INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

und der

Higher School of Economics
Moscow, Russia

vorgelegt von

Diplom Mathematikerin Anna Kozhina

aus Moscow

Tag der mündlichen Prüfung:

Parametrix method and its applications in probability theory

Betreuer: Prof. Dr. Enno Mammen
Prof. Dr. Valentin Konakov

Acknowledgments

I would like to express my gratitude Prof. Valentin Konakov and Prof. Enno Mammen for their trust and support and for their kind supervision. I also thank prof. Stephane Menozzi for very fruitful cooperation, helpful advice and kind support. Finally, I thank my family and my friends for being patient.

Zusammenfassung

In der vorliegenden Arbeit untersuchen wir Anwendungen und Weiterentwicklungen der Parametrix-Methode. Die Parametrix-Methode stammt ursprünglich aus der Theorie gewöhnlicher Differentialgleichungen und sie ermöglicht eine Reihendarstellung der Übergangsdichte der Lösung einer stochastischen Differentialgleichung, wobei die Summanden Funktionen der Übergangsdichte eines einfachen Markov-Prozesses sind. Diese Methode ist schon seit vielen Jahren bekannt, es gibt aber immer noch viele offene Probleme.

Die Arbeit ist unterteilt in drei Teile. Zuerst wird die Parametrix-Methode für Diffusionsprozesse und Markov-Prozesse in allgemeinen Situationen eingeführt. Nachdem wir die wichtigsten Konzepte und Begriffe eingeführt haben, werden wir zeigen, dass diese Technik auch angewandt werden kann, falls die Koeffizienten nicht glatt sind.

Im zweiten Teil studieren wir, wie anfällig die Übergangsdichten nicht-degenerierter Diffusionsprozesse und zugehöriger Markov-Prozesse für kleine Änderungen in den Koeffizienten sind. Diese Fragestellung taucht natürlicherweise z.B. bei Modellen mit falsch spezifizierten Koeffizienten oder bei der Untersuchung des schwachen Fehlers des Euler-Schemas mit irregulären Koeffizienten auf. Für den Beweis unseres Ergebnisses über die Kontrolle des Fehlers benötigen wir nur Hölder-Annahmen an die Stetigkeit der Koeffizienten. Diese Ergebnisse wurden von V. Konakov and S. Menozzi zur Analyse des schwachen Fehlers des Euler-Schemas angewandt [V. Konakov, S. Menozzi, 2017, Weak Error for the Euler Scheme Approximation of Diffusions with non-smooth Coefficients].

Motiviert von diesen Resultaten versuchen wir im dritten und herausforderndsten Teil der Arbeit, den schwachen und globalen Fehler im Fall von nicht-glatten Koeffizienten in Kolmogorovs degenerierter SDE zu kontrollieren. Solche Differentialgleichungen wurden zuerst 1933 von Kolmogorov eingeführt. Indem wir Techniken aus dem Artikel von V. Konakov, S. Menozzi und S. Molchanov (2010) (in dem die Autoren Lipschitz Koeffizienten betrachten) anpassen, ist es jetzt möglich degenerierte Kolmogorov-Diffusionsprozesse auch unter Hölder-Annahmen zu untersuchen. Um den schwachen und globalen Fehler in unserem Fall zu berechnen, führen wir die konkrete Version des Euler-Schemas für die degenerierte Kolmogorov-Gleichung ein, die auch als Itô-Prozess verstanden werden kann.

Unsere Ergebnisse über die Anfälligkeit der Übergangsdichten für Änderungen in den Koeffizienten lassen sich auf natürliche Weise vom nicht-degenerierten zum degenerierten Fall erweitern und sind somit auch für die Kontrolle des schwachen und globalen Fehlers geeignet. Bei dieser Erweiterung treten jedoch einige strukturelle Probleme auf, da die beiden Raumvariablen der Übergangsdichte unterschiedliche Zeitskalen haben. Diese Aspekte werden außerdem in dem Artikel [A. Kozhina, 2016, Stability of Densities for Perturbed Degenerate Diffusions] diskutiert.

Eines der Hauptergebnisse im letzten Teil erlaubt es uns, den schwachen Fehler

für degenerierte Diffusionsprozesse mit nicht-glatten Koeffizienten zu kontrollieren. Für den Beweis dieses Ergebnisses haben wir gezeigt, dass die Ableitungen des heat-kernels bzgl. nicht-degenerierter Variablen in der ersten Komponente der Übergangsdichte geeignet beschränkt werden können. Soweit uns bekannt, sind dies die ersten punktweisen Schranken an die Ableitungen bzgl. der nicht-degenerierten Variablen unter ausschließlich Hölderstetigkeitsannahmen an die Koeffizienten. Sie erweitern die bekannten

Schranken von Il'in et al. (1962) im Falle von Kolmogorov-Diffusionsprozessen. Das quantitative Verhalten der Ableitungen bzgl. der degenerierten Variablen unter minimalen Glattheitsannahmen zu untersuchen, verbleibt ein interessantes und offenes Problem. Schließlich haben wir auch versucht, die Differenz von der Übergangsdichte des Diffusionsprozesses und der Markov-Kette zu kontrollieren. Leider lässt es sich in diesem Fall nicht vermeiden, Annahmen über die Sensibilität des Kernes bzgl. der degenerierten Variablen zu stellen, weil Zeitsingularitäten höherer Ordnung sowie ein nicht beschränkter Transportterm entstehen. Die Zeitsingularitäten höherer Ordnung führen zu Restriktionen an den Hölder Exponenten, welche in unseren Annahmen auftauchen.

Abstract

The present thesis investigates applications and developments of the parametrix technique. The parametrix technique comes from the theory of ODEs. Now it reformulates as a continuity technique that provides a formal representation for the density of the SDE's solutions in terms of infinite series involving the density of another, simpler, Markov process. Although the method itself has been known for many years there are still many open problems.

The project is divided into three parts. Firstly, we are going to introduce the parametrix method for diffusions and Markov chains in general settings. After presenting main concepts and objects we emphasize that the technique can be fruitfully used also in case of non-smooth coefficients.

Secondly, we study the sensitivity of densities of non-degenerate diffusion processes and related Markov Chains with respect to a perturbation of the coefficients. Natural applications of these results appear, for example, in models with misspecified coefficients or for the investigation of the weak error of the Euler scheme with irregular coefficients. The stability controls have been derived under Hölder continuity assumptions on coefficients regularity only. Continuing the research, V. Konakov and S. Menozzi applied the results mentioned above to study the weak error of the Euler scheme approximations in their paper[V. Konakov, S. Menozzi, 2017, Weak Error for the Euler Scheme Approximation of Diffusions with non-smooth coefficients].

Motivated by these extensions, we continue with the most challenging and difficult part - the weak and global error controls for the case of rough coefficients to Kolmogorov's degenerate SDEs in the last part of the thesis. Such equations were first introduced in 1933 by Kolmogorov. Adapting the techniques, introduced in the paper written by V. Konakov, S Menozzi and S. Molchanov in 2010 (where authors considered Lipschitz coefficients), it is now possible to investigate the Holder settings for degenerate Kolmogorov diffusions. To specify the notation of the weak and global error in our framework, we also introduce the specific version of the Euler scheme for the degenerate Kolmogorov equation, which can be also seen as an Ito process.

The sensitivity analysis which we need to prove controls for the weak end global errors naturally extends from the non-degenerate case to the degenerate framework. However, some structural difficulties appear due to the different time scales for the first and the second space variable of the transition density. These aspects can be also found in the published article [A. Kozhina, 2016, Stability of densities for perturbed degenerate diffusions].

One of the key results in the last part provides the weak error controls for degenerate diffusions with non-smooth coefficients. To derive that we have proved the heat kernel derivatives bounds with respect to a non-degenerate first component of the transition density. Up to the best of our knowledge, these are the first pointwise bounds obtained on the derivatives w.r.t. the non-degenerate variables under the sole

Hölder continuity assumption on the coefficients. They extend the well-known controls derived by Il'in et al. in 1962 to Kolmogorov diffusions. Investigating the quantitative behaviour of the derivatives w.r.t. the degenerate variable under minimal smoothness assumptions remains a very interesting and open problem.

Finally, we studied the controls for the direct difference of transition densities of the diffusion and the Markov chain. Unfortunately, when handling directly the difference of the densities we cannot avoid to control sensitivities of the kernels w.r.t. to the degenerate variable. Such sensitivities lead to higher time singularities and make the unbounded transport term appear. The higher time-singularities yield the stated restriction on the Hölder exponent in our assumptions on the coefficients.

Contents

Chapter 1

Introduction

Modelling of many natural phenomena is still a challenging question. Data and observations which we receive from the real word normally contain a lot of inaccuracy and noisy factors. Using only deterministic models often makes predictions inefficient and imprecise. Thus, researchers in many fields are forced to use concepts with additional randomness inside.

The possible way to model the uncertainty is to describe the dynamics of the process in terms of Stochastic Differential Equations (SDE further). We are interested in studying Brownian SDEs of the following form

$$Z_t = z + \int_0^t b(s, Z_s)ds + \int_0^t \sigma(s, Z_s)dW_s, \tag{1.1}$$

where $(W_s)_{s\geq 0}$ is an $\mathbb{R}^k$-valued Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, Z_t is $\mathbb{R}^m$ valued, with $m \in \mathbb{N}$ possibly different of k. The coefficients b, σ are respectively $\mathbb{R}^m$ and $\mathbb{R}^m \otimes \mathbb{R}^k$ valued and s.t. a unique weak solution to (1.1) exists.

Equation (1.1) appears in many applicative fields from physics to finance. Let us for instance mention Hamiltonian mechanics [Tal02], financial mathematics [JYC10] and biology "simple epidemic model" ([Bai17]; [BY89]).

Except from some very specific cases, the SDE (1.1) cannot be solved explicitly and it therefore seems natural to investigate some related approximation procedures. One of the simplest effective computational methods is still the Euler - Maruyama method, introduced in the current SDE framework in [Mar55], and which is the analogue of the Euler method for ordinary differential equations. Fix a positive time horizon $T > 0$, for a given integer N, representing the number of time steps to be considered along the time interval $[0, T]$, introducing the time step $h = T/N$ we define for all $t \in [0, T]$:

$$Z_t^h = z + \int_0^t b(\phi(s), Z_{\phi(s)}^h)ds + \int_0^t \sigma(\phi(s), Z_{\phi(s)}^h)dW_s, \tag{1.2}$$

where setting $t_i := ih, i \in [0, N]$ we define $\phi(s) = t_i$ whenever $s \in [t_i, t_{i+1}[$. From the above dynamics, we have that Z^h is easily simulatable and can be viewed as an Itô process.

When studying the accuracy of the approximation of the scheme proposed in (1.2) for the initial SDE (1.1) two main types of errors are usually considered. Historically, the first one to be investigated (see e.g. [Mar55], Gikhman and Skorokhod [GS67], [GS82]) is the so-called *strong error*. Namely, for all $p \in [1, +\infty)$, with the usual Markovian notations for the processes Z_s^h, Z_s, started from z at the moment 0:

$$\mathcal{E}_S(T, z, h, p) := \Big(\mathbb{E}_z[\sup_{s\in[0,T]} |Z_s^{h,0,z} - Z_s^{0,z}|^p]\Big)^{1/p}. \tag{1.3}$$

This quantity is called *strong* in that it measures the distance between the whole paths. When the coefficients in (1.1) are Lipschitz continuous in space and at least 1/2-Hölder continuous in time, it is easily seen from usual stochastic analysis techniques, namely Itô' s formula Burkholder-Davies-Gundy inequalities and the Gronwall Lemma that:

$$\exists C_p(T, b, \sigma),\ \mathcal{E}_S(T, z, h, p) \le C_p h^{1/2}.$$

On the other hand, in many applications, such as e.g. some derivatives products in finance, one is only be interested in the so called *weak error* between the objects introduced in (1.1) and (1.2). For a *suitable* test function f (we remain here a bit vague about the function space to which f belongs to), one introduces:

$$\mathcal{E}_W(T, z, h, f) := \mathbb{E}_z[f(Z_T^{h,0,z})] - \mathbb{E}_z[f(Z_T^{0,z})]. \tag{1.4}$$

There are two sets of assumptions which guarantee that the convergence rate for $\mathcal{E}_W(T, z, h, f)$ is actually of order h. Namely, if

(i) b, σ, f are smooth and without any specific non-degeneracy assumptions

or

(ii) b, σ enjoy some structure property (i.e. the generator associated with (1.1) is elliptic or hypoelliptic) and some smoothness, and for f that enjoys suitable growth conditions (and that can even be a Dirac mass)

then

$$|\mathcal{E}_W(T, z, h, f)| = |\mathbb{E}_z[f(Z_T^{h,0,z} - \mathbb{E}_z[f(Z_T^{0,z})]| \le C(T, f, \sigma, b)h. \tag{1.5}$$

In both cases (i) and (ii) the main tool for the analysis is the correspondence between $\mathbb{E}_z[f(Z_T^h)]$ and the solution of a second order parabolic PDE. This correspondence is provided by the Feynman-Kac representation formula. Precisely, under the above assumptions we have that, with the usual Markovian notations, $v(t,z) := \mathbb{E}[f(Z_T^{t,z})]$ solves

$$\begin{cases}(\partial_t + L_t)v(t,z) = 0, \ (t,z) \in [0,T[\times \mathbb{R}^m, \\ v(T,z) = f(z), \ z \in \mathbb{R}^m,\end{cases} \tag{1.6}$$

where

$$L_t v(t,z) = \langle b(t,z), \nabla_z v(t,z)\rangle + \frac{1}{2}\mathrm{Tr}\Big(a(t,z)D_z^2 v(t,z)\Big), a(t,z) := \sigma\sigma^*(t,z),$$

is the generator associated with (1.1). Provided we have some smoothness on v, one can then write

$$\begin{aligned}
\mathcal{E}_W(T,z,h,f) &= \mathbb{E}[f(Z_T^{h,0,z})] - \mathbb{E}[f(Z_T^{0,z})] = \sum_{i=0}^{N-1}\mathbb{E}[v(t_{i+1}, Z_{t_{i+1}}^{h,0,z}) - v(t_i, Z_{t_i}^{h,0,z})] \quad (1.7)\\
&= \sum_{i=0}^{N-1}\mathbb{E}\Big[\int_{t_i}^{t_{i+1}}\Big\{\partial_s v(s, Z_s^{h,0,z}) + \nabla_z v(s, Z_s^{h,0,z}) b(t_i, Z_{t_i}^{h,0,z})\\
&+ \frac{1}{2}\mathrm{Tr}(D_z^2 v(s, Z_s^{h,0,z}) a(t_i, Z_{t_i}^{h,0,z}))\Big\}ds\Big] = \sum_{i=0}^{N-1}\mathbb{E}\Big[\int_{t_i}^{t_{i+1}}\Big\{\partial_s v + L_s v\Big\}(Z_s^{h,0,z})ds\Big]\\
&+ \mathbb{E}\Big[\int_{t_i}^{t_{i+1}}\Big\{\nabla_z v(s, Z_s^{h,0,z})\cdot(b(t_i, Z_{t_i}^{h,0,z}) - b(s, Z_s^{h,0,z}))\\
&+ \frac{1}{2}\mathrm{Tr}(D_z^2 v(s, Z_s^{h,0,z})(a(t_i, Z_{t_i}^{h,0,z}) - a(s, Z_s^{h,0,z})))\Big\}ds\Big]\\
&= \sum_{i=0}^{N-1}\mathbb{E}\Big[\int_{t_i}^{t_{i+1}}\Big\{\nabla_z v_\varepsilon(s, Z_s^{h,0,z})\cdot(b(t_i, Z_{t_i}^{h,0,z}) - b(s, Z_s^{h,0,z}))\\
&+ \frac{1}{2}\mathrm{Tr}(D_z^2 v(s, Z_s^{h,0,z})(a(t_i, Z_{t_i}^{h,0,z}) - a(s, Z_s^{h,0,z})))\Big\}ds\Big], \quad (1.8)
\end{aligned}$$

exploiting the PDE satisfied by v for the last equality and Itô formula for the third equality. For a function f in $C_b^{2+\beta}(\mathbb{R}^k, \mathbb{R}), \forall \beta \in (0,1]$, for example, the spatial derivatives of v up to order two are globally bounded on $[0,T]$. Through Taylor like expansions, when ever (i) or (ii) holds one can control (1.8), deriving that each contribution in (1.8) has order h^2. This leads to the error of order h achieved after summing from 0 to $N-1$.

In case (i), which is the one considered in the seminal paper by Talay and Tubaro [TT90], the smoothness of v is simply derived through stochastic flow techniques. In case (ii) let us mention that in the hypoelliptic setting, weak or strong (see Section 4.1.1 for additional details on hypoellipticity), Bally and Talay [BT96a], [BT96b] established (1.5) for bounded Borel functions f and Dirac masses respectively bases on the controls of Kusuoka and Stroock [KS84], [KS85] for the derivatives of the density of the diffusion process. We carefully mention that, for this method, which anyhow allows to consider a broad class of potential degeneracies, to apply, the coefficients are assumed to be smooth. The estimates on the tangent processes and Malliavin matrices in the works by Kusuoka and Stroock indeed require such a smoothness. In the uniformly elliptic case another approach has been developed by Konakov and Mammen [KM00], [KM02] based on parametrix expansions. The authors also manage to consider Dirac masses in (1.4).

Parametrix expansions, which roughly consists in approximating the density of a process with variable coefficients by the density of the corresponding dynamics with constant coefficients, have been a successful tool in many fields. In particular, when a good *proxy* is available (which is, for instance, the case if the coefficients b ad σ in (1.1) are non-degenerate and bounded), they allowed to derive the controls needed to analyse the weak error under rather mild assumptions. We can, for instance, mention the work of Il'in *et al.* [IKO62] who derived Gaussian heat kernel for the density of (1.1) for bounded Hölder coefficients when $\sigma\sigma^*$ is non-degenerate. Such bounds have been successfully exploited by Konakov and Menozzi [KM17] to derive, in that non-degenerate Hölder continuous setting, that for $b, \sigma \in C^{\gamma/2,\gamma}([0,T], \mathbb{R}^k)$, $\gamma \in (0,1]$ and $f \in C^\beta(\mathbb{R}^k, \mathbb{R}), \beta \in (0,1]$:

$$|\mathcal{E}_W(T, z, h, f)| = |\mathbb{E}_z[f(Z_s^h)] - \mathbb{E}_z[f(Z_s)]| \le C(T, f, \sigma, b)h^{\gamma/2}, \tag{1.9}$$

improving the previous result by Mikulevičius and Platen [MP91] who also obtained the bound (1.9) for a function $f \in C^{2+\gamma}(\mathbb{R}^k, \mathbb{R})$. This additional smoothness was due to the fact that they based their analysis and the associated Schauder estimates (which could already be found in [IKO62]). Going back to the heat-kernel directly allows to notably alleviate the smoothness assumptions on the final condition, which might be useful for applications.

Intuitively, the above convergence rate can be explained from the fact that, in the low regularity setting, the terms of order greater than one in the telescopic sum (1.7) cannot be expanded much further. Namely, we can only exploit the γ-Hölder continuity of the coefficients which leads to an error controlled by the increments

$$\mathbb{E}[|b(s, Z_s^h) - b(\phi(s), Z_{\phi(s)}^h)|] + \mathbb{E}[|a(s, Z_s^h) - a(\phi(s), Z_{\phi(s)}^h)|] \le C(b, \sigma)h^{\gamma/2}.$$

In other words, the convergence rate is closer to the one associated with the *strong error* in (1.3).

Now, for many applications, like e.g. neuro-sciences or diffusions in random media, it is even important to handle rougher coefficients, for instance piecewise smooth drifts in (1.1). In that case, the previously mentioned heat-kernel and bounds do not hold. Motivated by the investigation of the related weak error for Dirac masses test functions we have developed, with V. Konakov and S. Menozzi, a sensitivity analysis of the density of (1.1) (when suitable good Gaussian bounds exist) with respect to a perturbation of the coefficients. This is the first main result of the Thesis which led to the publication [KKM17] and is thoroughly developed in Chapter 3.

Namely, if we introduce the SDE of the form:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t,\ t \in [0, T], \tag{1.10}$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d,\ \sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded coefficients that are measurable in time and Hölder continuous in space (this last condition will be possibly relaxed for the drift term b). Also, $a(t, x) := \sigma\sigma^*(t, x)$ is assumed to be uniformly elliptic. In particular those assumptions guarantee that (1.10) admits a unique weak solution, see e.g. Bass and Perkins [BP09], [Men11] from which the uniqueness to the martingale problem for the associated generator can be derived under the current assumptions.

We now introduce, for a given parameter $\varepsilon > 0$, a perturbed version of (1.10) with dynamics:

$$dX_t^{(\varepsilon)} = b_\varepsilon(t, X_t^{(\varepsilon)})dt + \sigma_\varepsilon(t, X_t^{(\varepsilon)})dW_t,\ t \in [0, T], \tag{1.11}$$

where $b_\varepsilon : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d,\ \sigma_\varepsilon : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ satisfy at least the same assumptions as b, σ and are in some sense meant to be *close* to b, σ when ε is small.

It is known that, under the previous assumptions, the density of the processes $(X_t)_{t\geq 0}, (X_t^{(\varepsilon)})_{t\geq 0}$ exists and satisfies some Gaussian bounds, see e.g Aronson [Aro59] or [DM10] for extensions to some degenerate cases.

In the Chapter 3 we investigate, applying the parametrix technique, how the closeness of $(b_\varepsilon, \sigma_\varepsilon)$ and (b, σ) is reflected on the respective densities of the associated processes. Our stability results will also apply to two Markov chains with respective dynamics:

$$\begin{aligned} Y_{t_{k+1}} &= Y_{t_k} + b(t_k, Y_{t_k})h + \sigma(t_k, Y_{t_k})\sqrt{h}\xi_{k+1}, Y_0 = x, \\ Y_{t_{k+1}}^{(\varepsilon)} &= Y_{t_k}^{(\varepsilon)} + b_\varepsilon(t_k, Y_{t_k}^{(\varepsilon)})h + \sigma_\varepsilon(t_k, Y_{t_k}^{(\varepsilon)})\sqrt{h}\xi_{k+1}, Y_0^{(\varepsilon)} = x, \end{aligned} \tag{1.12}$$

where $h > 0$ is a given time step, for which we denote for all $k \geq 0,\ t_k := kh$ and the $(\xi_k)_{k\geq 1}$ are centered i.i.d. random variables satisfying some integrability conditions. Again, the key tool will be the parametrix representation for the densities of the chains and the Gaussian local limit theorem.

Let us specify the following assumptions **(A)** which we use in Chapter 3. Below, the parameter $\varepsilon > 0$ is fixed and the constants appearing in the assumptions **do not depend** on ε.

(A1) (Boundedness of the coefficients). The components of the vector-valued functions $b(t, x), b_\varepsilon(t, x)$ and the matrix-valued functions $\sigma(t, x), \sigma_\varepsilon(t, x)$ are bounded.

Specifically, there exist constants $K_1, K_2 > 0$ s.t.

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |b(t,x)| + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |b_\varepsilon(t,x)| \le K_1,$$
$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\sigma(t,x)| + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\sigma_\varepsilon(t,x)| \le K_2.$$

(**A2**) **(Uniform Ellipticity)**. The matrices $a := \sigma\sigma^*, a_\varepsilon := \sigma_\varepsilon\sigma_\varepsilon^*$ are uniformly elliptic, i.e. there exists $\Lambda \geq 1,\ \forall (t,x,\xi) \in [0,T]\times(\mathbb{R}^d)^2$,

$$\Lambda^{-1}|\xi|^2 \le \langle a(t,x)\xi,\xi\rangle \le \Lambda|\xi|^2, \Lambda^{-1}|\xi|^2 \le \langle a_\varepsilon(t,x)\xi,\xi\rangle \le \Lambda|\xi|^2.$$

(**A3**) **(Hölder continuity in space)**. For some $\gamma \in (0,1]$, $\kappa < \infty$, for all $t \in [0,T]$,

$$|\sigma(t,x) - \sigma(t,y)| + |\sigma_\varepsilon(t,x) - \sigma_\varepsilon(t,y)| \quad \le \quad \kappa\,|x-y|^\gamma\,.$$

Observe that the last condition also readily gives, thanks to the boundedness of $\sigma, \sigma_\varepsilon$ that a, a_ε are also uniformly γ-Hölder continuous.

For a given $\varepsilon > 0$, we say that assumption **(A)** holds when conditions **(A1)**-**(A3)** are in force. Let us now introduce, under **(A)**, the quantities that will bound the difference of the densities in our main results below. Set for $\varepsilon > 0$:

$$\Delta_{\varepsilon,b,\infty} := \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \{|b(t,x) - b_\varepsilon(t,x)|\},$$
$$\forall q \in (1,+\infty),\ \Delta_{\varepsilon,b,q} := \sup_{t\in[0,T]} \|b(t,.) - b_\varepsilon(t,.)\|_{L^q(\mathbb{R}^d)}.$$

Since $\sigma, \sigma_\varepsilon$ are both γ-Hölder continuous, see (**A3**), we also define

$$\Delta_{\varepsilon,\sigma,\gamma} := \sup_{u\in[0,T]} |\sigma(u,.) - \sigma_\varepsilon(u,.)|_\gamma,$$

where for $\gamma \in (0,1]$, $|.|_\gamma$ stands for the usual Hölder norm in space on $C_b^\gamma(\mathbb{R}^d, \mathbb{R}^d\otimes\mathbb{R}^d)$ (space of Hölder continuous bounded functions, see e.g. Krylov [Kry96]) i.e. :

$$|f|_\gamma := \sup_{x\in\mathbb{R}^d} |f(x)| + [f]_\gamma,\ [f]_\gamma := \sup_{x\neq y,(x,y)\in(\mathbb{R}^d)^2} \frac{|f(x)-f(y)|}{|x-y|^\gamma}.$$

We eventually set for $q \in (1,+\infty]$,

$$\Delta_{\varepsilon,\gamma,q} := \Delta_{\varepsilon,\sigma,\gamma} + \Delta_{\varepsilon,b,q}.$$

Theorem 3.2.1. *Fix $\varepsilon > 0$ and a final deterministic time horizon $T > 0$. Under assumptions **(A)**, specified before, for $q > d$, there exist $C := C(q) \geq 1, c := c(q) \in (0,1]$ s.t. for all $0 \leq s < t \leq T, (x,y) \in (\mathbb{R}^d)^2$:*

$$p_c(t-s, y-x)^{-1}|(p - p_\varepsilon)(s,t,x,y)| \leq C\Delta_{\varepsilon,\gamma,q},$$

where $p(s,t,x,.), p_\varepsilon(s,t,x,.)$ respectively stand for the transition densities at time t of equations (1.10), (1.11) *starting from x at time s. Also, we denote for a given $c > 0$ and for all $(u,z) \in \mathbb{R}^+ \times \mathbb{R}^d$, $p_c(u,z) := \frac{c^{d/2}}{(2\pi u)^{d/2}} \exp(-c\frac{|z|^2}{2u})$. If $q = \infty$, the constants C, c do not depend on q.*

This and the next theorem will be restated and discussed in Section 3.2.1.

Before stating our results for Markov Chains we introduce two kinds of innovations in (1.12). Namely:

(**IG**) The i.i.d. random variables $(\xi_k)_{k\geq 1}$ are Gaussian, with law $\mathcal{N}(0, I_d)$. In that case the dynamics in (1.12) correspond to the Euler discretization of equations (1.10) and (1.11).

(**IP**) For a given integer $M > 2d + 5 + \gamma$, the innovations $(\xi_k)_{k\geq 1}$ are centered and have C^5 density f_ξ which has, together with its derivatives up to order 5, at most polynomial decay of order M. Namely, for all $z \in \mathbb{R}^d$ and multi-index $\nu, |\nu| \leq 5$:

$$|D^\nu f_\xi(z)| \leq CQ_M(z),$$

where we denote for all $r > d,\ z \in \mathbb{R}^d,\ \ Q_r(z) := c_r \frac{1}{(1+|z|)^r},\ \int_{\mathbb{R}^d} dz Q_r(z) = 1$.

Theorem 3.2.2. *Fix $\varepsilon > 0$ and a final deterministic time horizon $T > 0$. For $h = T/N,\ N \in \mathbb{N}^*$, we set for $i \in \mathbb{N},\ t_i := ih$. Under **(A)**, assuming that either(**IG**) or (**IP**) holds, and for $q > d$ there exist $C := C(q) \geq 1, c := c(q) \in (0,1]$ s.t. for all $0 \leq t_i < t_j \leq T, (x,y) \in (\mathbb{R}^d)^2$:*

$$\chi_c(t_j - t_i, y-x)^{-1}|(p^h - p^h_\varepsilon)(t_i,t_j,x,y)| \leq C\Delta_{\varepsilon,\gamma,q},$$

where $p^h(t_i,t_j,x,.), p^h_\varepsilon(t_i,t_j,x,.)$ respectively stand for the transition densities at time t_j of the Markov Chains Y and $Y^{(\varepsilon)}$ in (1.12) *starting from x at time t_i. Also:*

- *If (**IG**) holds:*

$$\chi_c(t_j - t_i, y-x) := p_c(t_j - t_i, y-x),$$

with p_c as in Theorem 3.2.1.

