 2017. arXiv: 1709.04632.

[13] Konstantinos Dareiotis, Máté Gerencsér, and Khoa Lê. “Quantifying a Convergence Theorem of Gyöngy and Krylov”. In: *The Annals of Applied Probability* 33 (2023), pp. 2291–2323.

[14] Tiziano De Angelis, Maximilien Germain, and Elena Issoglio. “A Numerical Scheme for Stochastic Differential Equations with Distributional Drift”. In: *Stochastic Processes and their Applications* 154 (2022), pp. 55–90.

[15] François Delarue and Roland Diel. “Rough Paths and 1d SDE with a Time Dependent Distributional Drift: Application to Polymers”. In: *Probability Theory and Related Fields* 165 (2016), pp. 1–63.

[16] Franco Flandoli, Elena Issoglio, and Francesco Russo. “Multidimensional Stochastic Differential Equations with Distributional Drift”. In: *Transactions of the American Mathematical Society* 369 (2017), pp. 1665–1688.

[17] Franco Flandoli, Francesco Russo, and Jochen Wolf. “Some SDEs with Distributional Drift Part I: General Calculus”. In: *Osaka Journal of Mathematics* 40 (2003), pp. 493–542.

[18] Franco Flandoli, Francesco Russo, and Jochen Wolf. “Some SDEs with Distributional Drift.: Part II: Lyons-Zheng Structure, Itô’s Formula and Semimartingale Characterization”. In: *Random Operators and Stochastic Equations* 12 (2004), pp. 145–184.

[19] Ludovic Goudenège, El Mehdi Haress, and Alexandre Richard. *Numerical Approximation of SDEs with Fractional Noise and Distributional Drift*. 2023. arXiv: 2302.11455.

[20] Massimiliano Gubinelli, Peter Imkeller, and Nicolas Perkowski. “Paracontrolled Distributions and Singular PDEs”. In: *Forum of Mathematics, Pi* 3 (2015), p. 75.

[21] István Gyöngy and Nicolai Krylov. “Existence of Strong Solutions for Itô’s Stochastic Equations via Approximations”. In: *Probability Theory and Related Fields* 105.2 (1996), pp. 143–158.

[22] Yaozhong Hu, Khoa Lê, and Leonid Mytnik. “Stochastic Differential Equation for Brox Diffusion”. In: *Stochastic Processes and their Applications* 127.7 (2017), pp. 2281–2315.

[23] Elena Issoglio. “A non-linear parabolic PDE with a distributional coefficient and its applications to stochastic analysis”. In: *Journal of Differential Equations* 267.10 (2019), pp. 5976–6003.

[24] Elena Issoglio, Stefano Pagliarani, Francesco Russo, and Davide Trevisani. *Degenerate McKean-Vlasov Equations with Drift in Anisotropic Negative Besov Spaces*. 2024. arXiv: 2401.09165.

[25] Elena Issoglio and Francesco Russo. “A PDE with drift of negative Besov index and linear growth solutions”. In: *Differential and Integral Equations* 37.9 (2024), pp. 585–622.

[26] Elena Issoglio and Francesco Russo. “McKean SDEs with singular coefficients”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 59.3 (2023), pp. 1530–1548.

[27] Elena Issoglio and Francesco Russo. “Stochastic Differential Equations with Singular Coefficients: The Martingale Problem View and the Stochastic Dynamics View”. In: *Journal of Theoretical Probability* 37.3 (2024), pp. 2352–2393.

[28] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. 1991.

[29] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. 1999.

[30] Khoa Lê. “A Stochastic Sewing Lemma and Applications”. In: *Electronic Journal of Probability* 25 (2020), pp. 1–55.

[31] Gunther Leobacher and Michaela Szölgyenyi. “A Numerical Method for SDEs with Discontinuous Drift”. In: *BIT Numerical Mathematics* 56.1 (2016), pp. 151–162.

[32] Gunther Leobacher and Michaela Szölgyenyi. “A Strong Order 1/2 Method for Multidimensional SDEs with Discontinuous Drift”. In: *The Annals of Applied Probability* 27.4 (2017), pp. 2383–2418.

[33] Gunther Leobacher and Michaela Szölgyenyi. “Convergence of the Euler–Maruyama Method for Multidimensional SDEs with Discontinuous Drift and Degenerate Diffusion Coefficient”. In: *Numerische Mathematik* 138 (2018), pp. 219–239.

[34] Gunther Leobacher, Michaela Szölgyenyi, and Stefan Thonhauser. “On the Existence of Solutions of a Class of SDEs with Discontinuous Drift and Singular Diffusion”. In: *Electronic Communications in Probability* 20 (2015), pp. 1–14.

[35] Gisiro Maruyama. “Continuous Markov Processes and Stochastic Equations”. In: *Rendiconti del Circolo Matematico di Palermo* 4.1 (1955), pp. 48–90.

[36] Sylvie Méléard and Sylvie Roelly-Coppoletta. “A Propagation of Chaos Result for a System of Particles with Moderate Interaction”. In: *Stochastic Processes and their Applications* 26 (1987), pp. 317–332.

[37] Karl Oelschläger. “A Law of Large Numbers for Moderately Interacting Diffusion Processes”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 69.2 (1985), pp. 279–322.

[38] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 2007.

[39] Enrico Priola. “Pathwise Uniqueness for Singular SDEs Driven by Stable Processes”. In: *Osaka Journal of Mathematics* 49.2 (2012), pp. 421–447.

[40] Paweł Przybyłowicz and Michaela Szölgyenyi. “Existence, Uniqueness, and Approximation of Solutions of Jump-Diffusion SDEs with Discontinuous Drift”. In: *Applied Mathematics and Computation* 403 (2021).

[41] Michael Renardy and Robert C. Rogers. *An Introduction to Partial Differential Equations*. 2004.

[42] Christopher Rogers and David Williams. *Diffusions, Markov processes and martingales: Itô calculus*. 2nd ed. Vol. 2. 2013. 480 pp.

[43] Francesco Russo and Gerald Trutnau. "Some Parabolic PDEs Whose Drift Is an Irregular Random Noise in Space". In: *The Annals of Probability* 35.6 (Nov. 1, 2007).

[44] Anton A. Shardin and Michaela Szölgyenyi. "Optimal Control of an Energy Storage Facility Under a Changing Economic Environment and Partial Information". In: *International Journal of Theoretical and Applied Finance* 19.04 (2016), p. 1650026.

[45] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional Diffusion Processes*. 2nd ed. Grundlehren Der Mathematischen Wissenschaften 233. Springer, 1997. 338 pp.

[46] Hiroshi Tanaka, Masaaki Tsuchiya, and Shinzo Watanabe. "Perturbation of drift-type for Lévy processes". In: *Journal of Mathematics of Kyoto University* 14.1 (1974), pp. 73–92.

[47] Hans Triebel. *Theory of Function Spaces*. 2010.

[48] Alexander Yu. Veretennikov. "On Strong Solutions and Explicit Formulas for Solutions of Stochastic Integral Equations". In: *Mathematics of the USSR-Sbornik* 39.3 (1981), p. 387.

[49] Alexander K. Zvonkin. "A transformation of the phase space of a diffusion process that removes the drift". In: *Mathematics of the USSR-Sbornik* 22.1 (1974), pp. 129–149.

[50] David Zwicker. "py-pde: A Python package for solving partial differential equations". In: *Journal of Open Source Software* 5.48 (2020), p. 2158.