- *If (**IP**) holds:*

$$\chi_c(t_j - t_i, y-x) := \frac{c^d}{(t_j - t_i)^{d/2}} Q_{M-(d+5+\gamma)}\left(\frac{|y-x|}{(t_j-t_i)^{1/2}/c}\right).$$

Again, if $q = +\infty$ the constants C, c do not depend on q.

Continuing the research, V. Konakov and S. Menozzi applied results mentioned above to study the weak error of the Euler scheme approximations in the paper [KM17]. To investigate the weak error for rough drifts, the idea in [KM17] is then to mollify the drifts. The controls between the densities of the initial diffusion and the one with mollified coefficients is precisely controlled by the previous result. The same occurs with the Euler scheme. It therefore remains to control the difference between the densities of the mollified diffusion and schemes which can be addressed from previous results of [KM02] provided the high order derivatives (which explode with the mollifying parameter) are sharply controlled.

Let us mention that the previous strategy was also used in [KM17] to handle the Hölder weak error for Dirac masses. In that case, the result of [KM17] can be improved following the approach proposed by Frikha [Fri18] who avoided any smoothing procedure.

Motivated by the extension of the previous study, we continue with the weak error controls for the case of rough coefficients to Kolmogorov's degenerate SDEs in Chapter 4. Namely, we specify the model in (1.1) writing $Z_t = (X_t, Y_t)$ with:

$$\begin{cases} dX_t = b(X_t, Y_t)dt + \sigma(X_t, Y_t)dW_t, \\ dY_t = X_t dt, t \in [0, T], \end{cases} \tag{1.13}$$

where $b : \mathbb{R}^{2d} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^{2d} \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded coefficients that are Hölder continuous in space (this condition will be possibly relaxed for the drift term b) and W is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. In (1.13), $T > 0$ is a fixed deterministic final time. Also, $a(x, y) := \sigma\sigma^*(x, y)$ is assumed to be uniformly elliptic.

We point out that those assumptions (specified below) are actually sufficient to guarantee weak uniqueness for the solution of equation (1.13), see Remark 4.2.1.

Such equations were first introduced in the seminal paper [Kol34] by Kolmogorov. In that work, he found the explicit expression of the density when the coefficients are constants. The parametrix approach in that framework has then been applied by various authors, Weber [Web51], Sonin [Son67] to the more recent [KMM10] under various kinds of assumptions. Adapting the techniques introduced in the last quoted work, which deals with Lipschitz coefficients, it is now possible to consider the Hölder setting for the degenerate Kolmogorov diffusions of type (1.13). The sensitivity analysis naturally extends to this framework. These aspects are detailed in Chapter 4 (see as well the published article [Koz16]).

Precisely, let us introduce the Euler scheme for the SDE (1.13) first. For a fixed N and $T > 0$ we define a time grid $\{0, t_1, \ldots, t_N\}$ with a given step $h := T/N$, i.e.

$t_i = ih$, for $i = 0, \ldots, N$ and the scheme

$$\begin{cases} X_t^h = x + \int_0^t b(X_{\phi(s)}^h, Y_{\phi(s)}^h)ds + \int_0^t \sigma(X_{\phi(s)}^h, Y_{\phi(s)}^h)dW_s, \\ Y_t^h = y + \int_0^t X_s^h ds. \end{cases} \tag{1.14}$$

where $\phi(t) = t_i \ \forall t \in [t_i, t_{i+1})$. Observe that the above scheme is in fact well defined even though the non degenerate component of the scheme itself appears in the integral. On every time-step the increments of $(X_t^h, Y_t^h)_{t \in [t_i, t_{i+1}]}$, $i \geq 0$ are actually Gaussian. They indeed correspond to a suitable rescaling of the Brownian increment and its integral on the considered time step, see also Remark 4.2.3.

Let us also denote for a given $c > 0$ and for all $(x, y), (x', y') \in \mathbb{R}^{2d}$ the Kolmogorov-type density

$$p_{c,K}(t, (x, y), (x', y')) := \frac{c^d 3^{d/2}}{(2\pi t^2)^d} \exp\left(-c\left[\frac{|x' - x|^2}{4t} + 3\frac{|y' - y - (x + x')t/2|^2}{t^3}\right]\right) \tag{1.15}$$

The subscript K in the notation $p_{c,K}$ stands for Kolmogorov-like equations.

We would like to emphasize that in Chapter 4 we are considering time-homogeneous coefficients b, σ and specify assumptions precisely.

(**AD1**) **(Boundedness of the coefficients)**.

The components of the vector-valued function $b(x, y)$ and the matrix-valued function $\sigma(x, y)$ are bounded measurable. Specifically, there exists a constant K s.t.

$$\sup_{(x,y) \in \mathbb{R}^{2d}} |b(x, y)| + \sup_{(x,y) \in \mathbb{R}^{2d}} |\sigma(x, y)| \leq K.$$

(**AD2**) **(Uniform Ellipticity)**.

The matrix $a := \sigma\sigma^*$ is uniformly elliptic, i.e. there exists $\Lambda \geq 1$, $\forall (x, y, \xi) \in (\mathbb{R}^d)^3$,

$$\Lambda^{-1}|\xi|^2 \leq \langle a(x, y)\xi, \xi \rangle \leq \Lambda|\xi|^2.$$

(**AD3**) **(Hölder continuity in space)**.

For some $\gamma \in (0, 1]$, κ,

$$|b(x, y) - b(x', y')| + |\sigma(x, y) - \sigma(x', y')| \quad \leq \quad \kappa\left(|x - x'|^\gamma + |y - y'|^{\gamma/3}\right).$$

We say that assumption **(AD)** holds when conditions **(AD1)**-**(AD3)** are in force.

Under mentioned assumptions, we now introduce perturbed versions of (1.13) and (1.14). Namely, for $b_\varepsilon : \mathbb{R}^{2d} \to \mathbb{R}^d$, $\sigma_\varepsilon : \mathbb{R}^{2d} \to \mathbb{R}^d \otimes \mathbb{R}^d$ satisfy at least the same assumptions as b, σ and are in some sense meant to be *close* to b, σ for small values of $\varepsilon > 0$ one denote:

$$\begin{cases} dX_t^{(\varepsilon)} = b_\varepsilon(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dt + \sigma(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dW_t, \\ dY_t^{(\varepsilon)} = X_t^{(\varepsilon)}dt, t \in [0,T], \end{cases} \tag{1.16}$$

and similarly:

$$\begin{cases} X_t^{\varepsilon,h} = x + \int_0^t b_\varepsilon(X_{\phi(s)}^{\varepsilon,h}, Y_{\phi(s)}^{\varepsilon,h})ds + \int_0^t \sigma_\varepsilon(X_{\phi(s)}^{\varepsilon,h}, Y_{\phi(s)}^{\varepsilon,h})dW_s, \\ Y_t^{\varepsilon,h} = y + \int_0^t X_s^{\varepsilon,h}ds. \end{cases} \tag{1.17}$$

for $t \in [0, t_j), 0 < j \leq N$, where $\phi(t) = t_i \ \forall t \in [t_i, t_{i+1})$.

Considering as well a specific kind of Hölder continuity associated with the intrinsic scales of the system and the time-homogeneous case we set for $\varepsilon > 0$:

$$\forall q \in (1, +\infty], \ \Delta^d_{\varepsilon,b,q} := |b(.,.) - b_\varepsilon(.,.)|_{L^q(\mathbb{R}^{2d})}.$$

We also define

$$\Delta^d_{\varepsilon,\sigma,\gamma} := |\sigma(.,.) - \sigma_\varepsilon(.,.)|_{d,\gamma},$$

where for $\gamma \in (0,1]$, $|.|_{d,\gamma}$ stands for the Hölder norm in space on $C^\gamma_{b,\mathbf{d}}(\mathbb{R}^d \otimes \mathbb{R}^d)$, which denotes the space of Hölder continuous bounded functions with respect to the distance $\mathbf{d}$ defined as follows:

$$\forall (x,y), \ (x',y') \in (\mathbb{R}^d)^2, \ \mathbf{d}\big((x,y),(x',y')\big) := |x - x'| + |y' - y|^{1/3}.$$

Namely, a measurable function f is in $C^\gamma_{b,\mathbf{d}}(\mathbb{R}^d \otimes \mathbb{R}^d)$ if

$$|f|_{\mathbf{d},\gamma} := \sup_{x,y \in \mathbb{R}^{2d}} |f(x,y)| + [f]_{\mathbf{d},\gamma}, \ [f]_{d,\gamma} := \sup_{(x,y) \neq (x',y') \in \mathbb{R}^{2d}} \frac{|f(x,y) - f(x',y')|}{\mathbf{d}\big((x,y),(x',y')\big)^\gamma} < +\infty.$$

We eventually set $\forall q \in (1, +\infty]$,

$$\Delta^d_{\varepsilon,\gamma,q} := \Delta^d_{\varepsilon,\sigma,\gamma} + \Delta^d_{\varepsilon,b,q},$$

which will be the key quantity governing the error in our results.

Theorem 4.3.1. *Fix $T > 0$. Under **AD**, for $q \in (4d, +\infty]$, there exists $C := C(q) \geq 1, c \in (0,1]$ s.t. for all $0 < t \leq T, ((x,y),(x',y')) \in (\mathbb{R}^{2d})^2$:*

$$|(p - p_\varepsilon)(t,(x,y),(x',y'))| \leq C\Delta^d_{\varepsilon,\gamma,q} p_{c,K}(t,(x,y),(x',y')),$$

where $p(t,(x,y),(.,.)), p_\varepsilon(t,(x,y),(.,.))$ respectively stand for the transition densities at time t of equations (1.13), (1.16) *starting from (x,y) at time* 0.

Theorem 4.3.5. *Fix $T > 0$ and let us define a time-grid $\Lambda_h := \{(t_i)_{i\in[\![1,N]\!]}\}, N \in \mathbb{N}^*$. Under **AD**, there exists $C \geq 1, c \in (0,1]$ s.t. for all $0 < t_j \leq T, ((x,y),(x',y')) \in (\mathbb{R}^{2d})^2$:*

$$|p_h^\varepsilon - p_h|(t_j,(x,y),(x',y')) \leq C\Delta^d_{\varepsilon,\sigma,\gamma} p_{c,K}(t_j,(x,y),(x',y')),$$

where $p_h^\varepsilon(t,(x,y),(.,.)), p_h(t,(x,y),(.,.))$ respectively stand for the transition densities at time t of equations (1.14)*,* (1.17) *starting from (x,y) at time* 0.

These two theorems will be restated and discussed in Section 4.3.1.

The sensitivity analysis will then be applied, in the flavour of [KM17] to investigate the weak error associated to a specific Euler scheme which had already been considered in [LM10] for equations of type (1.13). However, to perform the analysis we need to change assumptions **(AD)** slightly. Precisely, we have to assume more about Hölder properties of coefficients than in **(AD)**.

Instead of (**AD3**), we assume for some $\gamma \in (0,1]$, κ,

$$|b(x,y) - b(x',y')| + |\sigma(x,y) - \sigma(x',y')| \quad \leq \quad \kappa\left(|x-x'|^\gamma + |y-y'|^{\gamma/2}\right).$$

and denote that as ($\hat{\mathbf{A}}$**D3**). Thus, we say that assumption **($\hat{\mathbf{A}}$D)** holds when conditions **(AD1)**,**(AD2)**, **($\hat{\mathbf{A}}$D3)** are in force.

Theorem 4.4.1. *Fix $T > 0$. Under assumptions **($\hat{A}$D)** for any test function $f \in C^{\beta,\beta/2}(\mathbb{R}^{2d})$ ($\beta-$Hölder in the first variable and $\beta/2-$Hölder in the second variable functions) for $\beta \in (0,1]$, there exists $C > 0$, such that:*

$$|\mathbb{E}_{(x,y)}[f(X_T^h, Y_T^h)] - \mathbb{E}_{(x,y)}[f(X_T,Y_T)]| \leq Ch^{\gamma/2}(1+|x|^{\gamma/2}).$$

where $\gamma \in (0,1]$ stands for the Hölder index of $\gamma, \gamma/2$ Hölder continuous time-homogeneous functions b,σ.

The theorem will be restated in Section 4.4.

We also would like to present our control for the direct difference of transition densities $p(t,(x,y),(x',y'))$ and $p_h(t,(x,y),(x',y'))$. The result below is in clear contrast with the one of Theorem 4.4.1 for the weak error, i.e. when additionally consider an integration of a Hölder function w.r.t. the final (or forward variable). We finally can reach a global error of order $h^\beta, \beta < \gamma - 1/2$ which is *close* to the expected one in $h^{\gamma/2}$ when γ goes to 1.

To improve the above result, we feel that some new advanced approaches to error analysis should be considered. This means that either the scheme would have to be modified or the error decomposed very differently than in the current huge literature (from the seminal papers of [KM00] and [KM02] the same lines are considered for the error decomposition, see e.g. [KM10], [KM17], [Fri18]). Eventually, a specific difficulty of the current model consists in dealing the unbounded transport term.

Theorem 4.5.1. *Fix a final time horizon $T > 0$ and a time step $h = T/N, N \in \mathbb{N}^*$ for the Euler scheme. Under assumptions **($\hat{\boldsymbol{A}}$D)**, for $\gamma \in (1/2, 1]$ and $\beta \in (0, \gamma - \frac{1}{2})$, for all t in the time grid $\Lambda_h := \{(t_i)_{i \in [\![1,N]\!]}\}$ and $(x, y), (x', y') \in \mathbb{R}^{2d}$ there exist $C := (T, b, a, \beta), c > 0$ such that :*

$$\begin{aligned} &|p(t,(x,y),(x',y')) - p_h(t,(x,y),(x',y')| \\ &\le Ch^\beta(1 + (|x| \wedge |x'|))^{1+\gamma}) \sup_{s\in[t-h,t]} p_{c,K}(s,(x,y),(x',y')), \end{aligned} \tag{1.18}$$

where $p_{c,K}(s, (x, y), (x', y'))$ stands for the Kolmogorov-type Gaussian density (1.15) *at time s.*

The theorem will be discussed in Section 4.5.

Chapter 2

Parametrix technique

Since the main topic of the thesis is the parametrix technique, its developments and applications – we would like to start with the short history review on it.

2.1 Review

The parametrix method itself is a classical method in order to construct fundamental solutions for parabolic type partial differential equations using an expansion argument. This method allows for coefficients to be less regular than in the Malliavin Calculus approach. However, the methodology is restricted to cases where the underlying process is Markov.

The parametix approach has been established in the beginning of XX century as a perturbation technique for partial differential equations theory by Levi [Lev07]. The original method has been used for approximations of the elliptic linear differential equation solution. In a nutshell the idea consisted of the appropriate separation of the “main” part and controlling the “remainder”. A common choice for the principal part consists in considering the solution of the underlying equation with constant coefficients.

The technique has been further developed by Hadamard [Had23]. The important modifications of the parametrix method, introduced by Il’in et. all in 1962 [IKO62], Friedman in 1964 [Fri64] and McKean and I. Singer in 1967 [MS67], provided the way to use it for SDEs theory. The main point is that the transition density of the SDE can be found through the fundamental solution of the Cauchy problem for the corresponding generator. Parametrix in a nutshell allows to get a representation of the SDE transition density as a sum, where each term contains the transition density of the more simple Gaussian process. Moreover the method of McKean and Singer does not require any regularity on the coefficients of the SDE, besides Hölder continuity (although it was initially presented under C^∞ assumptions).

As far as we know one of the first investigation of the parametrix method for Markov chains was presented in the paper by Konakov and Mammen in 2000 [KM00], however local limit theorems for homogeneous Markov chains with continuous state space have been already given in Konakov and Molchanov [KM85] but the last article was not really well-known being published in Russian. In the article [KM00] Konakov and Mammen applied the parametrix method for parabolic PDEs and a modification of the method – for discrete time Markov chain. As the result they achieved the convergence rate of order $O(n^{-1/2})$ for transition densities of triangular array of Markov Chains to the transition density of the limiting diffusion.

After that, in 2001, the same authors considered the situation of triangular arrays of Markov random walks that can be approximated by an accompanying sequence of diffusion processes. The main result consisted in proving that normalized transition probabilities differ from transition densities in the diffusion model by rate $O(n^{-1/2})$. In particular, local limit theorems for the case that the Markov random walks has been stated and proved. As in [KM00] the approach was based on application of the parametrix method.

In 2002 Konakov and Mammen studied the approximation of the density of the diffusion by the density of the Euler discretization with discretisation step $\frac{1}{n}$ in [KM02]. Using the parametrix approach they obtained an asymptotic expansion in powers of $\frac{1}{n}$.

After these improvements some results about Edgeworth type expansions for transition densities were achieved also by applying the parametrix expansions. In 2005 the article [KM05] appeared with the discussion on Edgeworth type expansion for the transition densities of triangular arrays of homogeneous Markov chains X_n^k that converge weakly to the diffusion process.

As the generalization for the previous article the paper [KM07] was published in 2007. The improvements have been done mainly in two directions: the time horizon T was allowed to converge to 0 and also cases are treated with non - homogeneous diffusion limit.

Now we come to the key paper on the parametrix topic which is also important for our current research. In 2010, Konakov, Menozzi and Molchanov presented the paper [KMM10], where parametrix method has been adapted for a larger class of processes - namely for degenerate diffusions with rank 2. However not only the density representation in terms of parametrix series has been given but also the explicit Gaussian upper and partial lower bounds has been derived. Due to the series representation, the authors have provided also a local limit theorem with the usual convergence rate for an associated Markov chains approximations.

As the continuation in this direction the article [DM10] was published by Delarue and Menozzi with the full summary on the existing parametrix techniques including the degenerate case. Two sided bounds for the density of the solution of a system of n differential equations of dimension d have been provided.

In the frame of degenerate diffusions case discussion we also would like to emphasize the paper [Men11] due to the proof of the uniqueness of the martingale problem associated to some degenerate operator inside. The key point is to exploit the strong parallel between the new technique introduced by Bass and Perkins [BP09] to prove uniqueness of the martingale problem in the framework of non degenerated elliptic operators and the parametrix approach to the density expansion.

Later, Konakov and Menozzi also considered a stable driven SDEs cases for the parametrix application [KM10]. Using a parametrix approach they derived an expansion for the difference between the diffusion and the Euler scheme densities.

Even more general framework has been treated in the paper by Menozzi and Huang [HM16] in 2016 where they considered a stable driven degenerate stochastic differential equation, whose coefficients satisfy a kind of weak Hörmander condition. Under mild smoothness assumptions they proved the uniqueness of the martingale problem for the associated generator using also the parametrix method as a tool.

2.2 Other developments in Parametrix

Bally and Kohatsu-Higa in their paper[BA09] introduced the parametrix method using a semigroup approach and obtain the probabilistic representation for the density of the solution to a diffusion equation or for Levy driven SDEs. It's worth to specifically emphasize that they have described two kinds of parametrix methods: the first one - "forward" and second one - "backward" parametrix. To use the first version it's necessary to assume that the coefficients are C_b^2. The second version converges if the drift coefficient is bounded and measurable and diffusion coefficient is bounded, uniformly elliptic and Hölder continuous.

In his further articles Kohatsu-Higa with his co-authors consider an unbiased simulation method for multidimensional diffusions based on the parametrix method for solving partial differential equations with Hölder continuous coefficients [AKH17].

And also studied the parametrix approach applied to so-called skew diffusions to obtain the existence and the regularity properties of the density and to provide a Gaussian upper bound [KHTZ16].

As a continuation of the topic we would like to mention a paper by N. Frikha [Fri18]. Applying the results obtained in [AKH17] the author studied the weak approximation error of a skew diffusion with bounded measurable drift and Hölder diffusion coefficient by an Euler-type scheme. A bound for the difference between the densities of the skew diffusion and its Euler approximation was obtained using the parametrix method for the skew diffusions.

Moreover we would like to emphasize other two papers as nice examples of the parametrix application. First, in [FH15] studying the development of the Richardson-Romberg extrapolation method for Monte Carlo linear estimator to the framework of stochastic optimization by means of stochastic approximation algorithm, authors also extended the parametrix expansion results to the derivatives of the densities.

Second, in [FKHL16] authors obtained properties of the law associated to the first hitting time of a threshold by a one-dimensional uniformly elliptic diffusion process and to the associated process stopped at the threshold. The methodology relied on the parametrix method that was applied to the associated Markov semigroup.

There is also an interesting direction for the parametrix applications developed mostly by A. Pascucchi. Although there is a big variety of papers done more or less in a same flavour it is worth to mention at least the key one. In [FP10] authors introduced their own view on the diffusions transition densities approximations deriving the technique also from the classical PDE theory. Moreover, authors provided a way to use the parametrix analogue for pricing and hedging of financial derivatives.

2.3 Diffusions

To introduce the technique we would like to start the non-degenerate SDEs. Namely, as in the Chapter 1, for a fixed given deterministic final time-horizon $T > 0$, we consider the following multidimensional SDE:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, t \in \ [0, T], \tag{2.1}$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded coefficients that are measurable in time and W is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Also, $a(t, x) := \sigma\sigma^*(t, x)$ is assumed to be uniformly elliptic, precisely, there exists $\lambda_0 \geq 1$ s.t. for $(t, x, \xi) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ we have $\lambda_0^{-1}|\xi|^2 \leq \langle a(t, x)\xi, \xi\rangle \leq \lambda_0|\xi|^2$ where $|.|$ stands for the Euclidean norm.

To begin with, we assume that there exists the so-called transition density of (2.1) which is a fundamental solution associated with the operator $\partial t + L_t$, where L_t is the generator of (2.1) (to get this one can assume Lipschitz continuity for coefficients in time and space, i.e.). Precisely, as in the Chapter 1, for all $\phi \in C_0^2(\mathbb{R}^d, \mathbb{R}), z \in \mathbb{R}^d$:

$$L_t\phi(z) = \frac{1}{2}\mathrm{Tr}(a(t, z)D_z^2\phi(z)) + \langle b(t, z), \nabla_z\phi(z)\rangle. \tag{2.2}$$

The existence of the Markov process with such a generator even in the case of β Hölder continuity, $0 < \beta \leq 1$ for a and b has been proven in [SV79]. The existence of the transition density $\mathbb{P}(X_s \in dy|X_t = x) = p(s, t, x, y)dy$ has been proved in the book [Fri64], for example.

As it was mentioned in the Chapter 1, are interested in the approximation of the solution X_t for the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, t \in \ [0, T], \tag{2.3}$$

For given $(s,x) \in \mathbb{R}^+ \times \mathbb{R}^d$, we use the standard Markov notation $(X_t^{s,x})_{t\geq s}$ to denote the solution of (2.3) starting from x at time s.

Assume that $(X_t^{s,x})_{t\geq s}$ has for all $t > 0$ a smooth density $p(t,x,.)$ (which is the case if the coefficients are smooth see e.g. Friedman [Fri64]). We would like to estimate this density at a given point $y \in \mathbb{R}^d$. To this end, we introduce the following Gaussian inhomogeneous process with spatial variable frozen at y. For all $(s,x) \in [0,T]\times\mathbb{R}^d,\ t \geq 0$ we set:

$$\tilde{X}_t^y = x + \int_s^t \sigma(u,y)dW_u,$$

which literally means we freeze the coefficient $\sigma()$ at the terminal point $x \in \mathbb{R}^d$. However it is worth to remark that the approach used by Levi [Lev07] and [IKO62] applied the freezing procedure to the initial point (which seems to be even more natural decision). The disadvantage of such a solution is that one needs to assume additional regularity in time. Also the specific form without a trend coefficient is used due to the boundedness assumption on $b(t,X_t)$. In the general case one should add $\int_s^t b(u,y)du$ to the definition of the frozen process.

The density of the frozen process $\tilde{p}^y$, which exists due to the uniform ellipticity assumptions on σ, seems to be the most natural "proxy" for the initial density $p(s,t,x,y)$. It's possible to quantify the distance between them using the Kolmogorov equations.

Assume for the beginning smoothness for coefficients in (2.3) and that $(X_t)_{t>0}$ has a smooth density. The density of the frozen process satisfies the Kolmogorov Backward equation:

$$\begin{cases} \partial_u \tilde{p}^y(u,t,z,y) + \tilde{L}_u^y \tilde{p}^y(u,t,z,y) = 0,\ \ s \leq u < t, z \in \mathbb{R}^d, \\ \tilde{p}^y(u,t,.,y) \underset{u\uparrow t}{\rightarrow} \delta_y(.), \end{cases} \tag{2.4}$$

where for all $\varphi \in C_0^2(\mathbb{R}^d,\mathbb{R}), z \in \mathbb{R}^d$:

$$\tilde{L}_u^y \varphi(z) = \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^*(u,y)D_z^2\varphi(z)\right),$$

stands for the generator of $\tilde{X}^y$ at time u.

On the other hand, since we have assumed the density of X to be smooth, it must satisfy the Kolmogorov forward equation (see e.g. Dynkin [Dyn65]). For a given starting point $x \in \mathbb{R}^d$ at time s,

$$\begin{cases} \partial_u p(s,u,x,z) = L_u^* p(s,u,x,z) = 0,\ \ s < u \leq t, z \in \mathbb{R}^d, \\ p(s,u,x,.) \underset{u\downarrow s}{\rightarrow} \delta_x(.), \end{cases} \tag{2.5}$$

where L_u^* stands for the *formal* adjoint (which is again well defined if the coefficients in (2.1) are smooth) of the generator of (2.1) (see (2.2)).

Using the Dirac convergences in (2.4) and (2.5) one can derive that:

$$(p - \tilde{p}^y)(s,t,x,y) = \int_s^t du \partial_u \int_{\mathbb{R}^d} dz p(s,u,x,z)\tilde{p}^y(u,t,z,y).$$

After a formal differentiating:

$$(p - \tilde{p}^y)(s,t,x,y) = \int_s^t du \int_{\mathbb{R}^d} dz \left(\partial_u p(s,u,x,z)\tilde{p}^y(u,t,z,y) + p(s,u,x,z)\partial_u \tilde{p}^y(u,t,z,y)\right).$$

Equations (2.4), (2.5) yield the formal expansion below which is initially due to McKean and Singer [MS67].

$$\begin{aligned}(p - \tilde{p}^y)(s,t,x,y) &= \int_s^t du \int_{\mathbb{R}^d} dz \left(L_u^* p(s,u,x,z)\tilde{p}^y(u,t,z,y) - p(s,u,x,z)\tilde{L}_u^y \tilde{p}^y(u,t,z,y)\right) \\ &= \int_s^t du \int_{\mathbb{R}^d} dz p(s,u,x,z)(L_u - \tilde{L}_u^y)\tilde{p}^y(u,t,z,y). \qquad (2.6)\end{aligned}$$

We eventually take the adjoint for the last equality. Note carefully that the differentiation under the integral is also here formal since we would need to justify that it can actually be performed using some growth properties of the density and its derivatives which we *a priori* do not know. Let us now introduce the notation

$$f \otimes g(s,t,x,y) = \int_s^t du \int_{\mathbb{R}^d} dz f(s,u,x,z) g(u,t,z,y)$$

for the time-space convolution and let us define $\tilde{p}(s,t,x,y) := \tilde{p}^y(s,t,x,y)$, that is in $\tilde{p}(s,t,x,y)$ we consider the density of the frozen process at the final point and observe it at *that specific* point. We now introduce the *parametrix* kernel:

$$H(s,t,x,y) := (L_s - \tilde{L}_s)\tilde{p}(s,t,x,y) := (L_s - \tilde{L}_s^y)\tilde{p}^y(s,t,x,y). \qquad (2.7)$$

With those notations equation (2.6) rewrites:

$$(p - \tilde{p})(s,t,x,y) = p \otimes H(s,t,x,y).$$

From this expression, the idea then consists in iterating this procedure for $p(s,u,x,z)$ in (2.6) introducing the density of a process with frozen characteristics in z which is here the integration variable. This yields to iterated convolutions of the kernel and leads to the formal expansion:

$$p(s,t,x,y) = \sum_{r=0}^{\infty} \tilde{p} \otimes H^{(r)}(s,t,x,y), \qquad (2.8)$$

where $\tilde{p} \otimes H^{(0)} = \tilde{p}, H^{(r)} = H \otimes H^{(r-1)},\ r \geq 1$. Obtaining estimates on p from the formal expression (2.8) requires to have good controls on the right-hand side. The remarkable property of this formal expansion is now that the right-hand-side of (2.8) only involves controls on Gaussian densities.

Observe that up to now we have used the smoothness assumption on the coefficients a lot, however it's possible to have the same representation under the Hölder assumptions only (see the Chapter 3).

The convergence of the series in (2.8) is in some sense *standard* (see e.g. [Men11] or Friedman [Fri64]). We recall for the sake of completeness the key steps.

From direct computations, there exist $c_1 \geq 1, c \in (0,1]$ s.t. for all $T > 0$ and all multi-index $\alpha, |\alpha| \leq 8$,

$$\forall 0 \leq u < t \leq T,\ (z,y) \in (\mathbb{R}^d)^2,\ |D_z^{\alpha}\widetilde{p}(u,t,z,y)| \leq \frac{c_1}{(t-u)^{|\alpha|/2}} p_c(t-u, y-z), \quad (2.9)$$

where

$$p_c(t-u, y-z) = \frac{c^{d/2}}{(2\pi(t-u))^{d/2}} \exp\left(-\frac{c}{2}\frac{|y-z|^2}{t-u}\right),$$

stands for the usual Gaussian density in $\mathbb{R}^d$ with 0 mean and covariance $(t-u)c^{-1}I_d$. From (2.9), the boundedness of the drift and the Lipschitz continuity in space of the diffusion matrix we readily get that there exists $c_1 \geq 1,\ c \in (0,1]$,

$$|H(u,t,z,y)| \leq \frac{c_1(1 \vee T^{1/2})}{(t-u)^{1/2}} p_c(t-u, z-y).$$

Now we present the property which usually called *the smoothing property of the parametrix kernel.* Let us illustrate this by deriving the time-singularuty of the first order-convolution.

$$\begin{aligned}|\widetilde{p} \otimes H(s,t,x,y)| &\leq ((1 \vee T^{1/2})c_1)^2 B(\frac{1}{2}, 1) p_c(t-s, y-x)(t-s)^{\frac{1}{2}} \\ &= \frac{((1 \vee T^{1/2})c_1)^2 \left[\Gamma(\frac{1}{2})\right]}{\Gamma(\frac{3}{2})} p_c(t-s, y-x)(t-s)^{\frac{1}{2}},\end{aligned}$$

where for $a, b > 0,\ B(a,b) = \int_0^1 t^{-1+a}(1-t)^{-1+b}dt$ stands for the $\beta-$ function, and using as well the identity $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ for the last inequality. Iterating the convolution operation, the exponent in time will grow with each iteration:

$$\begin{aligned}|\widetilde{p} \otimes H^{(r)}(s,t,x,y)| &\leq ((1 \vee T^{1/2})c_1)^{r+1} \prod_{i=1}^{r} B(\frac{1}{2}, \frac{i+1}{2}) p_c(t-s, y-x)(t-s)^{\frac{r}{2}} \\ &= \frac{((1 \vee T^{1/2})c_1)^{r+1} \left[\Gamma(\frac{1}{2})\right]^r}{\Gamma(1+\frac{r}{2})} p_c(t-s, y-x)(t-s)^{\frac{r}{2}}.\end{aligned}$$

These bound due to the asymptotic of the Gamma function readily yield the convergence of the series as well as a Gaussian upper-bound. Namely

$$p(s,t,x,y) \leq c_1 \exp((1 \vee T^{1/2}) c_1 [(t-s)^{1/2}]) p_c(t-s, y-x). \tag{2.10}$$

The upper bound enjoy the semigroup property, i.e. $\forall 0 \leq s < u < t \leq T$,

$$\int_{\mathbb{R}^d} p_c(u-t, z-x) p_c(s-u, y-z) dz = p_c(t-s, y-x),$$

which allows to propagate the upper bound (2.10) from small times to arbitrary but finite time.

2.4 Markov Chains

One of the main advantages of the formal expansion in (2.8) is that it has a direct discrete counterpart in the Markov chain setting. Indeed, denote by $(Y_{t_j}^{t_i,x})_{j\geq i}$ the Markov chain starting from x at time t_i with dynamics:

$$Y_{t_{k+1}} = Y_{t_k} + b(t_k, Y_{t_k})h + \sigma(t_k, Y_{t_k})\sqrt{h}\xi_{k+1}, Y_0 = x, \tag{2.11}$$

where $h > 0$ is a given time step, for which we denote for all $k \geq 0, t_k := kh$ and the $(\xi_k)_{k\geq 1}$ are centered i.i.d. random variables satisfying some integrability conditions. Observe first that if the innovations $(\xi_k)_{k\geq 1}$ have a density then so does the chain at time t_k.

Let us now introduce its generator at time t_i, i.e. for all $\varphi \in C_0^2(\mathbb{R}^d, \mathbb{R}),\ x \in \mathbb{R}^d$:

$$L_{t_i}^h \varphi(x) := h^{-1}\mathbb{E}[\varphi(Y_{t_{i+1}}^{t_i,x}) - \varphi(x)].$$

In order to give a representation of the density of $p^h(t_i, t_j, x, y)$ of $Y_{t_j}^{t_i,x}$ at point y for $j > i$, we introduce similarly to the continuous case, the Markov chain (or inhomogeneous random walk) with coefficients frozen in space at y. For given $(t_i, x) \in [0,T] \times \mathbb{R}^d,\ t_j \geq t_i$ we set:

$$\tilde{Y}_{t_j}^{t_i,x,y} := x + h^{1/2} \sum_{k=i}^{j-1} \sigma(t_k, y)\xi_{k+1},$$

and denote its density $\tilde{p}^{h,y}(t_i, t_j, x, .)$. Its generator at time t_i writes for all $\varphi \in C_0^2(\mathbb{R}^d, \mathbb{R}), x \in \mathbb{R}^d$:

$$\tilde{L}_{t_i}^{h,y} \varphi(x) = h^{-1}\mathbb{E}[\varphi(\tilde{Y}_{t_{i+1}}^{t_i,x,y}) - \varphi(x)].$$

Using the notation $\tilde{p}^h(t_i, t_j, x, y) := \tilde{p}^{h,y}(t_i, t_j, x, y)$, we introduce now for $0 \leq i < j \leq N$ the *parametrix* kernel:

$$H^h(t_i, t_j, x, y) := (L_{t_i}^h - \tilde{L}_{t_i}^{h,y})\tilde{p}^h(t_i + h, t_j, x, y).$$

Analogously to Lemma 3.6 in [KM00], which follows from a direct algebraic manipulation, we derive the following representation for the density. Assuming boundness and Lipschitz continuity for coefficients b and σ with the uniform ellipticity for σ, one can get for $0 \le t_i < t_j \le T$ that

$$p^h(t_i, t_j, x, y) = \sum_{r=0}^{j-i} \tilde{p}^h \otimes_h H^{h,(r)}(t_i, t_j, x, y), \tag{2.12}$$

where the discrete time convolution type operator $\otimes_h$ is defined by

$$f \otimes_h g(t_i, t_j, x, y) = \sum_{k=0}^{j-i-1} h \int_{\mathbb{R}^d} f(t_i, t_{i+k}, x, z) g(t_{i+k}, t_j, z, y) dz.$$

Also $g \otimes_h H^{h,(0)} = g$ and for all $r \ge 1, H^{h,(r)} := H^h \otimes_h H^{h,(r-1)}$ denotes the r-fold discrete convolution of the kernel H^h.

The key point to prove is the direct definition of the discrete kernel function. Since

$$H^h(t_i, t_j, x, y) = \int_{\mathbb{R}^d} h^{-1}[p^h(t_i, t_{i+1}, x, z) - \tilde{p}^h(t_i, t_{i+1}, x, z)]\tilde{p}^h(t_{i+1}, t_j, z, y) dz.$$

Using the Markov property we get the following identity:

$$\begin{aligned} p^h(t_i, t_j, x, y) - \tilde{p}^h(t_i, t_j, x, y) &= \sum_{k=i}^{j-1} h \int_{\mathbb{R}^d} p^h(t_i, t_k, x, z) \\ \times \int_{\mathbb{R}^d} h^{-1}[p^h(t_k, t_{k+1}, z, z') &- \tilde{p}^h(t_k, t_{k+1}, z, z')]\tilde{p}^h(t_{k+1}, t_j, z', y) dz' dz \\ &= \sum_{k=i}^{j-1} h \int_{\mathbb{R}^d} p^h(t_i, t_k, x, z) H^h(t_k, t_j, z, y) dz \\ &= (p^h \otimes_h H^h)(t_i, t_j, x, y). \end{aligned}$$

The expansion (2.12) follows by iterative application of this identity.

Chapter 3

Stability of diffusion transition densities

3.1 Introduction

In this Chapter we study the sensitivity of densities of non-degenerate diffusion processes and related Markov Chains with respect to a perturbation of the coefficients. Natural applications of these results appear in models with misspecified coefficients or for the investigation of the weak error of the Euler scheme with irregular coefficients.

3.2 Stability

For a fixed given deterministic final horizon $T > 0$, let us consider, as in the Chapter 2, the following multidimensional SDE:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t,\ t \in [0, T], \tag{3.1}$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d,\ \sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded coefficients that are measurable in time and Hölder continuous in space (this last condition will be possibly relaxed for the drift term b) and W is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Also, $a(t, x) := \sigma\sigma^*(t, x)$ is assumed to be uniformly elliptic. In particular those assumptions guarantee that (3.1) admits a unique weak solution, see e.g. Bass and Perkins [BP09], [Men11] from which the uniqueness to the martingale problem for the associated generator can be derived under the current assumptions.

We now introduce, for a given parameter $\varepsilon > 0$, a perturbed version of (3.1) with dynamics:

$$dX_t^{(\varepsilon)} = b_\varepsilon(t, X_t^{(\varepsilon)})dt + \sigma_\varepsilon(t, X_t^{(\varepsilon)})dW_t,\ t \in [0, T], \tag{3.2}$$

where $b_\varepsilon : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_\varepsilon : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ satisfy at least the same assumptions as b, σ and are in some sense meant to be *close* to b, σ when ε is small.

It is known that, under the previous assumptions, the density of the processes $(X_t)_{t\geq 0}, (X_t^{(\varepsilon)})_{t\geq 0}$ exists and satisfies some Gaussian bounds, see e.g Aronson [Aro59] or [DM10] for extensions to some degenerate cases.

The goal of this Chapter is to investigate how the closeness of $(b_\varepsilon, \sigma_\varepsilon)$ and (b, σ) is reflected on the respective densities of the associated processes. Important applications can for instance be found in mathematical finance. If the dynamics of (3.1) models the evolution of the (log-)price of a financial asset, it is often very useful to know how a perturbation of the volatility σ impacts the density, and therefore the associated option prices.

In the framework of parameter estimation it can be useful, having at hand estimators $(b_\varepsilon, \sigma_\varepsilon)$ of the true parameters (b, σ) and some controls for the differences $|b - b_\varepsilon|, |\sigma - \sigma_\varepsilon|$ in a suitable sense, to quantify the difference $p_\varepsilon - p$ of the densities corresponding respectively to the dynamics with the estimated parameters and the one of the model.

Another important application includes the case of mollification by spatial convolution. This specific kind of perturbation is useful to investigate the error between the densities of a non-degenerate diffusion of type (3.1) with Hölder coefficients (or with piecewise smooth bounded drift) and its Euler scheme. In this framework, some explicit convergence results can be found in [KM17].

More generally, this situation can appear in every applicative field for which the diffusion coefficient might be misspecified.

Our stability results will also apply to two Markov chains with respective dynamics:

$$\begin{aligned} Y_{t_{k+1}} &= Y_{t_k} + b(t_k, Y_{t_k})h + \sigma(t_k, Y_{t_k})\sqrt{h}\xi_{k+1}, Y_0 = x, \\ Y_{t_{k+1}}^{(\varepsilon)} &= Y_{t_k}^{(\varepsilon)} + b_\varepsilon(t_k, Y_{t_k}^{(\varepsilon)})h + \sigma_\varepsilon(t_k, Y_{t_k}^{(\varepsilon)})\sqrt{h}\xi_{k+1}, Y_0^{(\varepsilon)} = x, \end{aligned} \tag{3.3}$$

where $h > 0$ is a given time step, for which we denote for all $k \geq 0$, $t_k := kh$ and the $(\xi_k)_{k\geq 1}$ are centred i.i.d. random variables satisfying some integrability conditions. Again, the key tool will be the parametrix representation for the densities of the chains and the Gaussian local limit theorem.

3.2.1 Assumptions and Main Results.

For better readability let us now repeat assumptions, introduced in Chapter 1, which we use during this Section. Below, the parameter $\varepsilon > 0$ is fixed and the constants appearing in the assumptions **do not depend** on ε.

(A1) (Boundedness of the coefficients). The components of the vector-valued functions $b(t,x), b_\varepsilon(t,x)$ and the matrix-functions $\sigma(t,x), \sigma_\varepsilon(t,x)$ are bounded measurable. Specifically, there exist constants $K_1, K_2 > 0$ s.t.

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |b(t,x)| + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |b_\varepsilon(t,x)| \le K_1,$$
$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\sigma(t,x)| + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\sigma_\varepsilon(t,x)| \le K_2.$$

(**A2**) **(Uniform Ellipticity)**. The matrices $a := \sigma\sigma^*, a_\varepsilon := \sigma_\varepsilon\sigma_\varepsilon^*$ are uniformly elliptic, i.e. there exists $\Lambda \ge 1,\ \forall (t,x,\xi) \in [0,T] \times (\mathbb{R}^d)^2$,

$$\Lambda^{-1}|\xi|^2 \le \langle a(t,x)\xi, \xi\rangle \le \Lambda|\xi|^2, \Lambda^{-1}|\xi|^2 \le \langle a_\varepsilon(t,x)\xi, \xi\rangle \le \Lambda|\xi|^2.$$

(**A3**) **(Hölder continuity in space)**. For some $\gamma \in (0,1]$, $\kappa < \infty$, for all $t \in [0,T]$,

$$|\sigma(t,x) - \sigma(t,y)| + |\sigma_\varepsilon(t,x) - \sigma_\varepsilon(t,y)| \quad \le \quad \kappa\, |x-y|^\gamma\, .$$

Observe that the last condition also readily gives, thanks to the boundedness of $\sigma, \sigma_\varepsilon$ that a, a_ε are also uniformly γ-Hölder continuous.

For a given $\varepsilon > 0$, we say that assumption **(A)** holds when conditions **(A1)**-**(A3)** are in force. Let us now introduce, under **(A)**, the quantities that will bound the difference of the densities in our main results below. Set for $\varepsilon > 0$:

$$\begin{aligned} \Delta_{\varepsilon,b,\infty} &:= \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \{|b(t,x) - b_\varepsilon(t,x)|\},\ \forall q \in (1,+\infty), \\ \Delta_{\varepsilon,b,q} &:= \sup_{t\in[0,T]} \|b(t,.) - b_\varepsilon(t,.)\|_{L^q(\mathbb{R}^d)}. \end{aligned}$$

Since $\sigma, \sigma_\varepsilon$ are both γ-Hölder continuous, see (**A3.3**) we also define

$$\Delta_{\varepsilon,\sigma,\gamma} := \sup_{u\in[0,T]} |\sigma(u,.) - \sigma_\varepsilon(u,.)|_\gamma,$$

where for $\gamma \in (0,1]$, $|.|_\gamma$ stands for the usual Hölder norm in space on $C_b^\gamma(\mathbb{R}^d, \mathbb{R}^d \otimes \mathbb{R}^d)$ (space of Hölder continuous bounded functions, see e.g. Krylov [Kry96]) i.e. :

$$|f|_\gamma := \sup_{x\in\mathbb{R}^d} |f(x)| + [f]_\gamma,\ [f]_\gamma := \sup_{x\ne y, (x,y)\in(\mathbb{R}^d)^2} \frac{|f(x) - f(y)|}{|x-y|^\gamma}.$$

The previous control in particular implies for all $(u,x,y) \in [0,T] \times (\mathbb{R}^d)^2$:

$$|a(u,x) - a(u,y) - a_\varepsilon(u,x) + a_\varepsilon(u,y)| \le 2(K_2 + \kappa)\Delta_{\varepsilon,\sigma,\gamma}|x-y|^\gamma.$$

We eventually set for $q \in (1,+\infty]$,

$$\Delta_{\varepsilon,\gamma,q} := \Delta_{\varepsilon,\sigma,\gamma} + \Delta_{\varepsilon,b,q}, \tag{3.4}$$

which will be the key quantity governing the error in our results.

We will denote, from now on, by C a constant depending on the parameters appearing in **(A)** and T. We reserve the notation c for constants that only depend on **(A)** but not on T. The values of C, c may change from line to line and **do not depend on the considered parameter** ε. Also, for given integers $i, j \in \mathbb{N}$ s.t. $i < j$, we will denote by $[\![i, j]\!]$ the set $\{i, i+1, \cdots, j\}$.

We are now in position to state main results of this Chapter, which we have already mentioned in Chapter 1.

Theorem 3.2.1. *Fix $\varepsilon > 0$ and a final deterministic time horizon $T > 0$. Under **(A)** and for $q > d$, there exist $C := C(q) \geq 1, c := c(q) \in (0,1]$ s.t. for all $0 \leq s < t \leq T, (x,y) \in (\mathbb{R}^d)^2$:*

$$p_c(t-s, y-x)^{-1}|(p - p_\varepsilon)(s,t,x,y)| \leq C\Delta_{\varepsilon,\gamma,q}, \tag{3.5}$$

where $p(s,t,x,.), p_\varepsilon(s,t,x,.)$ respectively stand for the transition densities at time t of equations (3.1), (3.2) *starting from x at time s. Also, we denote for a given $c > 0$ and for all $(u,z) \in \mathbb{R}^+ \times \mathbb{R}^d$, $p_c(u,z) := \frac{c^{d/2}}{(2\pi u)^{d/2}} \exp(-c\frac{|z|^2}{2u})$.*

The proof will be given in Section 3.2.4.

Remark 3.2.1 (About the constants). We mention that the constant $C := C(q)$ in (3.5) explodes when $q \downarrow d$ and is decreasing in q. In particular, it can be chosen uniformly as soon as $q \geq q_0 > d$.

Before stating our results for Markov Chains we introduce two kinds of innovations in (3.3). Namely:

(**IG**) The i.i.d. random variables $(\xi_k)_{k\geq 1}$ are Gaussian, with law $\mathcal{N}(0, I_d)$, where I_d stands for the identity matrix of size $d \times d$. In that case the dynamics in (3.3) correspond to the Euler discretization of equations (3.1) and (3.2).

(**IP**) For a given integer $M > 2d + 5 + \gamma$, the innovations $(\xi_k)_{k\geq 1}$ are centred and have C^5 density f_ξ which has, together with its derivatives up to order 5, at most polynomial decay of order M. Namely, for all $z \in \mathbb{R}^d$ and multi-index $\nu, |\nu| \leq 5$:

$$|D^\nu f_\xi(z)| \leq CQ_M(z), \tag{3.6}$$

where we denote for all $r > d$, $z \in \mathbb{R}^d$, $Q_r(z) := c_r\frac{1}{(1+|z|)^r}$, $\int_{\mathbb{R}^d} dz Q_r(z) = 1$.

Theorem 3.2.2 (Stability Control for Markov Chains). *Fix $\varepsilon > 0$ and a final deterministic time horizon $T > 0$. For $h = T/N$, $N \in \mathbb{N}^*$, we set for $i \in \mathbb{N}$, $t_i := ih$. Under **(A)**, assuming that either(**IG**) or (**IP**) holds, and for $q > d$ there exist $C := C(q) \geq 1, c := c(q) \in (0,1]$ s.t. for all $0 \leq t_i < t_j \leq T, (x,y) \in (\mathbb{R}^d)^2$:*

$$\chi_c(t_j - t_i, y-x)^{-1}|(p^h - p^h_\varepsilon)(t_i, t_j, x, y)| \leq C\Delta_{\varepsilon,\gamma,q}, \tag{3.7}$$

where $p^h(t_i,t_j,x,.), p^h_\varepsilon(t_i,t_j,x,.)$ respectively stand for the transition densities at time t_j of the Markov Chains Y and $Y^{(\varepsilon)}$ in (3.3) *starting from x at time t_i. Also:*

- *If (**IG**) holds:*

$$\chi_c(t_j - t_i, y - x) := p_c(t_j - t_i, y - x),$$

with p_c as in Theorem 3.2.1.

- *If (**IP**) holds:*

$$\chi_c(t_j - t_i, y - x) := \frac{c^d}{(t_j - t_i)^{d/2}} Q_{M-(d+5+\gamma)}\left(\frac{|y - x|}{(t_j - t_i)^{1/2}/c}\right).$$

The proof will be given in Section 3.2.5.

3.2.2 On Some Related Applications.

Model Sensitivity for Option Prices.

Assume for instance that the (log)-price of a financial asset is given by the dynamics in (3.1). Under suitable assumptions the price of an option on that asset writes at time t and when $X_t = x$ as $\mathbb{E}[f(\exp(X_T^{t,x}))]$ up to an additional discounting factor. In the previous expression f is the pay-off function. For a rather large class of pay-offs, say measurable functions with polynomial growth, including irregular ones, Theorem 3.2.1 allows to specifically quantify how a perturbation of the coefficients impacts the option prices. Precisely for a given $\varepsilon > 0$, under **(A)**:

$$\begin{aligned} |\mathcal{E}_\varepsilon(t, T, x, f)| &:= |\mathbb{E}[f(\exp(X_T^{t,x}))] - \mathbb{E}[f(\exp(X_T^{t,x,(\varepsilon)}))]| \\ &\le C\Delta_{\varepsilon,\gamma,q} \int_{\mathbb{R}^d} f(\exp(y)) p_c(T - t, x, y) dy. \end{aligned}$$

This previous control can be as well exploited to investigate perturbations of a model which provides some closed formulas, e.g. a perturbation of the Black and Scholes model that would include a stochastic volatility taking for instance $\sigma_\varepsilon(x) = \sigma + \varepsilon\psi(x)$ for some bounded Hölder continuous function ψ and ε small enough. In that case, assuming that the drift is known and unperturbed, we have $\Delta_{\varepsilon,\gamma,\infty} = |\sigma_\varepsilon - \sigma|_\gamma = \varepsilon|\psi|_\gamma$.

In connection with this application, we can quote the work of Corielli *et al.* [CFP10] who give estimates on option prices through parametrix expansions truncating the series. Some of their results, see e.g. their Theorem 3.1, can be related to a perturbation analysis since they obtain an approximation of an option price for a local volatility model in terms of the Black–Scholes price and a correction term corresponding to the first order term in the parametrix series. A more probabilistic approach to similar problems can be found in Benhamou *et al.* [BGM10]. However, none of the indicated works indeed deals with the global perturbation analysis we perform here.

Weak Error Analysis

It is well known that if the coefficients b, σ in (3.1) are smooth and a satisfies the non-degeneracy condition **(A3.2)**, then weak error on the densities for the approximation by the Euler scheme is well controlled. Precisely, for a given time step $h > 0$, let us set for $i \in \mathbb{N}, t_i := ih$. Introduce now the Euler scheme $X_0^h = x, \ \forall i \geq 1, \ X_{t_{i+1}}^h = X_{t_i}^h + b(t_i, X_{t_i}^h)h + \sigma(t_i, X_{t_i}^h)(W_{t_{i+1}} - W_{t_i})$ and denote by $p^h(t_i, x, .)$ its density at time t_i. The dynamics of the Euler scheme clearly enters the scheme (3.3). It has been established in Konakov and Mammen [KM02] (see also Bally and Talay [BT96b] for an extension to the hypoelliptic setting) that:

$$|p - p^h|(t_i, t_j, x, y) \leq Chp_c(t_j - t_i, x, y).$$

If the coefficients in (3.1) are not smooth, it is then possible to use a mollification procedure, taking for $x \in \mathbb{R}^d$, $b_\varepsilon(t, x) := b(t, .) \star \rho_\varepsilon(x), \sigma_\varepsilon(t, x) := \sigma(t, .) \star \rho_\varepsilon(x)$ with $\rho_\varepsilon := \varepsilon^{-d}\rho(x/\varepsilon)$ and $\rho \in C^\infty(\mathbb{R}^d, \mathbb{R}^+), \int_{\mathbb{R}^d} \rho(x)dx = 1, |\mathrm{supp}(\rho)| \subset K$ for some compact set K of $\mathbb{R}^d$. For the mollifying kernel ρ_ε, one then easily checks that for γ-Hölder continuous in space coefficients b, σ there exists C s.t.

$$\sup_{t\in[0,T]} |b(t, .) - b_\varepsilon(t, .)|_\infty \leq C\varepsilon^\gamma, \ \sup_{t\in[0,T]} |\sigma(t, .) - \sigma_\varepsilon(t, .)|_\eta \leq C\varepsilon^{\gamma-\eta}, \ \eta \in (0, \gamma). \tag{3.8}$$

The important aspect is that we lose a bit with respect to the sup norm when investigating the Hölder norm. We then have by Theorems 3.2.1 and 3.2.2 and their proof, that, for γ-Hölder continuous in space coefficients b, σ and taking $p = \infty$, there exist c, C s.t. for all $0 \leq s < t \leq T, 0 \leq t_i < t_j \leq T, \ (x, y) \in (\mathbb{R}^d)^2$:

$$\begin{aligned} |(p - p_\varepsilon)(s, t, x, y)| &\leq CC_\eta\varepsilon^{\gamma-\eta}p_c(t - s, y - x), \\ |(p^h - p_\varepsilon^h)(t_i, t_j, x, y)| &\leq CC_\eta\varepsilon^{\gamma-\eta}p_c(t_j - t_i, y - x), \end{aligned}$$

where the constant C_η explodes when η tends to 0.

To investigate the global weak error $(p-p^h)(t_i, t_j, x, y) = \{(p-p_\varepsilon)+(p_\varepsilon-p_\varepsilon^h)+(p_\varepsilon^h-p^h)\}(t_i, t_j, x, y)$, it therefore remains to analyse the contribution $(p_\varepsilon - p_\varepsilon^h)(t_i, t_j, x, y)$. The results of [KM02] indeed apply but yield $|(p_\varepsilon-p_\varepsilon^h)(t_i, t_j, x, y)| \leq C_\varepsilon hp_c(t_j-t_i, y-x)$ where C_ε is explosive when ε goes to zero. The global error thus writes:

$$|(p - p^h)(t_i, t_j, x, y)| \leq C\{C_\eta\varepsilon^{\gamma-\eta} + C_\varepsilon h\}p_c(t_j - t_i, y - x),$$

and a balance is needed to derive a global error bound. This is precisely the analysis which is performed in [KM17]. In this Chapter, we extend to the densities (up to a slowly growing factor) the results previously obtained by Mikulevičius and Platen [MP91] on the weak error, i.e. they showed $|\mathbb{E}[f(X_T) - f(X_T^h)]| \leq Ch^{\gamma/2}$ provided $f \in C^{2+\gamma}(\mathbb{R}^d, \mathbb{R})$. Precisely, we obtain through a suitable analysis of the constants

C_η, C_ε, which respectively depend on behaviour of the parametrix series and of the derivatives of the heat kernel with mollified coefficients, that $|p - p^h|(t_i, t_j, x, y) \le Ch^{\gamma/2-\psi(h)}p_c(t_j - t_i, y - x)$ for a function $\psi(h)$ going to 0 as $h \to 0$. (which is induced by the previous loss of η in (3.8)). In the quoted work, we also obtain some error bounds for piecewise smooth drifts having a countable set of discontinuities. This part explicitly requires the stability result of Theorems 3.2.1, Theorems 3.2.2 for $q < +\infty$. The idea being that the difference between the piece-wise smooth drift and its smooth approximation (actually the mollification procedure is only required around the points of discontinuity), is well controlled in L^q norm, $q < +\infty$.

Extension to some Kinetic Models

The results of Theorems 3.2.1 and 3.2.2 should extend without additional difficulties to the case of degenerate diffusions of the form:

$$\begin{aligned} dX_t &= b(t, X_t, Y_t)dt + \sigma(t, X_t, Y_t)dW_t, \\ dY_t &= X_t dt, \end{aligned} \tag{3.9}$$

under the same previous assumptions on b, σ when we consider perturbations of the non-degenerate components, i.e. for a given $\varepsilon > 0$, $(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})$ where:

$$\begin{aligned} dX_t^{(\varepsilon)} &= b_\varepsilon(t, X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dt + \sigma_\varepsilon(t, X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dW_t, \\ dY_t^{(\varepsilon)} &= X_t^{(\varepsilon)}dt. \end{aligned} \tag{3.10}$$

Indeed, under **(A)**, the required parametrix expansions of the densities associated with the solutions of equation (3.9), (3.10) are mentioned in Chapter 4 (see also [KMM10]).

A posteriori controls in parameter estimation

Let us consider to illustrate this application a parametrized family of diffusions of the form:

$$dXt = b(t, Xt)dt + \sigma(\eta, t, Xt)dWt, \tag{3.11}$$

where $\eta \in \theta \subset \mathbb{R}^d$, the coefficients b, σ are smooth, bounded and the non-degeneracy condition **(A3.2)** holds. A natural practical problem consists in estimating the true parameter η from an observed discrete sample $(X_{t_i^n})_{i\in[0,n]}$ where the $(t_i)_{i\in[0,n]}$ form a partition of the observation interval, i.e. if $T = 1, 0 = t_0^n < t_1^n < \cdot < t_n^n = 1$.

Introducing the contrast:

$$U^n(\eta) := \frac{1}{n}\sum_{i=1}^{n}\left[\log(\det(a(\eta, t_{i-1}^n))) + \langle a^{-1}(\eta, t_{i-1}^n, X_{t_{i-1}^n})X_i^n, X_i^n\rangle\right],$$

$$\forall i \in [1,n], X_i^n := \frac{X_{t_i^n} - X_{t_{i-1}^n}}{\sqrt{t_i^n - t_{i-1}^n}},$$

and denoting by $\hat{\eta}_n$ the corresponding minimizer, it was shown by Genon-Catalot and Jacod [GCJ93] that under $\mathbb{P}^\eta$, $\sqrt{n}(\hat{\eta}_n - \eta)$ converges in law towards a mixed normal variable S which is, conditionally to $\mathcal{F}_1 := \sigma[(X_s)_{s\in[0,1]}]$, centred and Gaussian. For a precise expression of the covariance which depends on the whole path of $(X_t)_{t\in[0,1]}$ we refer to Theorem 3 and its proof in [GCJ93].

This means that, when n is large, conditionally to $\mathcal{F}_1$, we have on a subset $\bar{\Omega} \subset \Omega$ which has high probability, that $|\hat{\eta}_n - \eta| \le \frac{C}{\sqrt{n}}$ for a certain threshold C. Setting $\varepsilon_n = n^{-1/2}, \sigma_{\varepsilon_n}(t,x) := \sigma(\hat{\eta}_n, t, x)$ and with a slight abuse of notation $\sigma(t,x) := \sigma(\eta, t, x)$, one gets that, on $\bar{\Omega}$:

$$\begin{aligned} |\sigma(t,x) - \sigma_{\varepsilon_n}(t,x) - (\sigma(t,y) - \sigma_{\varepsilon_n}(t,y))| &\le |x-y| \wedge Cn^{-1/2} \\ \Rightarrow |\sigma - \sigma_\varepsilon|\vartheta &\le (Cn^{-1/2})^{1-\vartheta}, \vartheta \in (0,1]. \end{aligned}$$

We can then invoke our Theorem 3.2.2 to compare the densities of the diffusions with the estimated parameter and the exact one in (3.11).

3.2.3 Derivation of formal series expansion for densities

Parametrix Representation of the Density for Diffusions

In the following, for given $(s,x) \in \mathbb{R}^+ \times \mathbb{R}^d$, we use the standard Markov notation $(X_t^{s,x})_{t\ge s}$ to denote the solution of (3.1) starting from x at time s.

Assume again that $(X_t^{s,x})_{t\ge s}$ has for all $t > s$ a smooth density $p(s,t,x,.)$ (which is the case if additionally to **(A)** the coefficients are smooth see e.g. Friedman [Fri64]). We would like to estimate this density at a given point $y \in \mathbb{R}^d$. To this end, we again use the parametrix expansion with respect to the density of the frozen process with spatial variable frozen at y. For all $(s,x) \in [0,T] \times \mathbb{R}^d,\ t \ge s$ we set for this Chapter:

$$\tilde{X}_t^y = x + \int_s^t \sigma(u,y) dW_u.$$

Its density $\tilde{p}^y$ readily satisfies the Kolmogorov Backward equation:

$$\begin{cases} \partial_u \tilde{p}^y(u,t,z,y) + \tilde{L}_u^y \tilde{p}^y(u,t,z,y) = 0,\ s \le u < t, z \in \mathbb{R}^d, \\ \tilde{p}^y(u,t,.,y) \underset{u\uparrow t}{\to} \delta_y(.), \end{cases} \tag{3.12}$$

where for all $\varphi \in C_0^2(\mathbb{R}^d, \mathbb{R}), z \in \mathbb{R}^d$:

$$\tilde{L}_u^y \varphi(z) = \frac{1}{2} \mathrm{Tr}\left(\sigma\sigma^*(u,y) D_z^2 \varphi(z)\right),$$

stands for the generator of $\tilde{X}^y$ at time u.

On the other hand, since we have assumed the density of X to be smooth, it must satisfy the Kolmogorov forward equation (see e.g. Dynkin [Dyn65]). For a given starting point $x \in \mathbb{R}^d$ at time s,

$$\begin{cases} \partial_u p(s,u,x,z) = L_u^* p(s,u,x,z) = 0, \ \ s < u \le t, z \in \mathbb{R}^d, \\ p(s,u,x,.) \underset{u \downarrow s}{\rightarrow} \delta_x(.), \end{cases} \tag{3.13}$$

where L_u^* stands for the *formal* adjoint (which is again well defined if the coefficients in (3.1) are smooth) of the generator of (3.1) which for all $\varphi \in C_0^2(\mathbb{R}^d, \mathbb{R}), z \in \mathbb{R}^d$ writes:

$$L_u \varphi(z) = \frac{1}{2} \mathrm{Tr}\left(\sigma\sigma^*(u,z) D_z^2 \varphi(z)\right) + \langle b(u,z), D_z \varphi(z) \rangle.$$

Equations (3.12), (3.13) yield the formal expansion below which is initially due to McKean and Singer [MS67].

$$\begin{aligned} (p - \tilde{p}^y)(s,t,x,y) &= \int_s^t du \partial_u \int_{\mathbb{R}^d} dz p(s,u,x,z) \tilde{p}^y(u,t,z,y) \\ &= \int_s^t du \int_{\mathbb{R}^d} dz \left(\partial_u p(s,u,x,z) \tilde{p}^y(u,t,z,y) + p(s,u,x,z) \partial_u \tilde{p}^y(u,t,z,y) \right) \\ &= \int_s^t du \int_{\mathbb{R}^d} dz \Big(L_u^* p(s,u,x,z) \tilde{p}^y(u,t,z,y) - p(s,u,x,z) \tilde{L}_u^y \tilde{p}^y(u,t,z,y) \Big) \\ &= \int_s^t du \int_{\mathbb{R}^d} dz p(s,u,x,z) (L_u - \tilde{L}_u^y) \tilde{p}^y(u,t,z,y), \end{aligned} \tag{3.14}$$

using the Dirac convergence for the first equality, equations (3.13) and (3.12) for the second one. We eventually take the adjoint for the last equality. Note carefully that the differentiation under the integral is also here formal since we would need to justify that it can actually be performed using some growth properties of the density and its derivatives which we *a priori* do not know.

Let us remind the notation from the Chapter 2:

$$f \otimes g(s,t,x,y) = \int_s^t du \int_{\mathbb{R}^d} dz f(s,u,x,z) g(u,t,z,y)$$

for the time-space convolution and let us define $\tilde{p}(s,t,x,y) := \tilde{p}^y(s,t,x,y)$, that is in $\tilde{p}(s,t,x,y)$ we consider the density of the frozen process at the final point and observe it at *that specific* point. We also remind the notationn of the *parametrix* kernel:

$$H(s,t,x,y) := (L_s - \tilde{L}_s) \tilde{p}(s,t,x,y) := (L_s - \tilde{L}_s^y) \tilde{p}^y(s,t,x,y). \tag{3.15}$$

This yields to iterated convolutions of the kernel and leads to the formal expansion:

$$p(s,t,x,y) = \sum_{r=0}^{\infty} \tilde{p} \otimes H^{(r)}(s,t,x,y), \tag{3.16}$$

where $\tilde{p} \otimes H^{(0)} = \tilde{p}, H^{(r)} = H \otimes H^{(r-1)},\ r \geq 1$.

And now we come to the main difference from the prove which we have in Chapter 2. Since we are working under the Assumption (**A3.3**) of the just Hölder continuity of coefficients, it is not possible to refer directly to previously obtained results for the density existence or parametrix expansion. However, both facts are still true.

Proposition 3.2.3. *Under the sole assumption* ***(A)****, for $t > s$, the density of $X_t^{x,s}$ solving* (3.1) *exists and can be written as in* (3.16)*.*

Proof. The proof can already be derived from a sensitivity argument. We first introduce two parametrix series of the form (3.16). Namely,

$$p(s,t,x,y) := \widetilde{p}(s,t,x,y) + \sum_{r=1}^{\infty} \tilde{p} \otimes H^{(r)}(s,t,x,y) \tag{3.17}$$

and

$$p_\varepsilon(s,t,x,y) := \widetilde{p}_\varepsilon(s,t,x,y) + \sum_{r=1}^{\infty} \tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)}(s,t,x,y). \tag{3.18}$$

Let us point out that, at this stage, p and p_ε are defined as sum of series. The purpose is then to identify those sums with the densities of the processes $X_t^{s,x}, X_t^{(\varepsilon),s,x}$ at point y.

The convergence of the series (3.17) and (3.18) is in some sense *standard* (see e.g. [Men11] or Friedman [Fri64]) under **(A)**. We recall for the sake of completeness the key steps for (3.17).

From direct computations, there exist $c_1 \geq 1, c \in (0,1]$ s.t. for all $T > 0$ and all multi-index $\alpha, |\alpha| \leq 8$,

$$\forall 0 \leq u < t \leq T,\ (z,y) \in (\mathbb{R}^d)^2,\ |D_z^\alpha \widetilde{p}(u,t,z,y)| \leq \frac{c_1}{(t-u)^{|\alpha|/2}} p_c(t-u, y-z), \tag{3.19}$$

where

$$p_c(t-u, y-z) = \frac{c^{d/2}}{(2\pi(t-u))^{d/2}} \exp\left(-\frac{c}{2}\frac{|y-z|^2}{t-u}\right),$$

stands for the usual Gaussian density in $\mathbb{R}^d$ with 0 mean and covariance $(t-u)c^{-1}I_d$. From (3.19), the boundedness of the drift and the Hölder continuity in space of the diffusion matrix we readily get that there exists $c_1 \geq 1,\ c \in (0,1]$,

$$|H(u,t,z,y)| \leq \frac{c_1(1 \vee T^{(1-\gamma)/2})}{(t-u)^{1-\gamma/2}} p_c(t-u, z-y). \tag{3.20}$$

Now the key point is that the control (3.20) yields an integrable singularity giving a smoothing effect in time once integrated in space in the time-space convolutions appearing in (3.17) and (3.18). It follows by induction that:

$$
\begin{aligned}
|\widetilde{p} \otimes H^{(r)}(s,t,x,y)| &\le ((1 \vee T^{(1-\gamma)/2})c_1)^{r+1} \prod_{i=1}^{r} B(\frac{\gamma}{2}, 1 + (i-1)\frac{\gamma}{2}) p_c(t-s, y-x)(t-s)^{\frac{r\gamma}{2}} \\
&= \frac{((1 \vee T^{(1-\gamma)/2})c_1)^{r+1} \left[\Gamma(\frac{\gamma}{2})\right]^r}{\Gamma(1 + r\frac{\gamma}{2})} p_c(t-s, y-x)(t-s)^{\frac{r\gamma}{2}}.
\end{aligned}
\tag{3.21}
$$

These bounds readily yield the convergence of the series as well as a Gaussian upper-bound. Namely

$$
p(s,t,x,y) \le c_1 \exp((1 \vee T^{(1-\gamma)/2})c_1[(t-s)^{\gamma/2}]) p_c(t-s, y-x). \tag{3.22}
$$

An important application of the *stability* of the perturbation consists in considering coefficients $b_\varepsilon := b \star \zeta_\varepsilon, \sigma := \sigma \star \zeta_\varepsilon$ in (3.18), where ζ_ε is a mollifier in time and space. For mollified coefficients which satisfy the non-degeneracy assumptions **(A3.2)**, the existence and smoothness of the density p_ε for the associated process $X^{(\varepsilon)}$ in (3.2) can be derived from [IKO62]. Observe carefully that the previous Gaussian bounds also hold for p_ε uniformly in ε and independently of the mollifying procedure. This therefore gives that

$$
p_\varepsilon(s,t,x,y) \underset{\varepsilon \to 0}{\longrightarrow} p(s,t,x,y), \tag{3.23}
$$

boundedly and uniformly. Thus, for every continuous bounded function f we derive from the bounded convergence theorem and (3.22) that for all $0 \le s < t,\ x \in \mathbb{R}^d$:

$$
\mathbb{E}_{s,x}[f(X_t^{(\varepsilon)})] = \int_{\mathbb{R}^d} f(y) p_\varepsilon(s,t,x,y) dy \underset{\varepsilon \to 0}{\longrightarrow} \int_{\mathbb{R}^d} f(y) p(s,t,x,y) dy. \tag{3.24}
$$

In particular, taking $f = 1$ gives that $\int_{\mathbb{R}^d} p(s,t,x,y)dy = 1$ and the uniform convergence in (3.23) gives that $p(s,t,x,\dots)$ is non negative. We therefore derive that $p(s,t,x,\cdot)$ is a probability density on $\mathbb{R}^d$.

On the other hand, under **(A)**, we can derive from Theorem 11.3.4 of [SV79] that $(X_s^\varepsilon)_{s\in[0,T]} \underset{\varepsilon \to 0}{\Rightarrow}^{\text{law}} (X_s)_{s\in[0,T]}$. This gives that for any bounded continuous function f:

$$
\mathbb{E}[f(X_t^{(\varepsilon),s,x})] \underset{\varepsilon \to 0}{\longrightarrow} \mathbb{E}[f(X_t^{s,x})].
$$

This convergence and (3.24) then yield that the random variable $X_t^{s,x}$ admits $p(s,t,x,\cdot)$ as a density.

We can thus now conclude that the processes $X, X^{(\varepsilon)}$ in (3.1), (3.2) have transition densities given by the sum of the series (3.17),(3.17). $\square$

Parametrix for Markov Chains

One of the main advantages of the formal expansion in (3.16) is that it has a direct discrete counterpart in the Markov chain setting. Indeed, denote by $(Y_{t_j}^{t_i,x})_{j\geq i}$ the Markov chain with dynamics (3.3) starting from x at time t_i. Observe first that if the innovations $(\xi_k)_{k\geq 1}$ have a density then so does the chain at time t_k.

Let us now introduce its generator at time t_i, i.e. for all $\varphi \in C_0^2(\mathbb{R}^d,\mathbb{R}),\ x\in\mathbb{R}^d$:

$$L_{t_i}^h\varphi(x) := h^{-1}\mathbb{E}[\varphi(Y_{t_{i+1}}^{t_i,x}) - \varphi(x)].$$

In order to give a representation of the density of $p^h(t_i,t_j,x,y)$ of $Y_{t_j}^{t_i,x}$ at point y for $j > i$, we introduce similarly to the continuous case, the Markov chain (or inhomogeneous random walk) with coefficients frozen in space at y. For given $(t_i,x)\in [0,T]\times\mathbb{R}^d,\ t_j\geq t_i$ we set:

$$\tilde{Y}_{t_j}^{t_i,x,y} := x + h^{1/2}\sum_{k=i}^{j-1}\sigma(t_k,y)\xi_{k+1},$$

and denote its density $\tilde{p}^{h,y}(t_i,t_j,x,.)$. Its generator at time t_i writes for all $\varphi \in C_0^2(\mathbb{R}^d,\mathbb{R}), x\in\mathbb{R}^d$:

$$\tilde{L}_{t_i}^{h,y}\varphi(x) = h^{-1}\mathbb{E}[\varphi(\tilde{Y}_{t_{i+1}}^{t_i,x,y}) - \varphi(x)].$$

Using the notation $\tilde{p}^h(t_i,t_j,x,y) := \tilde{p}^{h,y}(t_i,t_j,x,y)$, we introduce now for $0\leq i < j\leq N$ the *parametrix* kernel:

$$H^h(t_i,t_j,x,y) := (L_{t_i}^h - \tilde{L}_{t_i}^{h,y})\tilde{p}^h(t_i+h,t_j,x,y).$$

Analogously to Lemma 3.6 in [KM00], which follows from a direct algebraic manipulation, we derive the following representation for the density which can be viewed as the Markov chain analogue of Proposition 3.2.3.

Proposition 3.2.4 (Parametrix Expansion for the Markov Chain)**.** *Assume* ***(A)*** *is in force. Then, for* $0\leq t_i < t_j \leq T$,

$$p^h(t_i,t_j,x,y) = \sum_{r=0}^{j-i}\tilde{p}^h\otimes_h H^{h,(r)}(t_i,t_j,x,y),$$

where the discrete time convolution type operator $\otimes_h$ *is defined by*

$$f\otimes_h g(t_i,t_j,x,y) = \sum_{k=0}^{j-i-1} h\int_{\mathbb{R}^d} f(t_i,t_{i+k},x,z)g(t_{i+k},t_j,z,y)dz.$$

Also $g\otimes_h H^{h,(0)} = g$ *and for all* $r\geq 1, H^{h,(r)} := H^h\otimes_h H^{h,(r-1)}$ *denotes the* r*-fold discrete convolution of the kernel* H^h.

3.2.4 Stability of Parametrix Series.

We will now investigate more specifically the sensitivity of the density w.r.t. the coefficients through the difference of the series. For a given fixed parameter ε, under **(A)** the densities $p(s,t,x,.), p_\varepsilon(s,t,x,\cdot)$ at time t of the processes in (3.1), (3.2) starting from x at time s both admit a parametrix expansion of the previous type.

Stability for Diffusions

Let us consider the difference between two parametrix expansions:

$$|p(s,t,x,y)-p_\varepsilon(s,t,x,y)| = |\sum_{r=0}^{\infty}\tilde p\otimes H^{(r)}(s,t,x,y)-\sum_{r=0}^{\infty}\tilde p_\varepsilon\otimes H_\varepsilon^{(r)}(s,t,x,y)|$$

$$\leq |(\tilde p-\tilde p_\varepsilon)(s,t,x,y)|+|\sum_{r=1}^{\infty}\tilde p\otimes H^{(r)}(s,t,x,y)-\sum_{r=1}^{\infty}\tilde p_\varepsilon\otimes H_\varepsilon^{(r)}(s,t,x,y)|. \quad (3.25)$$

The strategy to study the above difference, using some well known properties of the Gaussian kernels and their derivatives recalled in (3.19), consists in first studying the difference of the *main* terms.

We have the following Lemma.

Lemma 3.2.5 (Difference of the first terms and their derivatives)**.** *Under* ***(A)****, there exist* $c_1 \geq 1,\ c \in (0,1]$ *s.t. for all* $0 \leq s < t,\ (x,y) \in (\mathbb{R}^d)^2$ *and all multi-index* $\alpha,\ |\alpha| \leq 4$*,*

$$|D_x^\alpha \tilde p(s,t,x,y)-D_x^\alpha \tilde p_\varepsilon(s,t,x,y)| \leq \frac{c_1}{(t-s)^{|\alpha|/2}}\Delta_{\varepsilon,\sigma,\gamma}p_c(t-s,y-x).$$

Proof. Let us first consider $|\alpha|=0$ and introduce some notations. Set:

$$\Sigma(s,t,y) := \int_s^t a(u,y)du,\ \Sigma_\varepsilon(s,t,y) := \int_s^t a_\varepsilon(u,y)du. \quad (3.26)$$

Let us now identify the columns of the matrices $\Sigma(s,t,y), \Sigma_\varepsilon(s,t,y)$ with d-dimensional column vectors, i.e. for $\Sigma(s,t,y)$:

$$\Sigma(s,t,y) = \left(\ \Sigma^1 \mid \Sigma^2 \mid \cdots \mid \Sigma^d\ \right)(s,t,y).$$

We now rewrite:

$$\tilde p(s,t,x,y) = f_{x,y}(\Theta(s,t,y)),\ \Theta(s,t,y) = ((\Sigma^1)^*,\cdots,(\Sigma^d)^*)^*(s,t,y),$$
$$\tilde p_\varepsilon(s,t,x,y) = f_{x,y}(\Theta_\varepsilon(s,t,y)),\ \Theta_\varepsilon(s,t,y) = ((\Sigma_\varepsilon^1)^*,\cdots,(\Sigma_\varepsilon^d)^*)^*(s,t,y),$$

with

$$f_{x,y} : \mathbb{R}^{d^2} \to \mathbb{R}$$
$$\Gamma \mapsto f_{x,y}(\Gamma) = \frac{1}{(2\pi)^{d/2}\det(\Gamma^{1:d})^{1/2}} \exp\left(-\frac{1}{2}\langle(\Gamma^{1:d})^{-1}(y-x), y-x\rangle\right), \tag{3.27}$$

where $\Gamma := \begin{pmatrix} \Gamma^1 \\ \Gamma^2 \\ \vdots \\ \Gamma^d \end{pmatrix}$ and each $(\Gamma^i)_{i\in[\![1,d]\!]}$ belongs to $\mathbb{R}^d$. Also, we have denoted:

$$\Gamma^{1:d} := \left(\ \Gamma^1 \mid \Gamma^2 \mid \cdots \mid \Gamma^d\ \right),$$

the $d \times d$ matrix formed with the entries $(\Gamma_i)_{i\in[\![1,d]\!]}$, each entry being viewed as a column.

The multidimensional Taylor expansion now gives:

$$\begin{aligned}|(\tilde p - \tilde p_\varepsilon)(s,t,x,y)| &= |f_{x,y}(\Theta(s,t,y)) - f_{x,y}(\Theta_\varepsilon(s,t,y))| \\ &= \Bigg| \sum_{|\nu|=1} D_\nu f_{x,y}(\Theta(s,t,y))\{(\Theta_\varepsilon - \Theta)(s,t,y)\}^\nu \\ &+2\sum_{|\nu|=2} \frac{\{(\Theta_\varepsilon-\Theta)(s,t,y)\}^\nu}{\nu!}\int_0^1 (1-\delta)D^\nu f_{x,y}([\Theta+\delta(\Theta_\varepsilon-\Theta)](s,t,y))d\delta\Bigg|,\end{aligned} \tag{3.28}$$

where for a multi-index $\nu := (\nu_1, \cdots, \nu_{d^2}) \in \mathbb{N}^{d^2}$, we denote by $|\nu| := \sum_{i=1}^{d^2} \nu_i$ the length of the multi-index, $\nu! = \prod_{i=1}^{d^2} \nu_i!$ and for $h \in \mathbb{R}^{d^2}$, $h^\nu := \prod_{i=1}^{d^2} h_i^{\nu_i}$ (with the convention that $0^0 = 1$). With these notations, from (3.26), (3.27), (3.28) and Assumption **(A4)** we get:

$$\begin{aligned}|f_{x,y}(\Theta(s,t,y)) - f_{x,y}(\Theta_\varepsilon(s,t,y))| \le c\Bigg\{ &\sum_{|\nu|=1} |D^\nu f_{x,y}(\Theta(s,t,y))|\Delta_{\varepsilon,\sigma,\gamma}(t-s) \\ &+ \Delta^2_{\varepsilon,\sigma,\gamma}(t-s)^2 \max_{\delta\in[0,1]} \sum_{|\nu|=2} |D^\nu f_{x,y}([\Theta+\delta(\Theta_\varepsilon-\Theta)](s,t,y))| \Bigg\}.\end{aligned} \tag{3.29}$$

Since $f_{x,y}$ in (3.27) is a Gaussian density in the parameters x, y, we recall from Cramer and Leadbetter [CM04] (see eq. (2.10.3) therein), that for all $\Gamma \in \mathbb{R}^{d^2}$ and any multi index ν, $|\nu| \le 2$:

$$D^\nu f_{x,y}(\Gamma) = \frac{1}{2^{|\nu|}}\left(\prod_{i=1}^{d^2}\left(\frac{\partial^2}{\partial_{x_{\lfloor\frac{i-1}{d}\rfloor+1}}\partial_{x_{i-\lfloor\frac{i-1}{d}\rfloor d}}}\right)^{\nu_i} f_{x,y}(\Gamma)\right),$$

where $\lfloor \cdot \rfloor$ stands for the integer part. Hence, taking from (3.29), for all $\delta \in [0,1]$, $\Gamma_{\varepsilon,\delta}(s,t,y) := [\Theta + \delta(\Theta_\varepsilon - \Theta)](s,t,y)$ yields, thanks to the non-degeneracy conditions (see equation (3.19)):

$$|D^\nu f_{x,y}(\Gamma_{\varepsilon,\delta}(s,t,y))| \le \frac{\bar{c}_1}{(t-s)^{|\nu|}} f_{x,y}\big(\bar{c}\Gamma_{\varepsilon,\delta}(s,t,y)\big) \le \frac{\bar{c}_1}{(t-s)^{|\nu|}} p_{\bar{c}}(t-s, y-x), \quad (3.30)$$

for some $\bar{c}_1 \ge 1, \bar{c} \in (0,1]$.

Thus, from (3.27), (3.28), equations (3.29) and (3.30) give:

$$|\tilde{p}(s,t,x,y) - \tilde{p}_\varepsilon(s,t,x,y)| \le \bar{c}_1 \Delta_{\varepsilon,\sigma,\gamma} p_{\bar{c}}(s,t,x,y).$$

Up to a modification of $\bar{c}_1, \bar{c}$ or c_1, c in (3.19) we can assume that the statement of the lemma and (3.19) hold with the same constants c_1, c. The bounds for the derivatives are established similarly using the controls of (3.19). This concludes the proof. □

Remark 3.2.2. Observe from equation (3.28) that the previous Lemma still holds with $\Delta_{\varepsilon,\sigma,\gamma}$ replaced by $\Delta_{\varepsilon,\sigma,\infty} := \sup_{t\in[0,T]} |\sigma(t,.) - \sigma_\varepsilon(t,.)|_\infty$. The Hölder norm is required to control the differences of the parametrix kernels.

The previous lemma quantifies how close are the main parts of the expansions. To proceed we need to consider the difference between the one-step convolutions. Combining the estimates of Lemmas 3.2.5 and 3.2.6 below will yield by induction the result stated in Theorem 3.2.1.

Lemma 3.2.6 (Control of the one-step convolution)**.** *For all* $0 \le s < t \le T,\ (x,y) \in (\mathbb{R}^d)^2$ *and for* $q \in (d, +\infty]$*:*

$$\begin{aligned} &|\tilde{p} \otimes H^{(1)}(s,t,x,y) - \tilde{p}_\varepsilon \otimes H^{(1)}_\varepsilon(s,t,x,y)| \\ &\le c_1^2 p_c(s,t,x,y)\Big\{2(1 \vee T^{(1-\gamma)/2})^2[\Delta_{\varepsilon,\sigma,\gamma} + \mathbb{I}_{q=+\infty}\Delta_{\varepsilon,b,+\infty}]B(1,\frac{\gamma}{2})(t-s)^{\frac{\gamma}{2}} \\ &\quad + \mathbb{I}_{q\in(d,+\infty)}\Delta_{\varepsilon,b,q}B(\frac{1}{2}+\alpha(q),\alpha(q))(t-s)^{\alpha(q)}\Big\}, \end{aligned} \quad (3.31)$$

where c_1, c *are as in Lemma 3.2.5 and for* $q \in (d,+\infty]$ *we set* $\alpha(q) = \frac{1}{2}(1 - \frac{d}{q})$*. The above control then yields for a fixed* $q \in (d, +\infty]$*:*

$$\begin{aligned} &|\tilde{p} \otimes H^{(1)}(s,t,x,y) - \tilde{p}_\varepsilon \otimes H^{(1)}_\varepsilon(s,t,x,y)| \\ &\le 2\bar{C}^2\Delta_{\varepsilon,\gamma,q} p_c(s,t,x,y)(t-s)^{\frac{\gamma}{2}\wedge\alpha(q)}(B(1,\frac{\gamma}{2}) \vee B(\frac{1}{2}+\alpha(q),\alpha(q))),\ \bar{C} = c_1(1 \vee T^{(1-\gamma)/2}), \end{aligned} \quad (3.32)$$

which will be useful for the iteration (see Lemma 3.2.7).

Proof. Let us write:

$$\begin{aligned}&|\tilde p\otimes H^{(1)}(s,t,x,y)-\tilde p_\varepsilon\otimes H^{(1)}_\varepsilon(s,t,x,y)|\le\\ &|\tilde p-\tilde p_\varepsilon|\otimes|H|(s,t,x,y)+\tilde p_\varepsilon\otimes|H-H_\varepsilon|(s,t,x,y)=:I+II.\end{aligned}\tag{3.33}$$

From Lemma 3.2.5 and (3.19) we readily get for all $q\in(d,+\infty]$:

$$|\tilde p-\tilde p_\varepsilon|\otimes|H|(s,t,x,y)|\le((1\vee T^{(1-\gamma)/2})c_1)^2\Delta_{\varepsilon,\gamma,q}p_c(t-s,y-x)B(1,\frac{\gamma}{2})(t-s)^{\frac{\gamma}{2}}.\tag{3.34}$$

Now we will establish that for all $0\le u<t\le T,\ (z,y)\in(\mathbb{R}^d)^2$ and $q=+\infty$:

$$|(H-H_\varepsilon)(u,t,z,y)|\le\Delta_{\varepsilon,\gamma,\infty}\frac{(1\vee T^{(1-\gamma)/2})c_1}{(t-u)^{1-\frac{\gamma}{2}}}p_c(t-u,y-z).\tag{3.35}$$

Equations (3.35) and (3.18) give that II can be handled as I which yields the result for $q=+\infty$. It therefore remains to prove (3.35). Let us write with the notations of (3.27):

$$\begin{aligned}(H-H_\varepsilon)(u,t,z,y):=&\Big[\frac12\mathrm{Tr}\Big((a(u,z)-a(u,y))D_z^2f_{z,y}\big(\Theta(u,t,y)\big)\Big)+\langle b(u,z),D_zf_{z,y}\big(\Theta(u,t,y)\big)\rangle\Big]\\ &-\Big[\frac12\mathrm{Tr}\Big((a_\varepsilon(u,z)-a_\varepsilon(u,y))D_z^2f_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\Big)+\langle b_\varepsilon(u,z),D_zf_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\rangle\Big].\end{aligned}$$

Thus,

$$\begin{aligned}(H-H_\varepsilon)(u,t,z,y)\ &=\ \frac12\Big[\mathrm{Tr}\Big((a(u,z)-a(u,y))\{D_z^2f_{z,y}\big(\Theta(u,t,y)\big)-D_z^2f_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\}\Big)\\ &-\ \mathrm{Tr}\Big([(a_\varepsilon(u,z)-a_\varepsilon(u,y)-(a(u,z)-a(u,y))]D_z^2f_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\Big)\Big]\\ &+\ \Big[\langle b(u,z),\{D_zf_{z,y}\big(\Theta(u,t,y)\big)-D_zf_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\}\rangle\\ &-\ \langle[(b_\varepsilon(u,z)-b(u,z))],D_zf_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\rangle\Big].\end{aligned}\tag{3.36}$$

Observe now that, similarly to (3.30) one has for all $i\in\{1,2\}$:

$$\begin{aligned}|D_z^if_{z,y}\big(\Theta(u,t,y)\big)|+|D_z^if_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)|&\le\frac{\tilde c_1}{(t-u)^{i/2}}p_{\tilde c}(t-u,y-z),\\ |D_z^if_{z,y}\big(\Theta(u,t,y)\big)-D_z^if_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)|&\le\frac{\tilde c_1\Delta_{\varepsilon,\sigma,\gamma}}{(t-u)^{i/2}}p_{\tilde c}(t-u,y-z).\end{aligned}$$

Also,

$$\begin{aligned}|(a_\varepsilon(u,z)-a_\varepsilon(u,y)-(a(u,z)-a(u,y))|&\le c\Delta_{\varepsilon,\sigma,\gamma}|z-y|^\gamma,\\ |b(u,z)-b_\varepsilon(u,z)|&\le c\Delta_{\varepsilon,b,\infty}.\end{aligned}$$

Thus, provided that c_1, c have been chosen large and small enough respectively in Lemma 3.2.5, the definition in (3.4) gives:

$$|(H - H_\varepsilon)(u,t,z,y)| \le \frac{(1 \vee T^{(1-\gamma)/2})c_1\Delta_{\varepsilon,\gamma,\infty}}{(t-u)^{1-\gamma/2}} p_c(t-u, y-z).$$

This establishes (3.35) for $q = +\infty$. For $q \in (d, +\infty)$ we have to use Hölder's inequality in the time-space convolution involving the difference of the drifts (last term in (3.36)). Set:

$$D(s,t,x,y) := \int_s^t du \int_{\mathbb{R}^d} \tilde{p}_\varepsilon(s,u,x,z)\langle [(b_\varepsilon(u,z) - b(u,z))], D_z f_{z,y}\big(\Theta_\varepsilon(u,t,y)\big)\rangle dz.$$

Denoting by $\bar{q}$ the conjugate of q, i.e. $q, \bar{q} > 1, q^{-1} + \bar{q}^{-1} = 1$, we get from (3.18) and for $q > d$ that:

$$\begin{aligned}
&|D(s,t,x,y)| \\
&\le c_1^2 \int_s^t \frac{\|b(u,.) - b_\varepsilon(u,.)\|_{L^q(\mathbb{R}^d)} du}{(t-u)^{1/2}} \Big\{ \int_{\mathbb{R}^d} [p_c(u-s, z-x) p_c(t-u, y-z)]^{\bar{q}} dz \Big\}^{1/\bar{q}} \\
&\le c_1^2 \Delta_{\varepsilon,b,q} \int_s^t \frac{c^d \Big\{ \int_{\mathbb{R}^d} p_{c\bar{q}}(u-s, z-x) p_{c\bar{q}}(t-u, y-z) dz \Big\}^{1/\bar{q}} du}{(2\pi)^{d(1-\frac{1}{\bar{q}})} (c\bar{q})^{d/\bar{q}} (u-s)^{\frac{d}{2}(1-\frac{1}{\bar{q}})} (t-u)^{\frac{1}{2}+\frac{d}{2}(1-\frac{1}{\bar{q}})}} \\
&\le c_1^2 \left(\frac{c(t-s)}{2\pi}\right)^{\frac{d}{2}(1-\frac{1}{\bar{q}})} \bar{q}^{-\frac{d}{2\bar{q}}} \Delta_{\varepsilon,b,q} p_c(t-s, y-x) \int_s^t \frac{du}{(u-s)^{\frac{d}{2}(1-\frac{1}{\bar{q}})} (t-u)^{\frac{1}{2}+\frac{d}{2}(1-\frac{1}{\bar{q}})}}.
\end{aligned}$$

Now, the constraint $d < q < +\infty$ precisely gives that $1 < \bar{q} < d/(d-1) \Rightarrow \frac{1}{2} + \frac{d}{2}(1 - \frac{1}{\bar{q}}) < 1$ so that the last integral is well defined. We therefore derive:

$$|D(s,t,x,y)| \le c_1^2 (t-s)^{\frac{1}{2}-\frac{d}{2}(1-\frac{1}{\bar{q}})} \Delta_{\varepsilon,b,q} p_c(t-s, y-x) B(1 - \frac{d}{2}(1-\frac{1}{\bar{q}}), \frac{1}{2} - \frac{d}{2}(1 - \frac{1}{\bar{q}})).$$

In the case $d < q < +\infty$, recalling that $\alpha(q) = \frac{1}{2}(1 - \frac{d}{q})$, we eventually get :

$$\begin{aligned}
\tilde{p}_\varepsilon \otimes |H - H_\varepsilon|(s,t,x,y) \le c_1^2 p_c(t-s, y-x)\{&\Delta_{\varepsilon,b,q}(t-s)^{\alpha(q)} B(\frac{1}{2} + \alpha(q), \alpha(q)) \\
&+ 2\Delta_{\varepsilon,\sigma,\gamma}(1 \vee T^{(1-\gamma)/2})(t-s)^{\gamma/2} B(1, \gamma/2)\}.
\end{aligned} \tag{3.37}$$

The statement now follows in whole generality from (3.33), (3.34), equations (3.35), (3.18) for $q = \infty$ and (3.37) for $d < q < +\infty$. □

The following Lemma associated with Lemmas 3.2.5 and 3.2.6 allows to complete the proof of Theorem 3.2.1.

Lemma 3.2.7 (Difference of the iterated kernels)**.** *For all* $0 \le s < t \le T,\ (x,y) \in (\mathbb{R}^d)^2$ *and for all* $q \in (d,+\infty]$, $r \in \mathbb{N}$*:*

$$|(\tilde p \otimes H^{(r)} - \tilde p_\varepsilon \otimes H_\varepsilon^{(r)})(s,t,x,y)| \le (r+1)\Delta_{\varepsilon,\gamma,q} \frac{\bar C^{r+1}\left[\Gamma(\frac{\gamma}{2}\wedge\alpha(q))\right]^r}{\Gamma(1+r(\frac{\gamma}{2}\wedge\alpha(q)))} p_c(t-s,y-x)(t-s)^{r(\frac{\gamma}{2}\wedge\alpha(q))}. \tag{3.38}$$

where c, c_1 are as in Lemma 3.2.5 and $\bar C$ as in Lemma 3.2.6.

Proof. Observe that Lemmas 3.2.5 and 3.2.6 respectively give (3.38) for $r = 0$ and $r = 1$. Let us assume that it holds for a given $r \in \mathbb{N}^*$ and let us prove it for $r+1$.

Let us denote for all $r \ge 1$, $\eta_r(s,t,x,y) := |(\tilde p \otimes H^{(r)} - \tilde p_\varepsilon \otimes H_\varepsilon^{(r)})(s,t,x,y)|$. Write

$$\begin{aligned}\eta_{r+1}(s,t,x,y) &\le |[\tilde p \otimes H^{(r)} - \tilde p_\varepsilon \otimes H_\varepsilon^{(r)}] \otimes H(s,t,x,y)| + |\tilde p_\varepsilon \otimes H_\varepsilon^{(r)} \otimes (H - H_\varepsilon)(s,t,x,y)| \\ &\le \eta_r \otimes |H|(s,t,x,y) + |\tilde p_\varepsilon \otimes H_\varepsilon^{(r)}| \otimes |(H - H_\varepsilon)|(s,t,x,y).\end{aligned}$$

Recall now that under **(A)**, the terms $|H|(s,t,x,y)$ and $|\tilde p_\varepsilon \otimes H_\varepsilon^{(r)}|$ satisfy respectively and uniformly in ε the controls of equations (3.19), (3.20). The result then follows from the proof of Lemma 3.2.6 (see equation (3.35) for $q = \infty$ and (3.37) for $q \in (d,+\infty)$) and the induction hypothesis. □

Theorem 3.2.1 now simply follows from the controls of Lemma 3.2.7, the parametrix expansions (3.17) and (3.18) of the densities p, p_ε and the asymptotic of the gamma function.

3.2.5 Stability for Markov Chains.

In this Section we prove Theorem 3.2.2. The strategy is rather similar to the one of Section 3.2.4 thanks to the series representation of the densities of the chains given in Proposition 3.2.4.

Recall first from Section 3.2.3 that we have the following representations for the density p^h, p_ε^h of the Markov chains $Y, Y^{(\varepsilon)}$ in (3.3). For all $0 \le t_i < t_j \le T,\ (x,y) \in (\mathbb{R}^d)^2$:

$$\begin{aligned}p^h(t_i,t_j,x,y) &= \sum_{r=0}^{j-i} \tilde p^h \otimes_h H^{h,(r)}(t_i,t_j,x,y), \\ p_\varepsilon^h(t_i,t_j,x,y) &= \sum_{r=0}^{j-i} \tilde p_\varepsilon^h \otimes_h H_\varepsilon^{h,(r)}(t_i,t_j,x,y).\end{aligned}$$

Comparison of the frozen densities

The first key point for the analysis with Markov chains is the following Lemma.

Lemma 3.2.8 (Controls and Comparison of the densities and their derivatives)**.** *There exist c, c_1 s.t. for all $0 \le t_i < t_j \le T$, $(x,y) \in (\mathbb{R}^d)^2$ and for all multi-index α, $|\alpha| \le 4$:*

$$|D_x^\alpha \tilde{p}^h(t_i,t_j,x,y)| + |D_x^\alpha \tilde{p}_\varepsilon^h(t_i,t_j,x,y)| \le \frac{1}{(t_j-t_i)^{|\alpha|/2}} \psi_{c,c_1}(t_j-t_i, y-x),$$
$$|D_x^\alpha \tilde{p}^h(t_i,t_j,x,y) - D_x^\alpha \tilde{p}_\varepsilon^h(t_i,t_j,x,y)| \le \frac{\Delta_{\varepsilon,\sigma,\gamma}}{(t_j-t_i)^{|\alpha|/2}} \psi_{c,c_1}(t_j-t_i, y-x),$$

where

- *Under* ***(IG)****:*
$$\psi_{c,c_1}(t_j-t_i, y-x) := c_1 p_c(t_j-t_i, y-x),$$
- *Under* ***(IP)****:*
$$\psi_{c,c_1}(t_j-t_i, y-x) := \frac{c_1 c^d}{(t_j-t_i)^{d/2}} Q_{M-d-5}\left(\frac{|y-x|}{(t_j-t_i)^{1/2}/c}\right).$$

Proof. Note first that under (**IG**) the statement has already been proved in Lemma 3.2.5. We thus assume that (**IP**) holds. Introduce first the random vectors with zero mean:

$$\tilde{Z}_{k,j}^y := \frac{1}{(t_j-t_k)^{1/2}} \sum_{l=k}^{j-1} \sigma(t_l,y)\sqrt{h}\xi_{l+1},\ \tilde{Z}_{k,j}^{y,(\varepsilon)} := \frac{1}{(t_j-t_k)^{1/2}} \sum_{l=k}^{j-1} \sigma_\varepsilon(t_l,y)\sqrt{h}\xi_{l+1}.$$

Denoting by $q_{j-k}, q_{j-k,\varepsilon}$ their respective densities, one has:

$$\begin{aligned} D_x^\alpha \tilde{p}^h(t_k,t_j,x,y) &= \frac{1}{(t_j-t_k)^{(d+|\alpha|)/2}} (-1)^{|\alpha|} D_z^\alpha q_{j-k}(z)\big|_{z=\frac{y-x}{(t_j-t_k)^{1/2}}}, \\ D_x^\alpha \tilde{p}_\varepsilon^h(t_k,t_j,x,y) &= \frac{1}{(t_j-t_k)^{(d+|\alpha|)/2}} (-1)^{|\alpha|} D_z^\alpha q_{j-k,\varepsilon}(z)\big|_{z=\frac{y-x}{(t_j-t_k)^{1/2}}}. \end{aligned} \tag{3.39}$$

From the Edgeworth expansion of Theorem 19.3 in Bhattacharya and Rao [BR76], for $q_{j-k}, q_{j-k,\varepsilon}$, one readily derives under **(A)**, for $|\alpha| = 0$ that there exists c_1 s.t. for all $0 \le t_k < t_j \le T$, $(x,y) \in (\mathbb{R}^d)^2$,

$$\tilde{p}^h(t_k,t_j,x,y) + \tilde{p}_\varepsilon^h(t_k,t_j,x,y) \le \frac{c_1}{(t_j-t_k)^{d/2}} \frac{1}{\left(1 + \frac{|x-y|}{(t_j-t_k)^{1/2}}\right)^m}, \tag{3.40}$$

for all integer $m < M - d$, where we recall that M stands for the initial decay of the density f_ξ of the innovations bounded by Q_M (see equation (3.6)).

We can as well derive similarly to the proof of Theorem 19.3 in [BR76], see also Lemma 3.7 in [KM00], that for all $\alpha, |\alpha| \le 4$:

$$|D_x^\alpha \tilde{p}^h(t_k, t_j, x, y)| + |D_x^\alpha \tilde{p}_\varepsilon^h(t_k, t_j, x, y)| \le \frac{c_1}{(t_j - t_k)^{(d+|\alpha|)/2}} \frac{1}{\left(1 + \frac{|x-y|}{(t_j - t_k)^{1/2}}\right)^m}, \quad (3.41)$$

for all $m < M - d - 4$. Note indeed that differentiating in D_x^α the density and the terms of the Edgeworth expansion corresponds to a multiplication of the Fourier transforms involved by ζ^α, ζ standing for the Fourier variable. Hence, from our smoothness assumptions in (**IP**), after obvious modifications, the estimates of Theorem 9.11 and Lemma 14.3 from [BR76] apply for these derivatives. With these bounds, one then simply has to copy the proof of Theorem 19.3. Roughly speaking, taking derivatives deteriorates the concentration of the initial control in (3.40) up to the derivation order. On the other hand, the bound in (3.40) is itself deteriorated w.r.t. the initial concentration condition in (3.6). The key point is that the techniques of Theorem 19.3 in [BR76] actually provide concentration bounds for inhomogeneous sums of random variables with concentration as in (3.6) in terms of the moments of the innovations. To explain the bound in (3.40) let us observe that the m^{th} moment of ξ is finite for $m < M - d$.

Equations (3.40) and (3.41) give the first part of the lemma. Still from the proof of Theorem 19.3 in [BR76], one gets, under **(A)**, that there exists $C > 0$ s.t. for all multi-indexes $\bar{\alpha},\ |\bar{\alpha}| \le 4, \bar{\beta},\ |\bar{\beta}| \le m \le M - d - 5$ for all $j > k$:

$$\int_{\mathbb{R}^d} |\zeta^{\bar{\alpha}}| \Big\{ |D_\zeta^{\bar{\beta}} \hat{q}_{j-k}(\zeta)| + |D_\zeta^{\bar{\beta}} \hat{q}_{j-k,\varepsilon}(\zeta)| \Big\} d\zeta \le C, \quad (3.42)$$

where $\hat{q}_{j-k}(\zeta), \hat{q}_{j-k,\varepsilon}(\zeta)$ stand for the respective characteristic functions of the random variables $\tilde{Z}_{k,j}^y$, $\tilde{Z}_{k,j}^{y,(\varepsilon)}$ at point ζ.

To investigate the quantity $|D_x^\alpha p^h(t_k, t_j, x, y) - D_x^\alpha \tilde{p}_\varepsilon^h(t_k, t_j, x, y)|$ thanks to (3.39) define now for all $\alpha,\ |\alpha| \le 4, \beta,\ |\beta| \le m \le M - d - 5$:

$$\begin{aligned} \forall z \in \mathbb{R}^d,\ \Theta_{j-k,\varepsilon}(z) &:= z^\beta D_z^\alpha \left(q_{j-k}(z) - q_{j-k,\varepsilon}(z) \right), \\ \forall \zeta \in \mathbb{R}^d,\ \widehat{\Theta}_{j-k,\varepsilon}(\zeta) &:= (-i)^{|\alpha|+|\beta|} D_\zeta^\beta \left(\zeta^\alpha \left\{ \hat{q}_{j-k}(\zeta) - \hat{q}_{j-k,\varepsilon}(\zeta) \right\} \right). \end{aligned} \quad (3.43)$$

Let us now estimate the difference between the characteristic functions. From the Leibniz formula, we are led to investigate for all multi-indexes $\bar{\beta}, \bar{\alpha},\ |\bar{\beta}| \le |\beta|, |\bar{\alpha}| \le |\alpha|$ quantities of the form:

$$\begin{aligned} &(i^{\bar{\beta}})^{-1} \zeta^{\bar{\alpha}} (D_\zeta^{\bar{\beta}} \hat{q}_{j-k}(\zeta) - D_\zeta^{\bar{\beta}} \hat{q}_{j-k,\varepsilon}(\zeta)) \\ &= \zeta^{\bar{\alpha}} \mathbb{E} \left[(\tilde{Z}_{k,j}^y)^{\bar{\beta}} \exp[i\zeta \cdot \tilde{Z}_{k,j}^y] - (\tilde{Z}_{k,j}^{y,(\varepsilon)})^{\bar{\beta}} \exp[i\zeta \cdot \tilde{Z}_{k,j}^{y,(\varepsilon)}] \right]. \end{aligned}$$

Assume first that $j > k+1$. In that case, set now $\tilde Z^y_{k,j,1} := \tilde Z^y_{k,\lceil (j+k)/2\rceil}, \tilde Z^y_{k,j,2} := \tilde Z^y_{k,j} - \tilde Z^y_{k,j,1}$. Denoting similarly $\tilde Z^{y,(\varepsilon)}_{k,j,1} := \tilde Z^{y,(\varepsilon)}_{k,\lceil (j+k)/2\rceil}, \tilde Z^{y,(\varepsilon)}_{k,j,2} := \tilde Z^{y,(\varepsilon)}_{k,j} - \tilde Z^{y,(\varepsilon)}_{k,j,1}$ for the perturbed process, we get:

$$
\begin{aligned}
&(i^{\bar\beta})^{-1}\zeta^{\bar\alpha}(D^{\bar\beta}_\zeta \hat q_{j-k}(\zeta) - D^{\bar\beta}_\zeta \hat q_{j-k,\varepsilon}(\zeta)) = \\
&\zeta^{\bar\alpha}\Bigg\{\mathbb{E}\Big[(\tilde Z^y_{k,j,1}+\tilde Z^y_{k,j,2})^{\bar\beta}\exp[i\zeta\cdot\tilde Z^y_{k,j,1}]\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big] \\
&-\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1}+\tilde Z^{y,(\varepsilon)}_{k,j,2})^{\bar\beta}\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,2}]\Big]\Bigg\} \\
&=\zeta^{\bar\alpha}\Bigg\{\sum_{l,|l|\le|\bar\beta|} C^l_{\bar\beta}\mathbb{E}\Big[(\tilde Z^y_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^y_{k,j,1}]\Big]\mathbb{E}\Big[(\tilde Z^y_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big] \\
&-\sum_{l,|l|\le|\bar\beta|} C^l_{\bar\beta}\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\Big]\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,2}]\Big]\Bigg\} \\
&=\zeta^{\bar\alpha}\Bigg\{\sum_{l,|l|\le|\bar\beta|} C^l_{\bar\beta}\Big\{\Big[\mathbb{E}\Big[(\tilde Z^y_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^y_{k,j,1}]\Big]-\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\Big]\Big]\,\mathbb{E}\Big[(\tilde Z^y_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big] \\
&+\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\Big]\Big[\mathbb{E}\Big[(\tilde Z^y_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big]-\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,2}]\Big]\Big]\Big\}\Bigg\},
\end{aligned}
$$

where in the above expression we considered the binomial expansion for multi-indexes denoting by $C^l_{\bar\beta} := \frac{\bar\beta!}{(\bar\beta-l)!l!}$ with the corresponding definitions for factorials (see the proof of Lemma 3.2.5). Introduce now, for a multi-index $l, |l| \in [\![0,|\bar\beta|]\!]$, the functions:

$$
\Psi_1^{\bar\alpha,\bar\beta-l}(\zeta) := \zeta^{\bar\alpha}\mathbb{E}\Big[(\tilde Z^y_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big],\ \Psi_2^{\bar\alpha,l}(\zeta) := \zeta^{\bar\alpha}\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\Big],
$$

and

$$
\begin{aligned}
\mathcal{E}_{1,l}(\zeta) &:= \Big[\mathbb{E}\Big[(\tilde Z^y_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^y_{k,j,1}]\Big]-\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1}]\Big]\Big],\\
\mathcal{E}_{2,\bar\beta-l}(\zeta) &:= \Big[\mathbb{E}\Big[(\tilde Z^y_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^y_{k,j,2}]\Big]-\mathbb{E}\Big[(\tilde Z^{y,(\varepsilon)}_{k,j,2})^{\bar\beta-l}\exp[i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,2}]\Big]\Big].
\end{aligned}
$$

Thus, we can rewrite from the previous computations:

$$
(i^{\bar\beta})^{-1}\zeta^{\bar\alpha}(D^{\bar\beta}_\zeta \hat q_{j-k}(\zeta) - D^{\bar\beta}_\zeta \hat q_{j-k,\varepsilon}(\zeta)) = \sum_{l,|l|\le|\bar\beta|} C^l_{\bar\beta}\Big\{(\mathcal{E}_{1,l}\Psi_1^{\bar\alpha,\bar\beta-l})(\zeta)+(\mathcal{E}_{2,\bar\beta-l}\Psi_2^{\bar\alpha,l})(\zeta)\Big\}. \tag{3.44}
$$

Recall from (3.42) that we already have integrability for the contributions $\Psi_1^{\bar\alpha,\bar\beta-l}(\zeta)$ and $\Psi_2^{\bar\alpha,l}(\zeta)$. Let us thus start with the control of $\mathcal{E}_{1,l}(\zeta), \mathcal{E}_{2,\bar\beta-l}(\zeta)$. We only give details for $\mathcal{E}_{1,l}(\zeta)$, the contribution $\mathcal{E}_{2,\bar\beta-l}$ can be handled similarly. We also consider $|l|\ge 2$, since the cases $|l|\le 2$ can be handled more directly. Write:

$$
\begin{aligned}
&|\mathcal{E}_{1,l}(\zeta)|\le\\
&\mathbb{E}[|(\tilde Z^y_{k,j,1})^l-(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l|]+\mathbb{E}[|(\tilde Z^{y,(\varepsilon)}_{k,j,1})^l||\exp(i\zeta\cdot\tilde Z^y_{k,j,1})-\exp(i\zeta\cdot\tilde Z^{y,(\varepsilon)}_{k,j,1})|]\\
&\le C\Big\{\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|(|\tilde Z^y_{k,j,1}|^{|l|-1}+|\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|-1}]+\mathbb{E}[|\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|}|\zeta||\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|]\Big\}.
\end{aligned}
$$

Apply now Hölder's inequality with $p_1=|l|,q_1=|l|/(|l|-1)$ for the first term and $p_2=(|l|+1)/|l|$, $q_2=|l|+1$ for the second one so that all the contribution appear with the same power (in order to equilibrate the constraints concerning the intregrability conditions). One gets:

$$
\begin{aligned}
&|\mathcal{E}_{1,l}(\zeta)|\le\\
&C\Big\{\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|}]^{1/|l|}\{\mathbb{E}[|\tilde Z^y_{k,j,1}|^{|l|}]^{(|l|-1)/|l|}+\mathbb{E}[|\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|}]^{(|l|-1)/|l|}\}+\\
&|\zeta|\mathbb{E}[|\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|+1}]^{|l|/(|l|+1)}\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|^{|l|+1}]^{1/(|l|+1)}\Big\}. \qquad (3.45)
\end{aligned}
$$

The point is now to prove, since we have assumed $m\le M-d-5 \iff m+1\le M-d-4$, that there exists c s.t. for all $r\le m+1$,

$$
\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|^r]^{1/r}\le c\Delta_{\varepsilon,\sigma,\gamma},\ \mathbb{E}[|\tilde Z^y_{k,j,1}|^r]^{1/r}+\mathbb{E}[|\tilde Z^{y,(\varepsilon)}_{k,j,1}|^r]^{1/r}\le c. \qquad (3.46)
$$

Let us establish the point for the difference, the other bounds can be derived similarly. Define for all $i\in[\![k,j]\!]$, $\tilde M_i:=\sqrt{h}\sum_{r=k}^{i-1}(\sigma-\sigma_\varepsilon)(t_r,y)\xi_{r+1}$. The process $(\tilde M_i)_{i\in[\![k,j]\!]}$ is a square integrable martingale (in discrete time, w.r.t. $\mathcal{F}_i:=\Sigma(\xi_r,r\le i)$, Σ-field generated by the innovation up to the current time). Its quadratic variation writes $[\tilde M]_i=h\sum_{r=k}^{i-1}|(\sigma-\sigma_\varepsilon)(t_r,y)|^2|\xi_{r+1}|^2$ and the Burkholder-Davies-Gundy inequalities, see e.g. Shiryaev [Shi96], give for all $r\le M-d-4$:

$$
\mathbb{E}[\sup_{i\in[\![k,j]\!]}|\tilde M_i|^r]\le c_r\mathbb{E}[[\tilde M]_j^{r/2}]=c_rh^{r/2}\mathbb{E}[(\sum_{i=k}^{j-1}|(\sigma-\sigma_\varepsilon)(t_i,y)|^2|\xi_{i+1}|^2)^{r/2}]. \qquad (3.47)
$$

If $r=2$ one readily gets:

$$
\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|^2]\le\frac{c_2}{(t_j-t_k)}\mathbb{E}[\sup_{i\in[\![k,j]\!]}|\tilde M_i|^2]\le\frac{c_2h}{(t_j-t_k)}\Delta^2_{\varepsilon,\sigma,\gamma}\sum_{i=k}^{j-1}\mathbb{E}[|\xi_{i+1}|^2]\le\bar c_2\Delta^2_{\varepsilon,\sigma,\gamma}.
$$

Let us thus assume $r>2$ and derive from (3.47)

$$
\begin{aligned}
\mathbb{E}[|\tilde Z^y_{k,j,1}-\tilde Z^{y,(\varepsilon)}_{k,j,1}|^r]&\le\frac{c_r}{(t_j-t_k)^{r/2}}\mathbb{E}[\sup_{i\in[\![k,j]\!]}|\tilde M_i|^r]\\
&\le\frac{c_rh^{r/2}}{(t_j-t_k)^{r/2}}\mathbb{E}[(\sum_{i=k}^{j-1}|(\sigma-\sigma_\varepsilon)(t_i,y)|^r|\xi_{i+1}|^r)(\sum_{i=k}^{j-1}1)^{r/2(1-2/r)}],
\end{aligned}
$$

applying Hölder's inequality for the counting measure with $p = r/2, q = r/(r-2)$ for the last inequality. This finally gives:

$$\mathbb{E}[|\tilde{Z}^y_{k,j,1} - \tilde{Z}^{y,(\varepsilon)}_{k,j,1}|^r] \le \frac{c_r h^{r/2}}{(t_j - t_k)^{r/2}} (j-k)^{r/2-1} \Delta^r_{\varepsilon,\sigma,\gamma} \sum_{i=k}^{j-1} \mathbb{E}[|\xi_{i+1}|^r] \le \bar{c}_r \Delta^r_{\varepsilon,\sigma,\gamma}.$$

Since we have assumed $r \le m+1 \le M-d-4$, this gives the first control in (3.46). The other one readily follows replacing $\sigma - \sigma_\varepsilon$ by σ or σ_ε.

From equations (3.45), (3.46) and similar controls for $\mathcal{E}_{2,\bar{\beta}-l}(\zeta)$ we finally derive:

$$|\mathcal{E}_{1,l}(\zeta)| + |\mathcal{E}_{2,\bar{\beta}-l}(\zeta)| \le C_1 \Delta_{\varepsilon,\sigma,\gamma}(1+|\zeta|).$$

As a result we have from (3.43) and (3.44):

$$|D^\beta_\zeta(\zeta^\alpha(\hat{q}_{j-k}(\zeta) - D^\beta_\zeta \hat{q}_{j-k,\varepsilon}(\zeta)))|$$

$$\le C\Delta_{\varepsilon,\sigma,\gamma} \left\{ \sum_{\substack{\bar{\beta},\, |\bar{\beta}| \le |\beta| \\ \bar{\alpha} = \alpha - (\beta - \bar{\beta}).}} \sum_{l, |l| \le |\bar{\beta}|} (|\Psi_1^{\bar{\alpha},\bar{\beta}-l}(\zeta)| + |\Psi_2^{\bar{\alpha},l}(\zeta)|)(1+|\zeta|) \right\}.$$

We finally derive from (3.43) and (3.42) (which thanks to the smoothness assumption on Q_M in (**IP**) holds as well for a multi-index $\bar{\alpha}, |\bar{\alpha}| = 5$):

$$|\Theta_{j-k,\varepsilon}(z)| \le \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\hat{\Theta}_{j-k,\varepsilon}(\zeta)| d\zeta \le c\Delta_{\varepsilon,\sigma,\gamma}. \tag{3.48}$$

From (3.39) this concludes the proof for $j > k+1$. If $j = k+1$ the previous arguments can be simplified and lead to the same results. □

Comparison of the parametrix kernels

This step is crucial and actually the key to the result for the Markov chains. We focus for simplicity on the case $q = +\infty$, for which pointwise controls for the differences between the drift coefficients are available, and which already emphasizes all the difficulties. The case $q \in (d, +\infty)$ for the drifts could be handled as in Lemma 3.2.6, using similar Hölder inequalities.

We actually have the following Lemma.

Lemma 3.2.9 (Control of the One-Step Convolution for the Chain.)**.** *There exists c_1, c s.t. for all $q = +\infty$ and for $0 \le t_k < t_j \le T, (z,y) \in (\mathbb{R}^d)^2$:*

$$|(H^h - H^h_\varepsilon)(t_k, t_j, z, y)| \le \frac{\Delta_{\varepsilon,\gamma,\infty}}{(t_j - t_k)^{1-\gamma/2}} \Phi_{c,c_1}(t_j - t_k, z - y),$$

with

- $\Phi_{c,c_1}(t_j-t_k,z-y)=\psi_{c,c_1}(t_j-t_k,z-y)$ *under* ***(IG)***.

- $\Phi_{c,c_1}(t_j-t_k,z-y)=\psi_{c,c_1}(t_j-t_k,z-y)\left(1+\frac{|z-y|}{(t_j-t_k)^{1/2}}\right)^\gamma$, *under* ***(IP)*** ,

where ψ_{c,c_1} *is defined according to the assumptions on the innovations in Lemma 3.2.8.*

Proof. The case $k=j+1$ involves directly differences of densities and could be treated more directly than the case $k>j+1$. We thus focus on the latter. Introduce for $k\in [\![0,N]\!], (x,w)\in(\mathbb{R}^d)^2$ the one step transitions:

$$\begin{aligned}&T^h(t_k,x,w):=b(t_k,x)h+h^{1/2}\sigma(t_k,x)w,\ T^h_\varepsilon(t_k,x,w):=b_\varepsilon(t_k,x)h+h^{1/2}\sigma_\varepsilon(t_k,x)w,\\&T^h_0(t_k,x,w):=h^{1/2}\sigma(t_k,x)w,\ T^h_{\varepsilon,0}(t_k,x,w):=h^{1/2}\sigma_\varepsilon(t_k,x)w.\end{aligned}\tag{3.49}$$

From the definition of H^h,H^h_ε, recalling that f_ξ stands for the density of the innovation, the difference of the kernels writes:

$$\begin{aligned}&(H^h-H^h_\varepsilon)(t_k,t_j,z,y)\\&=h^{-1}\int_{\mathbb{R}^d}dwf_\xi(w)\Bigg[\Big\{\tilde p^h(t_{k+1},t_j,z+T^h(t_k,z,w),y)-\tilde p^h(t_{k+1},t_j,z+T^h_0(t_k,y,w),y)\Big\}\\&\qquad-\Big\{\tilde p^h_\varepsilon(t_{k+1},t_j,z+T^h_\varepsilon(t_k,z,w),y)-\tilde p^h_\varepsilon(t_{k+1},t_j,z+T^h_{0,\varepsilon}(t_k,y,w),y)\Big\}\Bigg].\end{aligned}\tag{3.50}$$

Let us now perform a Taylor expansion at order 2 with integral rest. To this end, let us first introduce for $\lambda\in[0,1]$ the mappings:

$$\begin{aligned}\varphi^h_\lambda:\ &\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}\\&(T_1,T_2)\longmapsto \mathrm{Tr}\Big(D^2_z\tilde p^h(t_{k+1},t_j,z+\lambda T_1,y)[T_2T_2^*]\Big),\\\varphi^h_{\lambda,\varepsilon}:\ &\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}\\&(T_1,T_2)\longmapsto \mathrm{Tr}\Big(D^2_z\tilde p^h_\varepsilon(t_{k+1},t_j,z+\lambda T_1,y)[T_2T_2^*]\Big),\end{aligned}\tag{3.51}$$

where T_2 is viewed as a column vector and T_2^* denotes its transpose. Recalling as well that ξ is centered we get:

$$\begin{aligned}
\Delta H^{h,\varepsilon}(t_k,t_j,z,y) &:= (H^h - H^h_\varepsilon)(t_k,t_j,z,y)\\
= \Bigg[\Big\langle D_z\tilde p^h(t_{k+1},t_j,z,y), b(t_k,z)\Big\rangle &- \Big\langle D_z\tilde p^h_\varepsilon(t_{k+1},t_j,z,y), b_\varepsilon(t_k,z)\Big\rangle\Bigg]\\
&+h^{-1}\int_{\mathbb{R}^d} dw f_\xi(w)\int_0^1 d\lambda(1-\lambda)\\
\times\Bigg[\Big\{\varphi^h_\lambda(T^h(t_k,z,w),T^h(t_k,z,w)) &- \varphi^h_\lambda(T^h_0(t_k,y,w),T^h_0(t_k,y,w))\Big\}\\
-\Big\{\varphi^h_{\lambda,\varepsilon}(T^h_\varepsilon(t_k,z,w),T^h_\varepsilon(t_k,z,w)) &- \varphi^h_{\lambda,\varepsilon}(T^h_{0,\varepsilon}(t_k,y,w),T^h_{0,\varepsilon}(t_k,y,w))\Big\}\Bigg]\\
&=: (\Delta_1 H^{h,\varepsilon} + \Delta_2 H^{h,\varepsilon})(t_k,t_j,z,y), \qquad (3.52)
\end{aligned}$$

where for $i \in \{1,2\}$, $\Delta_i H^{h,\varepsilon}$ is associated with the terms of order i. The idea is now to make $\Delta_{\varepsilon,\gamma,\infty}$ appear explicitly. The term $\Delta_1 H^{h,\varepsilon}$ is the easiest to handle. We can indeed readily write:

$$\begin{aligned}
&\Delta_1 H^{h,\varepsilon}(t_k,t_j,z,y)\\
= \Bigg[\Big\langle D_z\tilde p^h(t_{k+1},t_j,z,y), [b(t_k,z)-b_\varepsilon(t_k,z)]\Big\rangle &- \Big\langle (D_z\tilde p^h_\varepsilon - D_z\tilde p^h)(t_{k+1},t_j,z,y), b_\varepsilon(t_k,z)\Big\rangle\Bigg].
\end{aligned}$$

From Assumption **(A3.3)**, equation (3.4) and Lemma 3.2.8 we derive for $q=+\infty$:

$$|\Delta_1 H^{h,\varepsilon}(t_k,t_j,z,y)| \le \frac{C\Delta_{\varepsilon,\gamma,\infty}}{(t_j-t_k)^{1/2}}\psi_{c,c_1}(t_j-t_k,y-z). \qquad (3.53)$$

The term $\Delta_2 H^{h,\varepsilon}$ is trickier to handle. Define to this end:

$$\begin{aligned}
\Delta\varphi^{h,\varepsilon}_\lambda(t_k,z,y,w) := &\Big\{\varphi^h_\lambda(T^h(t_k,z,w),T^h(t_k,z,w)) - \varphi^h_\lambda(T^h_0(t_k,y,w),T^h_0(t_k,y,w))\Big\}\\
&-\Big\{\varphi^h_{\lambda,\varepsilon}(T^h_\varepsilon(t_k,z,w),T^h_\varepsilon(t_k,z,w)) - \varphi^h_{\lambda,\varepsilon}(T^h_{0,\varepsilon}(t_k,y,w),T^h_{0,\varepsilon}(t_k,y,w))\Big\}.
\end{aligned}$$

Let us then decompose:

$$\begin{aligned}
\Delta\varphi_\lambda^{h,\varepsilon}(t_k,z,y,w) := \Bigg[&\Big\{\varphi_\lambda^h(T^h(t_k,z,w),T^h(t_k,z,w)) - \varphi_\lambda^h(T^h(t_k,z,w),T_0^h(t_k,y,w))\Big\}\\
&-\Big\{\varphi_{\lambda,\varepsilon}^h(T_\varepsilon^h(t_k,z,w),T_\varepsilon^h(t_k,z,w)) - \varphi_{\lambda,\varepsilon}^h(T_\varepsilon^h(t_k,z,w),T_{0,\varepsilon}^h(t_k,y,w))\Big\}\Bigg]\\
&+\Bigg[\Big\{\varphi_\lambda^h(T^h(t_k,z,w),T_0^h(t_k,y,w)) - \varphi_\lambda^h(T_0^h(t_k,y,w),T_0^h(t_k,y,w))\Big\}\\
&-\Big\{\varphi_{\lambda,\varepsilon}^h(T_\varepsilon^h(t_k,y,w),T_{0,\varepsilon}^h(t_k,y,w)) - \varphi_{\lambda,\varepsilon}^h(T_{0,\varepsilon}^h(t_k,z,w),T_{0,\varepsilon}^h(t_k,y,w))\Big\}\Bigg]\\
&=: (\Delta_1\varphi_\lambda^{h,\varepsilon} + \Delta_2\varphi_\lambda^{h,\varepsilon})(t_k,z,y,w),
\end{aligned} \tag{3.54}$$

and write from (3.52):

$$\begin{aligned}
\Delta_2 H^{h,\varepsilon}(t_k,t_j,z,y) &= h^{-1}\int_{\mathbb{R}^d} dw f_\xi(w)\int_0^1 d\lambda(1-\lambda)(\Delta_1\varphi_\lambda^{h,\varepsilon} + \Delta_2\varphi_\lambda^{h,\varepsilon})(t_k,z,y,w)\\
&=: (\Delta_{21}H^{h,\varepsilon} + \Delta_{22}H^{h,\varepsilon})(t_k,t_j,z,y),
\end{aligned} \tag{3.55}$$

for the associated contributions in $\Delta_2 H^{h,\varepsilon}$. Again, we have to consider these two terms separately.

Term $\Delta_{21}H^{h,\varepsilon}$. We first write from (3.54):

$$\begin{aligned}
&\Delta_1\varphi_\lambda^{h,\varepsilon}(t_k,z,y,w)\\
=\Bigg[&\Big\{\varphi_\lambda^h(T^h(t_k,z,w),T^h(t_k,z,w)) - \varphi_\lambda^h(T^h(t_k,z,w),T_0^h(t_k,y,w))\Big\} -\\
&\Big\{\varphi_\lambda^h(T^h(t_k,z,w),T_\varepsilon^h(t_k,z,w)) - \varphi_\lambda^h(T^h(t_k,z,w),T_{0,\varepsilon}^h(t_k,y,w))\Big\}\Bigg]\\
+\Bigg[&\Big\{\varphi_\lambda^h(T^h(t_k,z,w),T_\varepsilon^h(t_k,z,w)) - \varphi_\lambda^h(T^h(t_k,z,w),T_{0,\varepsilon}^h(t_k,y,w))\Big\}\\
-&\Big\{\varphi_\lambda^h(T_\varepsilon^h(t_k,z,w),T_\varepsilon^h(t_k,z,w)) - \varphi_\lambda^h(T_\varepsilon^h(t_k,z,w),T_{0,\varepsilon}^h(t_k,y,w))\Big\}\Bigg]
\end{aligned}$$

$$-\Bigg[\Big\{\varphi^h_{\lambda,\varepsilon}(T^h_\varepsilon(t_k,z,w),T^h_\varepsilon(t_k,z,w))-\varphi^h_{\lambda,\varepsilon}(T^h_\varepsilon(t_k,z,w),T^h_{0,\varepsilon}(t_k,y,w))\Big\}$$
$$-\Big\{\varphi^h_{\lambda}(T^h_\varepsilon(t_k,z,w),T^h_\varepsilon(t_k,z,w))-\varphi^h_{\lambda}(T^h_\varepsilon(t_k,z,w),T^h_{0,\varepsilon}(t_k,y,w))\Big\}\Bigg]$$
$$=:\sum_{i=1}^3 \Delta_{1i}\varphi^{h,\varepsilon}_\lambda(t_k,z,y,w). \qquad (3.56)$$

We now state some useful controls for the analysis. Namely, setting:

$$\begin{aligned} D(t_k,z,y,w) &:= T^h(t_k,z,w)T^h(t_k,z,w)^* - T^h_0(t_k,y,w)T^h_0(t_k,y,w)^*,\\ D_\varepsilon(t_k,z,y,w) &:= T^h_\varepsilon(t_k,z,w)T^h_\varepsilon(t_k,z,w)^* - T^h_{0,\varepsilon}(t_k,y,w)T^h_{0,\varepsilon}(t_k,y,w)^*,\end{aligned}$$

we have from **(A3.3)** and equation (3.4) for $q=+\infty$:

$$\begin{aligned} (|D|+|D_\varepsilon|)(t_k,z,y,w) &\le \bar c(h^2+h^{3/2}|w|+h(1\wedge|z-y|)^\gamma|w|^2),\\ |D-D_\varepsilon|(t_k,z,y,w) &\le \bar c\Delta_{\varepsilon,\gamma,\infty}(h^2+h^{3/2}|w|+h(1\wedge|z-y|)^\gamma|w|^2). \quad (3.57)\end{aligned}$$

From the definition of φ^h_λ in (3.51), equation (3.56), the control (3.57) and Lemma 3.2.8, we get:

$$\begin{aligned} &|\Delta_{11}\varphi^{h,\varepsilon}_\lambda|(t_k,z,y,w)\\ &\le \bar c\Delta_{\varepsilon,\gamma,\infty}\frac{\psi_{c,c_1}(t_j-t_k,y-(z+\lambda T^h(t_k,z,w)))}{(t_j-t_k)}(h^2+h^{3/2}|w|+h(1\wedge|z-y|)^\gamma|w|^2).\end{aligned} \quad (3.58)$$

We would similarly get from Lemma 3.2.8 and (3.57):

$$\begin{aligned} &|\Delta_{13}\varphi^{h,\varepsilon}_\lambda|(t_k,z,y,w)\\ \le\ & \bar c\Delta_{\varepsilon,\gamma,\infty}\frac{\psi_{c,c_1}(t_j-t_k,y-(z+\lambda T^h_\varepsilon(t_k,z,w)))}{(t_j-t_k)}(h^2+h^{3/2}|w|+h(1\wedge|z-y|)^\gamma|w|^2),\\ &|\Delta_{12}\varphi^{h,\varepsilon}_\lambda|(t_k,z,y,w)\\ \le\ & \frac{\psi_{c,c_1}(t_j-t_k,y-(z+\theta\lambda T^h(t_k,z,w)+(1-\theta)\lambda T^h_\varepsilon(t_k,z,w)))}{(t_j-t_k)^{3/2}}\\ \times\ & |(T^h-T^h_\varepsilon)(t_k,z,w)||D_\varepsilon|\\ \le\ & \bar c\Delta_{\varepsilon,\gamma,\infty}\frac{\psi_{c,c_1}(t_j-t_k,y-(z+\theta\lambda T^h(t_k,z,w)+(1-\theta)\lambda T^h_\varepsilon(t_k,z,w)))}{(t_j-t_k)^{3/2}}\\ \times\ & (h^2+h^{3/2}|w|+h(1\wedge|z-y|)^\gamma|w|^2)(h+h^{1/2}|w|), \end{aligned} \quad (3.59)$$

for some $\theta\in(0,1)$, using as well (3.49) and (3.4) for the last inequality. The point is now to get rid of the transitions appearing in the function ψ_{c,c_1}. We separate here the two assumptions at hand.

- Under (**IG**), it suffices to remark that by the convexity inequality $|z-y-\Theta|^2 \ge \frac{1}{2}|z-y|^2 - |\Theta|^2$, for all $\Theta \in \mathbb{R}^d$:

$$\psi_{c,c_1}(t_j-t_k, y-z-\Theta) \le c_1 \frac{c^{d/2}}{(2\pi(t_j-t_k))^{d/2}} \exp\left(-\frac{c}{4}\frac{|z-y|^2}{t_j-t_k}\right) \exp\left(\frac{c}{2}\frac{|\Theta|^2}{t_j-t_k}\right).$$

Now, if Θ is one of the above transitions or linear combination of transitions, we get from (3.49):

$$\psi_{c,c_1}(t_j-t_k, y-z-\Theta) \le c_1 \frac{(c/2)^{d/2}}{(2\pi(t_j-t_k))^{d/2}} \exp\left(-\frac{c}{4}\frac{|z-y|^2}{t_j-t_k}\right) \exp(\frac{c}{2}K_2^2|w|^2), \quad (3.60)$$

up to a modification of c_1 observing that $h/(t_j-t_k) \le 1$ and with K_2 as in **(A3.1)**. Since c can be chosen small enough in the previous controls, up to deteriorating the concentration properties in Lemma 3.2.8, the last term can be integrated by the standard Gaussian density f_ξ appearing in (3.55). We thus derive, from (3.60), (3.58), (3.59) and the definition in (3.56), up to modifications of c, c_1:

$$|\Delta_1 \varphi_\lambda^{h,\varepsilon}|(t_k,z,y,w) \le$$
$$\Delta_{\varepsilon,\gamma,\infty} h\bar{c}\psi_{c,c_1}(t_j-t_k, z-y)\exp(c|w|^2)\left\{1 + \frac{|w|}{(t_j-t_k)^{1/2}} + \frac{|z-y|^\gamma |w|^2}{t_j-t_k}\right\},$$

which plugged into (3.55) yields up to modifications of $\bar{c}, c, c_1$:

$$|\Delta_{21}H^{h,\varepsilon}(t_k,t_j,z,y)| \le \bar{c}\frac{\Delta_{\varepsilon,\gamma,\infty}(1\vee T^{(1-\gamma)/2})\psi_{c,c_1}(t_j-t_k,z-y)}{(t_j-t_k)^{1-\gamma/2}}. \quad (3.61)$$

- Under(**IP**), we only detail the computations for the off diagonal regime $|z-y| \ge c(t_j-t_k)^{1/2}$ which is the most delicate to handle. In this case, we have to discuss according to the position of w w.r.t. $y-z$. With the notations of **(A2)**, introduce $\mathcal{D} := \{\bar{w} \in \mathbb{R}^d : \{\Lambda h\}^{1/2}|\bar{w}| \le |z-y|/2\}$. If $w \in \mathcal{D}$, then, still from (3.58), (3.59),

$$(|\Delta_{11}\varphi_\lambda^{h,\varepsilon}| + |\Delta_{13}\varphi_\lambda^{h,\varepsilon}|)(t_k,z,y,w)$$
$$\le \bar{c}\Delta_{\varepsilon,,\gamma,\infty}\frac{\psi_{c,c_1}(t_j-t_k,y-z)}{(t_j-t_k)}(h^2 + h^{3/2}|w| + h(1\wedge|z-y|)^\gamma|w|^2),$$
$$|\Delta_{12}\varphi_\lambda^{h,\varepsilon}|(t_k,z,y,w)$$
$$\le c\Delta_{\varepsilon,\gamma,\infty}\frac{\psi_{c,c_1}(t_j-t_k,y-z)}{(t_j-t_k)^{3/2}}(h^2 + h^{3/2}|w| + h(1\wedge|z-y|)^\gamma|w|^2)(h + h^{1/2}|w|).$$

On the other hand, when $w \notin \mathcal{D}$ we use f_ξ to make the off-diagonal bound of $\psi_{c,c_1}(t_j - t_k, y-z)$ appear. Namely, we can write:

$$\begin{aligned} f_\xi(w) &\le c\frac{1}{(1+|w|)^M} \le c\frac{1}{(1+\frac{|z-y|}{h^{1/2}})^{M-(d+4)}}\frac{1}{(1+|w|)^{d+4}} \\ &\le c\frac{1}{(1+\frac{|z-y|}{(t_j-t_k)^{1/2}})^{M-(d+4)}}\frac{1}{(1+|w|)^{d+4}}, \end{aligned} \quad (3.62)$$

where the last splitting is performed in order to integrate the contribution in $|w|^3$ coming from the upper bound for $|\Delta_{12}\varphi_\lambda^{h,\varepsilon}|$ in (3.59). Plugging the above controls in (3.55) yields:

$$|\Delta H_{21}^{h,\varepsilon}(t_k,t_j,z,y)| \le \frac{\Delta_{\varepsilon,\gamma,\infty}\Phi_{c,c_1}(t_j-t_k,z-y)}{(t_j-t_k)^{1-\gamma/2}}. \tag{3.63}$$

We emphasize that in the case of innovations with polynomial decays, the control on the difference of the kernels again induces a loss of concentration of order γ in order to equilibrate the time singularity.

Term $\Delta_{22}H^{h,\varepsilon}$. This term can be handled with the same arguments as $\Delta_{21}H^{h,\varepsilon}$. For the sake of completeness we anyhow specify how the different contributions appear. Namely, with the notations of (3.54) and (3.55):

$$\begin{aligned}
&\Delta_2\varphi_\lambda^{h,\varepsilon}(t_k,z,y,w) = \\
&\int_0^1 d\mu\Big\{\Big\langle D_{T_1}\varphi_\lambda^h(T_0^h(t_k,y,w)+\mu(T^h(t_k,z,w)-T_0^h(t_k,y,w)),T_0^h(t_k,y,w)) \\
&\qquad ,T^h(t_k,z,w)-T_0^h(t_k,y,w)\Big\rangle \\
&-\Big\langle D_{T_1}\varphi_{\lambda,\varepsilon}^h(T_{0,\varepsilon}^h(t_k,y,w)+\mu(T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w)),T_{0,\varepsilon}^h(t_k,y,w)) \\
&\qquad ,T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w)\Big\rangle\Big\} \\
&=\Big\{\int_0^1 d\mu\Big\{\Big\langle D_{T_1}\varphi_\lambda^h(T_0^h(t_k,y,w)+\mu(T^h(t_k,z,w)-T_0^h(t_k,y,w)),T_0^h(t_k,y,w)), \\
&\qquad \big[(T^h(t_k,z,w)-T_0^h(t_k,y,w))-(T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w))\big]\Big\rangle\Big\} \\
&-\Big\{\int_0^1 d\mu\Big[\Big\langle D_{T_1}\varphi_{\lambda,\varepsilon}^h(T_{0,\varepsilon}^h(t_k,y,w)+\mu(T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w)),T_{0,\varepsilon}^h(t_k,y,w)) \\
&\qquad -D_{T_1}\varphi_\lambda^h(T_0^h(t_k,y,w)+\mu(T^h(t_k,z,w)-T_0^h(t_k,y,w)),T_0^h(t_k,y,w))\Big], \\
&\qquad T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w)\Big\rangle\Big\} \\
&=: (\Delta_{21}\varphi_\lambda^{h,\varepsilon}+\Delta_{22}\varphi_\lambda^{h,\varepsilon})(t_k,z,y,w).
\end{aligned}$$

In $\Delta_{21}\varphi_\lambda^{h,\varepsilon}$ we have sensitivities of order 3 for the density, giving time singularities in $(t_j-t_k)^{-3/2}$, which are again equilibrated by the the multiplicative factor:

$$\begin{aligned}
&|T_0^h(t_k,y,w)[T_0^h(t_k,y,w)]^*| \\
&\times|(T^h(t_k,z,w)-T_0^h(t_k,y,w))-(T_\varepsilon^h(t_k,z,w)-T_{0,\varepsilon}^h(t_k,y,w))| \\
&\le \bar c(h^2+h^{3/2}|w|+h|w|^2)\Delta_{\varepsilon,\gamma,\infty}(h+h^{1/2}(1\wedge|z-y|)^\gamma|w|),
\end{aligned}$$

where the last inequality is obtained similarly to (3.57) using as well (3.4). The same kind of controls can be established for $\Delta_{22}\varphi_\lambda^{h,\varepsilon}$. Anyhow, the analysis of this term leads to investigate the difference of third order derivatives, which finally yields contributions involving derivatives of order four. This is what induces the final concentration loss under (**IP**), i.e. we need to integrate a term in $|w|^4$ (see also equation (3.62) in which we performed the splitting of f_ξ on the *off-diagonal* region to integrate a contribution in $|w|^3$).

We can thus claim that

$$|\Delta H_{22}^{h,\varepsilon}(t_k,t_j,z,y)| \le \frac{\Delta_{\varepsilon,\gamma,\infty}\Phi_{c,c_1}(t_j-t_k,z-y)}{(t_j-t_k)^{1-\gamma/2}}.$$

Plugging the above control and (3.63) (or (3.61) under (**IG**) into (3.55) we derive:

$$|\Delta H_{2}^{h,\varepsilon}(t_k,t_j,z,y)| \le \frac{\Delta_{\varepsilon,\gamma,\infty}\Phi_{c,c_1}(t_j-t_k,z-y)}{(t_j-t_k)^{1-\gamma/2}},$$

which together with (3.53) and the decomposition (3.52) completes the proof. □

From Lemmas 3.2.8 and 3.2.9 the proof of Theorem 3.2.2 is achieved, under (**IG**), following the steps of Lemmas 4.3.3 and 3.2.7, using the Hölder inequalities for the differences of the drift terms for $q \in (d,+\infty)$.

The point is that we want to justify the following inequality under (**IP**) and $q = +\infty$:

$$\begin{aligned} |(\tilde p^h \otimes_h H^{h,(r)} &- \tilde p^h_\varepsilon \otimes_h H^{h,(r)}_\varepsilon)(t_i,t_j,x,y)| \\ &\le (r+1)\Delta_{\varepsilon,\gamma,\infty}\frac{\{(1\vee T^{(1-\gamma)/2})c_1\}^{r+1}\left[\Gamma(\frac{\gamma}{2})\right]^r}{\Gamma(1+r\frac{\gamma}{2})} \\ &\times \frac{c^{d/2}}{(t_j-t_i)^{d/2}}Q_{M-(d+5+\gamma)}\left(\frac{y-x}{(t_j-t_i)^{1/2}/c}\right)(t_j-t_i)^{\frac{r\gamma}{2}}. \end{aligned} \tag{3.64}$$

The only delicate point, w.r.t. the analysis performed for diffusions, consists in controlling the convolutions of the densities with polynomial decay. To this end, we can adapt a technique used by Kolokoltsov [Kol00] to investigate convolutions of "stable like" densities. Set $m := M-(d+5+\gamma)$ and denote for all $0 \le i < j \le N,\ x \in \mathbb{R}^d$ by $q_m(t_j-t_i,x) := \frac{c^d}{(t_j-t_i)^{d/2}}Q_{M-(d+5+\gamma)}\left(\frac{x}{(t_j-t_i)^{1/2}/c}\right)$ the density with polynomial decay appearing in Lemmas 3.2.8 and 3.2.9. Let us consider for fixed $i<k<j, (x,y) \in (\mathbb{R}^d)^2$ the convolution:

$$I^1_{t_k}(t_i,t_j,x,y) := \int_{\mathbb{R}^d} dz q_m(t_k-t_i,z-x)q_m(t_j-t_k,y-z). \tag{3.65}$$

- If $|x-y| \le c(t_j-t_i)^{1/2}$ (diagonal regime for the parabolic scaling), it is easily seen that one of the two densities in the integral (3.65) is homogeneous to $q_m(t_j-t_i,y-x)$.

Namely, if $(t_k - t_i) \geq (t_j - t_i)/2$, $q_m(t_k - t_i, z - x) \leq \frac{c^{d/2} c_m}{(t_k - t_i)^{d/2}} \leq \frac{(2c)^{d/2} c_m}{(t_j - t_i)^{d/2}} \leq \tilde{c} q_m(t_j - t_i, y - x)$. Thus,

$$I^1_{t_k}(t_i, t_j, x, y) \leq \tilde{c} q_m(t_j - t_i, y - x) \int_{\mathbb{R}^d} dz q_m(t_j - t_k, y - z) = \tilde{c} q_m(t_j - t_i, y - x).$$

If $(t_k - t_i) < (t_j - t_i)/2$, the same operation can be performed taking $q_m(t_j - t_k, y - z)$ out of the integral, observing again that in that case $q_m(t_j - t_k, y - z) \leq \tilde{c} q_m(t_j - t_i, y - x)$.

- If $|x - y| > c(t_j - t_i)^{1/2}$ (off-diagonal regime), we introduce $A_1 := \{z \in \mathbb{R}^d : |x - z| \geq \frac{1}{2}|x - y|\}$, $A_2 := \{z \in \mathbb{R}^d : |z - y| \geq \frac{1}{2}|x - y|\}$. Every $z \in \mathbb{R}^d$ belongs at least to one of the $\{A_i\}_{i \in \{1,2\}}$. Let us assume w.l.o.g. that $z \in A_2$. Then $|z - y| \geq \frac{c}{2}(t_j - t_i)^{1/2} \geq \frac{c}{2}(t_j - t_k)^{1/2}$ so that the density $q_m(t_j - t_k, y - z)$ is itself in the off-diagonal regime. Write:

$$\begin{aligned} \int_{A_2} dz q_m(t_k - t_i, z - x) q_m(t_j - t_k, y - z) &\leq \int_{A_2} dz q_m(t_k - t_i, z - x) \frac{c_m (t_k - t_i)^{(m-d)/2}}{|z - y|^m} \\ &\leq \frac{c_m 2^m (t_j - t_i)^{(m-d)/2}}{|x - y|^m} \int_{A_2} dz q_m(t_k - t_i, z - x) \leq \bar{c} q_m(t_j - t_i, y - x), \end{aligned}$$

recalling that, under (**IP**), $m > d$ for the last but one inequality. The same operation could be performed on A_1.

We have thus established that, there exist $\bar{c} > 1$ s.t. for all $0 \leq i < k < j, (x, y) \in (\mathbb{R}^d)^2$:

$$I^1_{t_k}(t_i, t_j, x, y) \leq \bar{c} q_m(t_j - t_i, y - x).$$

From the controls of Lemma 3.2.9 and following the strategy of Lemma 3.2.7, we will be led to consider convolutions of the previous type involving Γ functions. The above strategy thus yields (3.64) by induction.

Chapter 4

Degenerate diffusions

4.1 Introduction

4.1.1 Hypoellipticity

We would like to study the development and applications of the parametrix technique for a certain class of *degenerate* diffusions.

We will specifically focus on the Kolmogorov like diffusions (named after the seminal work of Kolmogorov [Kol34] which later on inspired Hörmander's general theory of hypoellipticity [H67̈]). Discussing the hypoellipticity concepts, we would like first to introduce the class of *hypoelliptic differential operators*.

A partial differential operator L with C^∞ coefficients in an open set $\Omega \subset \mathbb{R}^d$ is called hypoelliptic (on Ω) in case for every distribution u in Ω we have that u is a C^∞ function in every open set where Lu is a C^∞ function.

Although necessary and sufficient conditions for constant coefficients for L to be hypoelliptic have been known for quite some time before [H67̈] , see e.g. [H63̈], the technique obviously was not adapted for the general case. For instance, Kolmogorov [Kol34] constructed an example of the fundamental solution of the equation:

$$\frac{\partial^2 u}{\partial x^2} + x\frac{\partial u}{\partial y} - \frac{\partial u}{\partial t} = f, \tag{4.1}$$

which is a C^∞ function outside the diagonal. This means, the corresponding operator is hypoelliptic which was not in the framework of the existing sufficient conditions, derived before [H67̈].

Inspired by [Kol34], Hörmander in his paper [H67̈] has studied a characterization of hypoelliptic second order differential operators L with real C^∞ coefficients. Namely, if $A_0, \dots, A_n$ denote a collection of smooth vector fields on $\mathbb{R}^d$, regarded also as first order differential operators, one can define the second-order differential operator:

$$L := \frac{1}{2}\sum_{i=1}^{k} A_i^2 + A_0,\ k < d. \tag{4.2}$$

Assume that the vector fields $A_1, \ldots, A_k, [A_l, A_m]_{(l,m)\in[0,k]^2}, [A_l, [A_m, A_n]]_{(l,m,n)\in[0,k]^3}, \cdots$ where $[\cdot,\cdot]$ stands for the Lie bracketing, span $\mathbb{R}^d$. In this case, Hörmander proved that the operator L is hypoelliptic.

4.1.2 Kolmogorov's example

As we have already mentioned, in 1934 A. Kolmogorov published the paper [Kol34] in which he explicitly found the fundamental solution for the parabolic operator:

$$L_x = k\partial^2_{xx} + b\partial_x + x\partial_y,\ b \in \mathbb{R},\ k \in \mathbb{R}_+,$$

for scalar variables x, y, which precisely writes as:

$$\tilde{p}(t,(x,y),(x',y')) = \frac{\sqrt{3}}{2\pi k t^2} \times \exp\left(-\frac{|x'-x-bt|^2}{4kt} - \frac{3|y'-y-\frac{x'+x}{2}t|^2}{kt^3}\right). \tag{4.3}$$

With the modern language of stochastic calculus it is readily seen that $\tilde{p}(t,(x,y),(x',y'))$ actually corresponds to the transition density of the Gaussian process with the following dynamics:

$$\begin{cases} X_t^{s,(x,y)} = x + b(t-s) + (2k)^{1/2}(W_t - W_s), \\ Y_t^{s,(x,y)} = y + \int_s^t X_u^{s,(x,y)} du, \end{cases} \tag{4.4}$$

In Hörmander's form , with the notations of the previous paragraph $N = 2$, $L = \frac{1}{2}A_1^2 + A_0, A_1 = \begin{pmatrix} (2k)^{1/2}\partial_x \\ 0 \end{pmatrix}, A_0 = \begin{pmatrix} b\partial_x \\ x\partial_y \end{pmatrix}$ so that $[A_1, A_0] = \begin{pmatrix} 0 \\ \partial_y \end{pmatrix}$ and thus, $A_1, [A_1, A_0]$ have together rank 2.

The corresponding dynamics in (4.4) equivalently rewrites as

$$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = A_0 \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + A_1 dW_t. \tag{4.5}$$

It is clear that $A_1A_1^* = \begin{pmatrix} 2k & 0 \\ 0 & 0 \end{pmatrix}$ is a *degenerate* diffusion matrix on $\mathbb{R}^2 \otimes \mathbb{R}^2$. This is the reason why such systems as (4.4) are usually called *degenerate*. Analogously, it can be seen from (4.5) that, in the differential dynamics, the noise only acts on the first component of the system.

4.1.3 Degeneracy and Hörmander conditions

There are two main families of degenerate diffusions which are considered in modern analysis, the ones who do fulfill the *strong* Hörmander condition, namely those for which the iterated Lie brackets of the diffusive vector fields, i.e. those with indexes in $[1, k]$ in (4.2), span the whole space (like e.g. the Brownian motion on the Heisenberg group see Gaveau [Gav77]), and the ones who satisfy the so-called *weak* Hörmander condition, for which the drift vector field, A_0 in (4.2), needs to be considered in the Lie bracketing to span the whole underlying space. As emphasized above, the Kolmogorov diffusions belong to the second category. Roughly speaking "strong Hörmander" means that the noise propagates inside the system through the diffusive part only. In contrast, under the weak Hörmander conditions the drift has a key role in the noise propagation.

A striking fact, which will appear clearly, leading to specific difficulties, in the analysis, is that the weak Hörmander framework intrinsically leads to multi-scale behaviours of the underlying diffusion. This is already clearly seen in (4.3), which exhibits the two characteristic time scales of the Brownian motion and its integral (namely $t^{1/2}$ and $t^{3/2}$ for the standard deviations respectively). We also refer to the works [DM10] or [CMP15] which deal with two-sided heat kernel estimates which also reflect multi-scale behaviors in the weak Hörmander setting (noting as well that the coefficients in [DM10] are already Hölder continuous and therefore not in the previously indicated smooth case) for models generalizing the Kolmogorov example. On the other hand, we recall that, in the strong Hörmander setting we have a kind a separation between space and time. It is known from Kusuoka and Stroock [KS85] that for this family of diffusions, two sided bounds with the usual parabolic scaling in $t^{1/2}$ holds for the off-diagonal terms in the heat kernel estimates when considering the spatial distance induced by the Carnot metric associated with the vector fields.

Our research here provides a way how to adapt the parametrix expansion technique to Kolmogorov-type degenerate diffusions, even for non-smooth Hölder coefficients in the dynamics of the SDE, and to the corresponding numerical approximations through suitable Euler scheme discretizations.

We will in this chapter investigate, for various classes of test functions (namely Hölder continuous ones and Dirac masses) the so-called weak error. We refer to the global introduction and to Section 3.2.2 and Section 4.4 for details.

Although Bally and Talay [BT96a], [BT96b] already have investigated the weak error behavior for the Euler scheme approximation in a *general* hypoelliptic setting (*weak* or *strong*) for time homogeneous coefficients in the SDE, they only considered the case of smooth coefficients which lead to a usual convergence rate of order h, with h being the time discretization step of the scheme.

However, we will focus on rough coefficients. In this framework, we rely, even more than in the case of smooth coefficients, on the controls of the density of the underlying SDE. Precisely, we need to establish sharp heat-kernel and gradient bounds.

Such bounds have naturally been obtained for smooth coefficients through Malliavin calculus techniques, see [KS84], [KS85]. In the current setting of Hölder coefficients, the parametrix approach seems more adapted, since we cannot hope to handle tangent flows or Malliavin covariance matrices. To perform the analysis we will establish in the Kolmogorov case the analogue to the heat kernel and gradient estimates achieved in [IKO62] for the non-degenerate case in the Hölder setting for the coefficients.

To derive convergence rates for the weak error we also establish some stability results for the diffusion and scheme transition densities with respect to small perturbations of the coefficients. The result is of interest in itself. It is in the current context crucial in the sense that our main controls on the derivatives of the underlying heat kernel (see Theorem 4.4.2) only provide gradient bounds in the non-degenerate directions. The smoothing procedure of the coefficients (mollification) allows to directly exhibit some underlying PDE which kills the first order terms in the error analysis.

4.1.4 General models

The mentioned article [Kol34]has led to many research on the topic of so-called *degenerate Kolmogorov SDEs.* One of the most general model which allow to apply Gaussian bounds to the transition density has been studied by F. Delarue and S. Menozzi in their paper [DM10]. Precisely, it has a form of:

$$\begin{aligned} dX_t^1 &= F_1(t, X_t^1, \dots, X_t^n)dt + \sigma(t, X_t^1, \dots, X_t^n)dW_t, \\ dX_t^2 &= F_2(t, X_t^1, \dots, X_t^n)dt, \\ dX_t^3 &= F_3(t, X_t^2, \dots, X_t^n)dt, \\ \dots & \\ dX_t^n &= F_n(t, X_t^{n-1}, X_t^n)dt, \end{aligned} \tag{4.6}$$

where W_t stands for the d-dimensional Brownian motion and $\forall i, 1 \le i \le n, (X_t^i)_{t\ge 0} \in \mathbb{R}^d$.

A typical example for (4.6) is a system of n coupled oscillators, each moving vertically and being connected to the nearest neighbours directly, the first oscillator being forced by a random noise.

In the general case when $n \ge 2$, such systems appear in heat conduction models (see for example the original papers by Eckmann et al. [EPRB99] and Rey-Bellet and Thomas [RBL00] when the chain is forced by two heat baths; see also the paper by Bodineau and Lefevere [BL08]).

For the case of smooth coefficients in (4.6) the existence of a density for $(X_t^1, \dots, X_t^n)$, seen as an $\mathbb{R}^{nd}$ -valued vector, may be seen as a consequence of Hörmander's theorem. Understanding the structure of the density under general hypoelliptic conditions (i.e. for more general systems than (4.6)) is something very difficult. The reason may be explained as follows: there may be many ways for the underlying noise to propagate into the whole system, and therefore, many different time scales for the propagation phenomenon.

In the article [DM10] the authors considered uniformly Lipschitz continuous $(F_i)_{i\in[1,n]}$ (or suppose for $i = 1$ that the drift of the non degenerate component F_1 is measurable and bounded) and uniformly Hölder continuous in space uniformly elliptic diffusion matrix. Under such assumptions they established Gaussian Aronson like estimates for the density of (4.6) over compact time interval for Hölder index γ greater than 1/2. To derive this a "formal" parametrix expansion has been used, considering a sequence of equations with smooth coefficients for which Hörmander's theorem guaranteed the existence of the density, see e.g. Hörmander [H67̈].The estimates did not depend on the derivatives of the mollified coefficients but only on the γ- Hölder continuity assumed. However to pass to the limit in the described procedure some uniqueness in law is needed. It was precisely derived in [DM10] through viscosity type techniques which do not exploit the underlying smoothing effects of the parametrix kernel. This is what led to the restriction on the Hölder exponent.

Exploiting such a smoothing effect, S. Menozzi proved in [Men11] the well posedness of the martingale problem for the generator associated with the SDE (4.6) in the Hölder setting without any restriction on the Hölder index γ. Thus, together with results from [DM10] it gives that for all γ in $(0,1]$ the unique weak solution for (4.6) exists and admits for all $t > 0$ a density that satisfies Aronson like bounds.

4.1.5 Model [KMM10]

For simplicity we would like to come back to another model, considered by V. Konakov, S. Molchanov and S. Menozzi in [KMM10] and describe the parametrix derivation in details. Worth to emphasize that results mentioned in [KMM10] have been achieved under Lipschitz continuity assumptions on coefficients which partly can be relaxed to Hölder continuity according to [DM10], [Men11]. The parametrix representation achieved in [KMM10] allows to give a local limit theorem with the usual convergence rate for the associated Markov chain approximation. We anyhow mention that, the generic Markov chain approximation setting leads to additional technicalities, namely an aggregation procedure of the randomness is needed in order that the transitions at the considered aggregated time have a density. We refer to the quoted work for additional details.

Namely, in [KMM10] the following diffusions has been studied:

$$\begin{cases} X_t = x + \int_0^t b(X_s, Y_s)ds + \int_0^t \sigma(X_s, Y_s)dW_s, \\ Y_t = y + \int_0^t X_s ds, \end{cases} \tag{4.7}$$

with the generator such that: $\forall \phi \in C_0^2(\mathbb{R}^{2d}), \forall (x,y) \in \mathbb{R}^{2d}$,

$$L\phi(x,y) = \frac{1}{2}\mathrm{Tr}\Big(a(x,y)D_x^2\phi(x,y)\Big) + \langle b(x,y), \nabla_x \phi(x,y)\rangle + \langle x, \nabla_y \phi(x,y)\rangle, \quad (4.8)$$

where $(W_t)_{t\geq 0}$ is a standard $d-$ dimensional Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual assumptions.

Exactly this model can be used dealing with Asian options, X_t can be associated with the dynamics of the underlying asset and its integral Y_t is involved in the option payoff. Typically, the price of such options writes $\mathbb{E}_x[\psi(X_T, T^{-1}Y_T)]$, where for the put (resp. call) option the function $\psi(x,y) = (x-y)^+$ (resp.$(y-x)^+$), see [BPV01]. It is, thus, useful in this framework to specifically quantify how a perturbation of the coefficients impacts the option prices.

The cross dependence of the dynamics of X_t in Y_t is also important handling kinematic models or Hamiltonian systems. For a given Hamilton function of the form $H(x,y) = V(y) + \frac{|x|^2}{2}$, where V is a potential and $\frac{|x|^2}{2}$ the kinetic energy of a particle with unit mass, the associated stochastic Hamiltonian system would correspond to $b(X_s, Y_s) = -(\partial_y V(Y_s) + F(X_s, Y_s)X_s)$ in (4.7), where F is a friction term. When $F > 0$ natural questions arise concerning the asymptotic behaviour of (X_t, Y_t), for instance, the geometric convergence to equilibrium for the Langevin equation is discussed in Mattingly and Stuart [MSH02], numerical approximations of the invariant measures in Talay [Tal02], the case of high degree potential V is investigated in Hérau and Nier [HN04].

4.2 Parametrix in the degenerate case

In this section we would like to introduce the parametrix technique which is possible to perform even under Hölder continuity assumptions for coefficients applying some regularization procedure.

The unboundedness of the first order term imposes a more subtle strategy than in non-degenerate case for the choice of the frozen Gaussian density. We have to take into consideration the "geometry" of the deterministic differential equation associated to the first order terms of the operator. In other words, the corresponding flow must appear in the frozen density.

We would like to keep considering the model similar to (4.7) but with Hölder continuity assumptions for coefficients instead of Lipschitz, namely, we consider $\mathbb{R}^d \times \mathbb{R}^d-$valued processes that follow the dynamics:

$$\begin{cases} dX_t = b(X_t, Y_t)dt + \sigma(X_t, Y_t)dW_t, \\ dY_t = X_t dt, t \in [0,T], \end{cases} \quad (4.9)$$

where $b : \mathbb{R}^{2d} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^{2d} \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded coefficients that are Hölder continuous in space (this condition will be possibly relaxed for the drift term b) and W is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. In (4.9), $T > 0$ is a fixed deterministic final time. Also, $a(x,y) := \sigma\sigma^*(x,y)$ is assumed to be uniformly elliptic.

We point out that those assumptions (specified below) are actually sufficient to guarantee weak uniqueness for the solution of equation (4.9), see Remark 4.2.1.

4.2.1 Assumptions

For better readability we now repeat assumptions for this Chapter, which we have introduced in 1.

(**AD1**) **(Boundedness of the coefficients)**.

The components of the vector-valued function $b(x,y)$ and the matrix-valued function $\sigma(x,y)$ are bounded measurable. Specifically, there exists a constant K s.t.

$$\sup_{(x,y)\in\mathbb{R}^{2d}} |b(x,y)| + \sup_{(x,y)\in\mathbb{R}^{2d}} |\sigma(x,y)| \leq K.$$

(**AD2**) **(Uniform Ellipticity)**.

The matrix $a := \sigma\sigma^*$ is uniformly elliptic, i.e. there exists $\Lambda \geq 1$, $\forall (x,y,\xi) \in (\mathbb{R}^d)^3$,

$$\Lambda^{-1}|\xi|^2 \leq \langle a(x,y)\xi, \xi\rangle \leq \Lambda|\xi|^2.$$

(**AD3**) **(Hölder continuity in space)**.

For some $\gamma \in (0,1]$, $0 < \kappa < \infty$,

$$|b(x,y) - b(x',y')| + |\sigma(x,y) - \sigma(x',y')| \;\; \leq \;\; \kappa \left(|x-x'|^\gamma + |y-y'|^{\gamma/3} \right).$$

Observe that the last condition also readily gives, thanks to the boundedness of σ that the diffusion matrix a is also uniformly γ and $\gamma/3$ -Hölder continuous with respect to the variables x and y respectively.

We say that assumption **(AD)** holds when conditions **(AD1)**-**(AD3)** are in force.

Remark 4.2.1. We point out that **(AD)** actually guarantees the well posedness of the martingale problem for the generator associated with the SDE (4.9) which in turns imply weak well-posedness for (4.9). If $b = 0$ this readily follows from [Men11]. The weak well-posedness would in fact hold for any $\gamma_1, \gamma_2 \in (0,1]$, meant to be the respective Hölder continuity indexes for the variables x, y, in **(A3)** (see e.g. Theorem 2.1 therein). Similarly, the well posedness would still hold for (4.9) for any bounded measurable b. The key point in the approach of [Men11] is indeed to have a so-called smoothing effect in time of an underlying “parametrix" kernel (introduced in the current work in definition (4.18) below), which precisely holds for bounded drifts in the non-degenerate component (see again Theorem 2.1 in [Men11] and controls in Lemma (4.2.1) below).

We will denote, from now on, by C a constant depending on the parameters appearing in **(AD)** and T. We reserve the notation c for constants that only depend on **(AD)** but not on T. The values of C, c may change from line to line.

4.2.2 Parametrix expansion. Diffusion

Although we have a lot in common with the parametrix technique presented in Chapter 2, there are also some special properties which we need due to the structure of (4.9).

Namely, the first step of the parametrix for degenerate diffusions also consists in approximating the transition density $p(T,(\cdot,\cdot),(x',y'))$ by a known Gaussian density $\tilde{p}^{T,(x',y')}(T,(\cdot,\cdot),(x',y'))$. The choice obeys the following idea: in short time, $p(T,(\cdot,\cdot),(x',y'))$ and $\tilde{p}^{T,(x',y')}(T,(\cdot,\cdot),(x',y'))$ are to be close. We would like to emphasize that in our case, due to the measurability and boundedness of the function $b(\cdot,\cdot)$ it is just enough to use such a specific form of the Gaussian proxy, where the initial system does not depend on the trend function. Similarly to the non-degenerate case in Chapter 2 we have not incorporated the dependency on $b(\cdot,\cdot)$ into the frozen process directly.

For non-smooth coefficients in (4.9) but satisfying (**AD**), it is then possible to use a mollification procedure, taking $b_\eta(x,y) := b \star \rho_\eta(x,y),\ \sigma_\eta(x,y) := \sigma \star \rho_\eta(x,y),\ x,y \in \mathbb{R}^d$ where ρ_η is a smooth mollifying kernel and $\star$ stands for the usual convolution operation and $\eta \in [0,1]$,«the case $\eta = 0$ by definition will correspond to the initial process in (4.9).

For mollified coefficients, the existence and smoothness of the density p_η for the associated process (X_s^η, Y_s^η) follows from the Hörmander theorem (see e.g. [H$\ddot{6}$7]). Thus, we can apply the parametrix technique directly for p_η.

Fixing the terminal point (x',y') at time T, we finally introduce the Gaussian system of the form:

$$\begin{cases} d\tilde{X}_{\eta,t}^{T,(x',y')} = x + \sigma_\eta(x', y' - x'(T-t))dW_t, \\ d\tilde{Y}_{\eta,t}^{T,(x',y')} = y + \tilde{X}_{\eta,t}^{T,(x',y')}dt. \end{cases} \tag{4.10}$$

Since the model (4.10) defines the Gaussian process, the transition density of (4.10) $(\tilde{p}_\eta^{T,(x',y')}(t,(x,y),(\hat{x},\hat{y})))_{0<t\le T;(x,y),(\hat{x},\hat{y})\in\mathbb{R}^{2d}}$ exists.

Observe that in our settings the SDE (4.10) itself integrates as

$$\begin{pmatrix} \tilde{X}_{\eta,t}^{0,(x,y),T,(x',y')} \\ \tilde{Y}_{\eta,t}^{0,(x,y),T,(x',y')} \end{pmatrix} = R_t \begin{pmatrix} x \\ y \end{pmatrix} + \int_0^t R_u B \sigma_\eta(R_{T-u} \begin{pmatrix} x' \\ y' \end{pmatrix})dW_u, \tag{4.11}$$

where $R_t = \begin{pmatrix} I_{d\times d} & 0_{d\times d} \\ tI_{d\times d} & I_{d\times d} \end{pmatrix}$ – the resolvent matrix associated with the linear system and $B = \begin{pmatrix} I_{d\times d} \\ 0_{d\times d} \end{pmatrix}$.– the embedding matrix from R^d to $\mathbb{R}^{2d}$.

In particular, for a fixed $t > 0$ and a given starting point (x, y) in (4.11), we can write now the exact form of the transition density at time t for the frozen process:

$$\begin{aligned} &\tilde{p}_\eta^{t,(x',y')}(t,(x,y),(x',y')) \\ &\quad = \frac{1}{(2\pi)^d \det(C_t^\eta)^{1/2}} \exp\left(-\tfrac{1}{2}\langle (C_t^\eta)^{-1}(R_t \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x' \\ y' \end{pmatrix}), R_t \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x' \\ y' \end{pmatrix}\rangle\right), \end{aligned}$$

where $C_t^\eta = \int_0^t R_{t-u} B \sigma_\eta \sigma_\eta^*(x', y' - x'(t-u)) B^* R_{t-u}^* du$.

We have already introduced in (4.8) the generator for (4.9) and now it comes to the definition of the frozen process (4.11) generator $(\tilde{L}_s^{t,(x',y'),\eta})_{0\le s<t\le T}$:

$$\tilde{L}_s^{t,(x',y'),\eta}\phi(x,y) = \frac{1}{2}\mathrm{Tr}\Big(a_\eta(x', y' - x'(t-s)) D_x^2 \phi(x,y)\Big) + \langle x, \nabla_y \phi(x,y)\rangle. \quad (4.12)$$

The density $\tilde{p}_\eta$ then readily satisfies the Kolmogorov Backward equation:

$$\begin{cases} \partial_u \tilde{p}_\eta(t-u,(x,y),(x',y')) + \tilde{L}_u^{t,(x',y'),\eta} \tilde{p}_\eta(t-u,(x,y),(x',y')) = 0, \\ 0 < u < t, (x,y),(x',y') \in \mathbb{R}^{2d}, \\ \tilde{p}_\eta(t-u,(\cdot,\cdot),(x',y')) \underset{t-u\downarrow 0}{\rightarrow} \delta_{(x',y')}(.). \end{cases} \quad (4.13)$$

On the other hand, since the density of (X_s^η, Y_s^η) is smooth, it must satisfy the Kolmogorov forward equation (see e.g. Dynkin [Dyn65]). For a given starting point $(x, y) \in \mathbb{R}^{2d}$ at time 0,

$$\begin{cases} \partial_u p_\eta(u,(x,y),(x',y')) - L^* p_\eta(u,(x,y),(x',y')) = 0, \ 0 < u \le t, (x,y) \in \mathbb{R}^{2d}, \\ p_\eta(u,(x,y),.) \underset{u\downarrow 0}{\rightarrow} \delta_{(x,y)}(.), \end{cases} \quad (4.14)$$

where L^* stands for the adjoint (which is well defined since the coefficients are smooth) of the generator L in (4.8).

Let us remind for a given $c > 0$ and for all $(x, y), (x', y') \in \mathbb{R}^{2d}$ the Kolmogorov-type density, introduced in Chapter 1:

$$p_{c,K}(t,(x,y),(x',y')) := \frac{c^d 3^{d/2}}{(2\pi t^2)^d} \exp\left(-c\left[\frac{|x'-x|^2}{4t} + 3\frac{|y'-y-(x+x')t/2|^2}{t^3}\right]\right), \quad (4.15)$$

which also enjoys the semigroup property, i.e. for any $0 \le s < t \le T$,

$$\int_{\mathbb{R}^{2d}} p_{c,K}(s,(x,y),(w,z))p_{c,K}(t-s,(w,z),(x',y'))dwdz = p_{c,K}(t,(x,y),(x',y')). \quad (4.16)$$

The subscript K in the notation $p_{c,K}$ stands for Kolmogorov-like equations, and $p_{c,K}(t,(x,y),(\cdot,\cdot))$ denotes the transition density of

$$\left(X_t^{c,K}, Y_t^{c,K}\right) := \left(x + \frac{\sqrt{2}W_t}{c^{1/2}}, y + \int_0^t X_s^{c,K} ds\right).$$

Observe carefully that the density in (4.15) exhibits a *multiscale* behaviour. The non degenerate component has at time t the usual diffusive scale in $t^{1/2}$ corresponding to the self-similarity index or typical scale of the Brownian motion, whereas the degenerate one has, in small time, a "faster" typical behaviour in $t^{3/2}$ corresponding to the standard scale of the integral $\int_0^t W_s ds$. By "faster", we mean that the time normalization in the exponential deviation bounds appearing in (4.15) are bigger in small time, i.e. $t^{-3/2} \ge t^{-1/2}$ for the typical scales or standard deviations.

From direct computations on Gaussian density, it follows that for any indexes α, β, such that $|\alpha| \le 4, |\beta| \le 2$:

$$\begin{gathered}\exists C > 0,\ \forall \alpha = (\alpha_1, \alpha_2),\ |\alpha| \le 4, \\ |D_x^\alpha D_y^\beta \tilde{p}_\eta^{t,x',y'}(t,(x,y),(x',y'))| \le \tfrac{C}{t^{|\alpha|/2+3|\beta|/2}} p_{c,K}(t,(x,y),(x',y')).\end{gathered} \quad (4.17)$$

We adopt the following convention: $\tilde{p}_\eta(T-s,(x,y),(x',y'))$ stands for $\tilde{p}_\eta^{T,(x',y')}(T-s,(x,y),(x',y'))$.

The key quantity in the parametrix method is the kernel function which writes similarly as in the non-degenerate case:

$$\forall \eta \in [0,1] \quad H_\eta(t,(x,y),(x',y')) := (L^\eta - \tilde{L}^{t,(x',y'),\eta})\tilde{p}_\eta(t,(x,y),(x',y')), \quad (4.18)$$

where L^η denotes the same operator as in (4.8), but with mollified coefficients b_η and σ_η.

Note carefully that in the above kernel H_η, because of the linear structure of the degenerate component in the model, the most singular terms, i.e. those involving derivatives w.r.t. y, i.e. the *fast* variable, vanish.

Let us now remind the notation for

$$f \otimes g(t,(x,y),(x',y')) = \int_0^t du \int_{\mathbb{R}^{2d}} dzdw f(u,(x,y),(w,z))g(t-u,(w,z),(x',y'))$$

as the time-space convolution.

Using the standard mollification argument and applying forward and backward Kolmogorov equations one can derive

$$
\begin{aligned}
&(p_\eta - \tilde{p}_\eta)(t,(x,y),(x',y')) = p_\eta \otimes H_\eta(t,(x,y),(x',y')) \\
&= \int_0^t du \int_{\mathbb{R}^{2d}} p_\eta(u,(x,y),(w,z)) H_\eta(t-u,(w,z),(x',y')) dw dz,
\end{aligned}
$$

and after the iteration procedure one get the formal expansion:

$$
p_\eta(t,(x,y),(x',y')) = \sum_{r=0}^{\infty} \tilde{p}_\eta \otimes H_\eta^{(r)}(t,(x,y),(x',y')), \tag{4.19}
$$

Obtaining estimates on p_η from the formal expression (4.19) requires to have good controls on the right-hand side. Precisely thanks to (4.17), we first get that there exist $c_1 > 1,\ c > 0$ s.t. for all $u \in [0,t)$,

$$
\begin{aligned}
&|H_\eta(t-u,(w,z),(x',y'))| \\
&\le \frac{1}{2} Tr\left\{a_\eta(w,z) - a_\eta(x', y' - x'(t-u))\right\} D_w^2 \tilde{p}_\eta(t-u,(w,z),(x',y')) \\
&\quad + \langle b_\eta(w,z), D_w \tilde{p}_\eta(t-u,(w,z),(x',y'))\rangle \\
&\le \left[\frac{C|w-x'|^\gamma + |z-(y'-x'(t-u))|^{\gamma/3}}{2(t-u)} + \frac{C}{(t-u)^{1/2}}\right] \\
&\quad \times p_{c,K}(t-u,(w,z),(x',y')) \\
&\le c_1\left(1 \vee T^{(1-\gamma)/2}\right) \frac{p_{c,K}(t-u,(w,z),(x',y'))}{(t-u)^{1-\gamma/2}}.
\end{aligned} \tag{4.20}
$$

We can establish by induction the following key result.

Lemma 4.2.1. *There exist constants $C \ge 1, c > 0$ s.t. for all $\eta \in [0,1]$ one has for all $r \in \mathbb{N}^*, (t,(x,y),(x',y')) \in (0,T] \times (\mathbb{R}^{2d})^2$:*

$$
\begin{aligned}
&\left|\tilde{p}_\eta \otimes H_\eta^{(r)}(t,(x,y),(x',y'))\right| \\
&\le C^{r+1} t^{r\gamma/2} B\left(1, \frac{\gamma}{2}\right) \times B\left(1+\frac{\gamma}{2}, \frac{\gamma}{2}\right) \times \cdots \times B\left(1 + \frac{(r-1)\gamma}{2}, \frac{\gamma}{2}\right) \\
&\quad \times p_{c,K}(t,(x,y),(x',y')),
\end{aligned}
$$

recalling that $H_\eta^{(r)} := H_\eta^{(r-1)} \otimes H_\eta$.

Proof. The result (4.17) in particular yields that $\exists C_2 > 0, \forall u \in (0,t], \tilde{p}_\eta(t-u,(x,y),(w,z)) \le C_2 p_{c,K}(t-u,(x,y),(w,z))$ uniformly w.r.t. $\eta \in [0,1]$.

Setting $C := c_1 \left(1 \vee T^{(1-\gamma)/2}\right) \vee C_2$, we also obtain uniformly in η

$$
\begin{aligned}
&|\tilde{p}_\eta \otimes H_\eta(t,(x,y),(x',y'))| \\
&\le \int_0^t du \int_{\mathbb{R}^{2d}} \tilde{p}_\eta(u,(x,y),(w,z))|H_\eta(t-u,(w,z),(x',y'))|dwdz, \\
&\le \int_0^t du \int_{\mathbb{R}^{2d}} C^2 p_{c,K}(u,(x,y),(w,z)) \frac{1}{(t-u)^{1-\gamma/2}} p_{c,K}(t-u,(w,z),(x',y'))dwdz \\
&\le C^2 t^{\gamma/2} B\left(1,\frac{\gamma}{2}\right) p_{c,K}(t,(x,y),(x',y')),
\end{aligned}
$$

using the semigroup property (4.16) in the last inequality and where $B(p,q) = \int_0^1 u^{p-1}(1-u)^{q-1}du$ denotes the $\beta-$function. By induction on r:

$$
\begin{aligned}
&\left|\tilde{p}_\eta \otimes H_\eta^{(r)}(t,(x,y),(x',y'))\right| \\
&\le C^{r+1} t^{r\gamma/2} B\left(1,\frac{\gamma}{2}\right) \times B\left(1+\frac{\gamma}{2},\frac{\gamma}{2}\right) \times \cdots \times B\left(1+\frac{(r-1)\gamma}{2},\frac{\gamma}{2}\right) \\
&\times p_{c,K}(t-s,(x,y),(x',y')), \quad r \in \mathbb{N}^*,
\end{aligned}
$$

which means that the sum of the series (4.19) is uniformly controlled w.r.t. $\eta \in [0,1]$. □

These bounds imply that the series representing the density of the initial process $p_\eta(t,(x,y),(x',y'))$ could be expressed as (4.19) yield and the following bound uniformly in $\eta \in [0,1]$: $p_\eta(t,(x,y),(x',y')) \le c_1 p_{c,K}(t,(x,y),(x',y'))$.

From the bounded convergence theorem one can derive that

$$
p_\eta(t,(x,y),(x',y')) \underset{\eta\to 0}{\longrightarrow} \sum_{r=0}^{\infty} \tilde{p} \otimes H^{(r)}(t,(x,y),(x',y')) := p(t,(x,y),(x',y')), \tag{4.21}
$$

uniformly in $(t,(x,y),(x',y'))$, where $\tilde{p}(u,(x,y),(w,z)) := \tilde{p}_0(u,(x,y),(w,z))$ and $H^{(r)}(t-u,(w,z),(x',y')) := H_0^{(r)}(t-u,(w,z),(x',y'))$.

Due to the uniform convergence in η (which implies the uniqueness in law):

$$
\int_{\mathbb{R}^{2d}} f(z,w) p_\eta(t,(x,y),(w,z))dwdz \underset{\eta\to 0}{\longrightarrow} \int_{\mathbb{R}^{2d}} f(z,w) p(t,(x,y),(w,z))dwdz,
$$

for all continuous and bounded f. The well-posedness of the martingale problem and the same technique as in Theorem 11.4.2 from [SV79] then give that the process (X_t,Y_t) has the transition density which is exactly the sum of the parametrix series $p(t,(x,y),(x',y'))$.

Remark 4.2.2. Although our model (4.9) does not totally fulfill the framework of Theorem 11.4.2 from [SV79] one can derive the same result following every step of the proof. Namely, the only Theorem which is used in the proof by Stroock and Varadhan and not so clear in our framework is Th. 9.2.12 in [SV79]. Namely, we need to derive that in our case that $\forall \eta \in [0,1]$, $t \in [0,T]$, $(x,y) \in \mathbb{R}^{2d}$ and $h = (h_1,h_2) \in \mathbb{R}^{2d}$ small:

$$\lim_{|h|\to 0} \sup_{\eta} \int_{\mathbb{R}^{2d}} |p_\eta(t,(x,y),(w,z)) - p_\eta(t,(x,y),(w+h_1,z+h_2))| dwdz = 0. \quad (4.22)$$

The equation (4.22) can be proved in the same technique as in [SV79] taking into account the fact transition densities of the SDEs with mollified coefficients b_η, σ_η are smooth and the limit of the parametrix sum is a continues function.

Thus, we have proved the below proposition.

Proposition 4.2.2. *Under the sole assumption (**AD**), for* $t > 0$*, the transition density of the process* (X_t, Y_t) *solving* (4.9) *exists and can be written as the series in* (4.19) *with* $\eta = 0$.

Parametrix expansion. Scheme

Let us introduce the approximation scheme for (4.9). For a fixed N and $T > 0$ we define a time grid $\{0, t_1, \dots, t_N\}$ with a given step $h := T/N$, i.e. $t_i = ih$, for $i = 0, \dots, N$ and the scheme

$$\begin{cases} X_t^h = x + \int_0^t b(X_{\phi(s)}^h, Y_{\phi(s)}^h)ds + \int_0^t \sigma(X_{\phi(s)}^h, Y_{\phi(s)}^h)dW_s, \\ Y_t^h = y + \int_0^t X_s^h ds. \end{cases} \quad (4.23)$$

where $\phi(t) = t_i \ \forall t \in [t_i, t_{i+1})$. Observe that the above scheme is in fact well defined even though the non degenerate component of the scheme itself appears in the integral. On every time-step the increments of $(X_t^h, Y_t^h)_{t\in[t_i,t_{i+1}]}$, $i \geq 0$ are actually Gaussian. They indeed correspond to a suitable rescaling of the Brownian increment and its integral on the considered time step, see also Remark 4.2.3.

Remark 4.2.3. The specific version of the Euler scheme we perform for the model (4.9) gives us directly the existence of the one step Gaussian transition density whereas in the generic Markov settings of [KMM10] some "aggregations" are needed. Namely, under assumptions **(A)** the discretization scheme (4.23) admits a Gaussian transition density: for all $(x,y) \in \mathbb{R}^{2d}, 0 < j \leq N, A \in \mathcal{B}(\mathbb{R}^{2d})$ (where $\mathcal{B}(\mathbb{R}^{2d})$ stands for the

Borel $\sigma-$field of $\mathbb{R}^{2d}$) we get:

$$\begin{aligned}
&\mathbb{P}[(X^h_{t_j}, Y^h_{t_j}) \in A | (X^h_0, Y^h_0) = (x,y)] \\
= &\int_{(\mathbb{R}^{2d})^{j-1}\times A} p_h(h,(x,y),(x_1,y_1))p_h(h,(x_1,y_1),(x_2,y_2)) \times \dots \\
&\times p_h(h,(x_{j-1},y_{j-1}),(x_j,y_j))dx_1dy_1dx_2dy_2\dots dx_jdy_j \\
&=: \int_A p_h(t_j,(x,y),(x_j,y_j))dx_jdy_j,
\end{aligned}$$

where the notation $p_h(h,(x_i,y_i),(x_{i+1},y_{i+1})), i \in [0, N-1]$, stands for the density of a Gaussian random variable with mean $\begin{pmatrix} x_i + b(x_i,y_i)h \\ y_i + x_ih + b(x_i,y_i)h^2/2 \end{pmatrix}$ and non degenerate covariance matrix $\begin{pmatrix} a(x_i,y_i)h & a(x_i,y_i)h^2/2 \\ a(x_i,y_i)h^2/2 & a(x_i,y_i)h^3/3 \end{pmatrix}$. Two-sided Gaussian bounds of Kolmogorov type for the scheme transition density $p_h(t_j,(x,y),(x_j,y_j))$ have been established in [LM10].

As for the diffusion density, we would like to take the advantage of applying the parametrix technique to the discretization scheme transition density (4.23) as in [LM10]. We first need to introduce the *frozen* version for the scheme (4.23) and the discrete counterpart of the time-space convolution kernel. From this we can derive the parametrix representation for the density of the discretization scheme.

For fixed points $(x,y),(x',y') \in \mathbb{R}^{2d}$, the fixed final time t_j,$0 \le j \le j' \le N$ we define

$$\left(\tilde{X}^h_t, \tilde{Y}^h_t\right)_{t\in[0,t_j]} \left(\equiv \left(\tilde{X}^{h,(x',y',t_j)}_t, \tilde{Y}^{h,(x',y',t_j)}_t\right)_{t\in[0,t_j]}\right)$$

by $\left(\tilde{X}^h_0, \tilde{Y}^h_0\right) = (x,y)$, and $\forall t \in (0,t_j)$:

$$\begin{cases} \tilde{X}^h_t = x + \int_0^t \sigma(x', y' - x'(t_j - \phi(s)))dW_s, \\ \tilde{Y}^h_t = y + \int_0^t \tilde{X}^h_v dv = y + xt + \int_0^t\int_0^v \sigma(x', y' - x'(t_j - \phi(s)))dW_s dv, \end{cases} \tag{4.24}$$

where $\phi(t) = t_i,\ \forall t \in [t_i, t_{i+1})$.

Let us emphasise that $\int_0^t\int_0^v \sigma(x', y' - x'(t_j - \phi(s)))dW_sdv = \int_0^t (t-v)\sigma(x', y' - x'(t_j - \phi(s)))dW_s := \int_0^t (t-v)\tilde{\sigma}_{\phi(s)}dW_s$ (the equality means that processes are equal in distributions) as two Gaussian processes with zero-means and the same covariance matrices. Setting $\forall s \in [0,t_j], \tilde{a}_{\phi(s)} = \tilde{\sigma}_{\phi(s)}\tilde{\sigma}^*_{\phi(s)}$, recall from **(AD2)** condition that $\tilde{a}_{\phi(s)}$ is symmetric, one can finally obtain that the covariance matrix $\Sigma^h_{t_j}$ of the vector $(\tilde{X}^h_{t_j}, \tilde{Y}^h_{t_j})$ is equal to

$$\Sigma^h_{t_j} = \begin{pmatrix} \int_0^{t_j} \tilde{a}_{\phi(s)}ds & \int_0^{t_j}(t_j - s)\tilde{a}_{\phi(s)}ds \\ \int_0^{t_j}(t_j - s)\tilde{a}_{\phi(s)}ds & \int_0^{t_j}(t_j-s)^2\tilde{a}_{\phi(s)}ds \end{pmatrix}$$

The frozen process also depends on t_j through an additional term in the diffusion coefficient. From now on, $\tilde{p}^{h,t_{j'},(x',y')}$ denotes the transition density of the discretization scheme (4.24) and let us emphasize that for the frozen coefficients, we will denote for simplicity $\tilde{p}^{h,t_j,(x',y')}(t_{j'},(x,y),(\cdot,\cdot)) =: \tilde{p}^h(t_{j'},(x,y),(\cdot,\cdot))$ - the transition density between times 0 and $t_{j'} \le t_j$ of the frozen Markov chain.

Let us now introduce the discrete counterpart of the parametrix kernel considered for the continuous objects in (4.18). To this end, for a sufficiently smooth function $\psi : \mathbb{R}^{2d} \to \mathbb{R}$ and fixed $(x',y') \in \mathbb{R}^{2d}, j \in (0,N]$ define operators $\left(L^h\right)$ and $\left(\tilde{L}^h\right)\left(\equiv \left(\tilde{L}^{h,t_j,(x',y')}\right)\right)$

$$L^h f(t_j,(x,y),(x',y')) = h^{-1}\left[\int_{\mathbb{R}^{2d}} p^h(h,(x,y),(u,v)) f(t_j-h,(u,v),(x',y'))dudv - f(t_j-h,(x,y),(x',y'))\right],$$

$$\tilde{L}^h f(t_j,(x,y),(x',y')) = h^{-1}\left[\int_{\mathbb{R}^{2d}} \tilde{p}^h(h,(x,y),(u,v)) f(t_j-h,(u,v),(x',y'))dudv - f(t_j-h,(x,y),(x',y'))\right].$$

Define the discrete kernel H_h by

$$H_h(t_j,(u,v),(x',y')) = \left(L^h - \tilde{L}^h\right)\tilde{p}^h(t_j-h,(u,v),(x',y')), \quad 0 \le j \le N. \tag{4.25}$$

From the previous definition, for all $0 \le j \le N$

$$H_h(t_j,(u,v),(x',y')) = h^{-1}\int_{\mathbb{R}^{2d}} \left[p^h - \tilde{p}^h\right](h,(u,v),(w,z))\tilde{p}^h(t_j-h,(w,z),(x',y'))dwdz.$$

Analogously to Lemma 3.6 in [KM00], which follows from a direct algebraic manipulation, it has been derived in [LM10] that the transition density of the scheme admits the following representation.

Proposition 4.2.3 (Parametrix Expansion for the Euler scheme). *Assume that the assumptions* ***(AD)*** *are in force. Then*

$$p^h(t_j,(x,y),(x',y')) = \sum_{r=0}^{j}\left(\tilde{p} \otimes_h H_h^{(r)}\right)(t_j,(x,y),(x',y')), \tag{4.26}$$

for the discrete time convolution type operator $\otimes_h$ *defined by*

$$(g \otimes_h f)(t_j, (x, y), (x', y')) = \sum_{i=0}^{j-1} h \int_{\mathbb{R}^{2d}} g(t_i, (x, y), (u, v)) f(t_j - t_i, (u, v), (x', y')) du dv,$$

where $g \otimes_h H_h^{(0)} := g$, *and for all* $r \geq 1, H_h^{(r)} = H_h \otimes_h H_h^{(r-1)}$ *denotes the* $r-$*fold discrete convolution of the kernel* H_h. *W.r.t. the above definition, we use the convention that* $\tilde{p}^h \otimes_h H_h^{(r)}(0, (x, y), (x', y')) = 0, r \geq 1$.

4.3 Stability results

In this section we are going to study the sensitivity of the transition densities of some Kolmogorov like degenerate diffusion processes with respect to a perturbation of the coefficients of the non-degenerate component.

4.3.1 Stability for perturbed diffusions

We now introduce a perturbed version of (4.9) with dynamics:

$$\begin{cases} dX_t^{(\varepsilon)} = b_\varepsilon(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dt + \sigma_\varepsilon(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})dW_t, \\ dY_t^{(\varepsilon)} = X_t^{(\varepsilon)} dt, t \in [0, T], \end{cases} \tag{4.27}$$

where $b_\varepsilon : \mathbb{R}^{2d} \to \mathbb{R}^d$, $\sigma_\varepsilon : \mathbb{R}^{2d} \to \mathbb{R}^d \otimes \mathbb{R}^d$ satisfy at least the same assumptions as b, σ and are in some sense meant to be *close* to b, σ for small values of $\varepsilon > 0$. In particular, from Proposition 4.2.3 we have that $(X_t^{(\varepsilon)}, Y_t^{(\varepsilon)})$ admits a density.

The goal of this Section is to investigate how the closeness of $(b_\varepsilon, \sigma_\varepsilon)$ and (b, σ) is reflected on the respective densities of the associated processes.

In many applications (misspecified volatility models or calibration procedures) it can be useful to know how the controls on the differences $|b - b_\varepsilon|, |\sigma - \sigma_\varepsilon|$ (for suitable norms) impact the difference $p_\varepsilon - p$ of the densities corresponding respectively to the dynamics with the perturbed parameters and the one of the model.

Let us now introduce, under **(AD)**, the quantities that will bound the difference of the densities in our main results below. Set for $\varepsilon > 0$:

$$\forall q \in (1, +\infty], \ \Delta^d_{\varepsilon, b, q} := |b(.,.) - b_\varepsilon(.,.)|_{L^q(\mathbb{R}^d)}.$$

Since $\sigma, \sigma_\varepsilon$ are both γ-Hölder continuous, see (**A3**), we also define

$$\Delta^d_{\varepsilon, \sigma, \gamma} := |\sigma(.,.) - \sigma_\varepsilon(.,.)|_{d, \gamma},$$

where $\gamma \in (0,1]$, $|.|d,_\gamma$ stands for the Hölder norm in space on $C^\gamma_{b,\mathbf{d}}(\mathbb{R}^d, \mathbb{R}^d \otimes \mathbb{R}^d)$, which denotes the space of Hölder continuous bounded functions with respect to the distance $\mathbf{d}$ defined as follows:

$$\forall (x,y),\ (x',y') \in (\mathbb{R}^d)^2,\ \mathbf{d}\big((x,y),(x',y')\big) := |x-x'| + |y'-y|^{1/3}. \tag{4.28}$$

Namely, a measurable function f is in $C^\gamma_{b,\mathbf{d}}(\mathbb{R}^d, \mathbb{R}^d \otimes \mathbb{R}^d)$ if

$$|f|_{\mathbf{d},\gamma} := \sup_{x\in\mathbb{R}^d} |f(x)| + [f]_{\mathbf{d},\gamma},\ [f]_{d,\gamma} := \sup_{(x,y)\neq(x',y')\in\mathbb{R}^{2d}} \frac{|f(x,y)-f(x',y')|}{\mathbf{d}\big((x,y),(x',y')\big)^\gamma} < +\infty.$$

The previous control in particular implies for all $((x,y),(x',y')) \in (\mathbb{R}^{2d})^2$:

$$|a(x,y) - a(x',y') - a_\varepsilon(x,y) + a_\varepsilon(x',y')| \ \le\ 2^{2-\gamma}(K+\kappa)\Delta^d_{\varepsilon,\sigma,\gamma}\mathbf{d}^\gamma\big((x,y),(x',y')\big).$$

We eventually set $\forall q \in (1,+\infty]$,

$$\Delta^d_{\varepsilon,\gamma,q} := \Delta^d_{\varepsilon,\sigma,\gamma} + \Delta^d_{\varepsilon,b,q},$$

which will be the key quantity governing the error in our results.

Theorem 4.3.1 (Stability Control). *Fix $T > 0$. Under (**AD**), for $q \in (4d,+\infty]$, there exists $C := C(q) \ge 1, c \in (0,1]$ s.t. for all $0 < t \le T, ((x,y),(x',y')) \in (\mathbb{R}^{2d})^2$:*

$$|(p-p_\varepsilon)(t,(x,y),(x',y'))| \le C\Delta^d_{\varepsilon,\gamma,q}p_{c,K}(t,(x,y),(x',y')),$$

where $p(t,(x,y),(.,.)), p_\varepsilon(t,(x,y),(.,.))$ respectively stand for the transition densities at time t of equations (4.9), (4.27) starting from (x,y) at time 0.

Proof. We will now investigate more specifically the sensitivity of the density w.r.t. the coefficients perturbation through the difference of the series. From Proposition 4.2.2 , for a given fixed parameter ε, under (A) the densities $p(t,(x,y),(\cdot,\cdot)), p_\varepsilon(t,(x,y),(\cdot,\cdot))$ at time t of the processes in (4.9), (4.27) starting from (x,y) at time 0 both admit a parametrix expansion of the previous type.

Let us consider the difference between the two parametrix expansions for (4.9) and (4.27) in the form (4.19):

$$\begin{aligned} &|p(t,(x,y),(x',y')) - p_\varepsilon(t,(x,y),(x',y'))| \\ \le &\sum_{r=0}^{+\infty} |\tilde p \otimes H^{(r)}(t,(x,y),(x',y')) - \tilde p_\varepsilon \otimes H^{(r)}_\varepsilon(t,(x,y),(x',y'))|. \end{aligned}$$

Since we consider perturbations of the densities with respect to the non-degenerate component, following the same steps as in [KKM17] one can show that the Lemma below holds:

Lemma 4.3.2 (Difference of the first terms and their derivatives)**.** *There exist $c_1 \geq 1,\ c \in (0,1]$ s.t. for all $0 < t,\ (x,y),(x',y') \in \mathbb{R}^{2d}$ and all multi-index $\alpha,\ |\alpha| \leq 4$,*

$$|D_x^\alpha \tilde{p}(t,(x,y),(x',y')) - D_x^\alpha \tilde{p}_\varepsilon(t,(x,y),(x',y'))| \leq \frac{c_1 \Delta^d_{\varepsilon,\sigma,\gamma} p_{c,K}(t,(x,y),(x',y'))}{t^{|\alpha|/2}}.$$

Lemma 4.3.3 (Control of the one-step convolution)**.** *For all $0 < t,\ (x,y),(x',y') \in \mathbb{R}^{2d}$:*

$$\begin{aligned}
&|\tilde{p} \otimes H^{(1)}(t,(x,y),(x',y')) - \tilde{p}_\varepsilon \otimes H^{(1)}_\varepsilon(t,(x,y),(x',y'))| \\
&\leq c_1^2 \Big\{ (1 \vee T^{(1-\gamma)/2})^2 [\Delta^d_{\varepsilon,\sigma,\gamma} + \mathbb{I}_{q=+\infty} \Delta^d_{\varepsilon,b,+\infty}] B(1, \frac{\gamma}{2}) t^{\frac{\gamma}{2}} \\
&+ \mathbb{I}_{q \in (4d,+\infty)} \Delta^d_{\varepsilon,b,q} B(\frac{1}{2} + \alpha(q), \alpha(q)) t^{\alpha(q)} \Big\} p_{c,K}(t,(x,y),(x',y')),
\end{aligned} \tag{4.29}$$

where c_1, c are as in Lemma 4.3.2 and for $q \in (4d,+\infty)$ we set $\alpha(q) = \frac{1}{2} - \frac{2d}{q}$.

Proof. Let us write:

$$\begin{aligned}
&|\tilde{p} \otimes H^{(1)}(t,(x,y),(x',y')) - \tilde{p}_\varepsilon \otimes H^{(1)}_\varepsilon(t,(x,y),(x',y'))| \leq \\
&|(\tilde{p} - \tilde{p}_\varepsilon) \otimes H(t,(x,y),(x',y'))| + |\tilde{p}_\varepsilon \otimes \Big(H - H_\varepsilon\Big)(t,(x,y),(x',y'))| := I + II.
\end{aligned} \tag{4.30}$$

From Lemma 4.3.2 and (4.20) we readily get for all $q \in (4d,+\infty]$:

$$I \leq ((1 \vee T^{(1-\gamma)/2}) c_1)^2 \Delta^d_{\varepsilon,\gamma,q} p_{c_2,K}(t,(x,y),(x',y')) B(1, \frac{\gamma}{2}) t^{\frac{\gamma}{2}}. \tag{4.31}$$

To estimate (II) let us first consider $H - H_\varepsilon$ more precisely:

$$\begin{aligned}
&(H - H_\varepsilon)(t-u,(w,z),(x',y')) \qquad (4.32)\\
&= \frac{1}{2} \mathrm{Tr}\Big\{ a(w,z) - a(x', y' - x'(t-u)) - a_\varepsilon(w,z) + a_\varepsilon(x', y' - x'(t-u) \Big\} \\
&\quad \times D_w^2 \tilde{p}(t-u,(w,z),(x',y')) \\
&+ \frac{1}{2} \mathrm{Tr}\Big\{ a_\varepsilon(w,z) - a_\varepsilon(x', y' - x'(t-u)) \Big\} \Big[D_w^2 (\tilde{p} - \tilde{p}_\varepsilon) \Big] (t-u,(w,z),(x',y')) \\
&+ \langle b(w,z) - b_\varepsilon(w,z), D_w \tilde{p}(t-u,(w,z),(x',y')) \rangle \\
&+ \langle b_\varepsilon(w,z), D_w (\tilde{p} - \tilde{p}_\varepsilon)(t-u,(w,z),(x',y')) \rangle \\
&:= \Big(\Delta^1_\varepsilon H + \Delta^2_\varepsilon H \Big)(t-u,(w,z),(x',y')) \\
&+ \langle b(w,z) - b_\varepsilon(w,z), D_w \tilde{p}(t-u,(w,z),(x',y')) \rangle \\
&+ \langle b_\varepsilon(w,z), (D_w \tilde{p} - D_w \tilde{p}_\varepsilon)(t-u,(w,z),(x',y')) \rangle.
\end{aligned}$$

Since functions $a(w,z), a_\varepsilon(w,z)$ are Hölder uniformly continuous and (4.17) holds than:

$$\begin{aligned}&|\Delta_\varepsilon^1 H|(t-u,(w,z),(x',y'))|\\ \le\ & \frac{c\Delta^d_{\varepsilon,\gamma,\infty}\left(|w-x'|^\gamma+|z-y'+x'(t-u)|^{\gamma/2}\right)p_{c,K}(t-u,(w,z),(x',y'))}{(t-u)}\\ &\le c\Delta^d_{\varepsilon,\gamma,\infty}\frac{p_{c_2,K}(t-u,(w,z),(x',y'))}{(t-u)^{1-\gamma/2}}.\end{aligned}$$

From Lemma 4.3.2 and Hölder uniform continuity of the function $a_\varepsilon(x,y)$ it follows:

$$\begin{aligned}&|\Delta_\varepsilon^2 H|(t-u,(w,z),(x',y'))\\ \le\ & \frac{c\Delta^d_{\varepsilon,\gamma,\infty}\left(|w-x'|^\gamma+|z-y'+x'(t-u)|^{\gamma/3}\right)p_{c,K}(t-u,(w,z),(x',y'))}{(t-u)}\\ &\le c\Delta^d_{\varepsilon,\gamma,\infty}\frac{p_{\tilde c_2,K}(t-u,(w,z),(x',y'))}{(t-u)^{1-\gamma/2}}.\end{aligned}$$

Thus, the fact that $|b(w,z)-b_\varepsilon(w,z)|\le c\Delta^d_{\varepsilon,b,\gamma}$ and (4.17) give the control for $q=+\infty$. Namely,

$$|(H-H_\varepsilon)(t-u,(w,z),(x',y'))|\le\left(1\vee T^{(1-\gamma)/2}\right)c_1\Delta^d_{\varepsilon,\gamma,\infty}\left[\frac{p_{c,K}(t-u,(w,z),(x',y'))}{(t-u)^{1-\gamma/2}}\right]. \tag{4.33}$$

For $q\in(4d,+\infty)$ we use Hölder inequality in the time-space convolution involving the difference of the drifts (last term in (4.32)). Set

$$\begin{aligned}&D(t,(x,y),(x',y'))\\ :=\ &\int_0^t du\int_{\mathbb{R}^{2d}}\tilde p_\varepsilon(u,(x,y),(w,z))\langle[b_\varepsilon(w,z)-b(w,z)],D_w\tilde p(t-u,(w,z)(x',y'))\rangle dwdz.\end{aligned}$$

Denoting by $\bar q$ the conjugate of q, i.e. $q,\bar q>1, q^{-1}+\bar q^{-1}=1$, we get from (4.17) and for $q > d$ that:

$$|D(t,(x,y),(x',y'))| \le c_1^2 \int_0^t \frac{du}{(t-u)^{1/2}} \|b(.,.) - b_\varepsilon(.,.)\|_{L^q(\mathbb{R}^d)}$$
$$\times\Big\{ \int_{\mathbb{R}^{2d}} [p_{c,K}(u,(x,y),(w,z)) p_{c,K}(t-u,(w,z),(x',y'))]^{\bar q} dwdz \Big\}^{1/\bar q}$$
$$\le c_1^2 \Delta^d_{\varepsilon,b,q} \int_0^t \frac{3^{d/q} c^{2d}}{(2\pi)^{2d/q} (c\bar q)^{2d/\bar q}}$$
$$\times\Big\{ \int_{\mathbb{R}^{2d}} p_{c\bar q,K}(u,(x,y),(w,z)) p_{c\bar q,K}(t-u,(w,z),(x',y')) dwdz \Big\}^{1/\bar q} \frac{du}{u^{2d/q}(t-u)^{\frac12 + 2d/q}}$$
$$\le c_1^2 \left(\frac{\sqrt{3} c t^2}{2\pi}\right)^{d/q} \bar q^{\frac{d}{\bar q}} \Delta^d_{\varepsilon,b,q} p_{c,K}(t,(x,y),(x',y')) \int_0^t \frac{du}{u^{2d/q}(t-u)^{\frac12 + 2d/q}}.$$

Now, the constraint $4d < q < +\infty$ precisely gives that $\frac12 + 2d(1 - \frac{1}{\bar q}) < 1$ so that the last integral is well defined. We therefore derive:

$$|D(t,(x,y),(x',y')|$$
$$\le c_1^2 t^{\frac12 - 2d/q} \Delta^d_{\varepsilon,b,q} p_{c,K}(t,(x,y),(x',y')) B(1 - 2d/q, \frac12 - 2d/q).$$

In the case $4d < q < +\infty$, recalling that $\alpha(q) = \frac12 - \frac{2d}{q}$, we eventually get :

$$\begin{aligned} &|\tilde p_\varepsilon(s,(x,y),(w,z)) \otimes \Big(H - H_\varepsilon\Big)(t-u,(w,z),(x',y'))| \\ &\le c_1^2 p_{c,K}(t,(x,y),(x',y'))\{\Delta^d_{\varepsilon,b,q} t^{\alpha(q)} B(\frac12 + \alpha(q), \alpha(q)) \\ &\quad + 2\Delta^d_{\varepsilon,\sigma,\gamma}(1 \vee T^{(1-\gamma)/2}) t^{\gamma/2} B(1,\gamma/2)\}. \end{aligned} \tag{4.34}$$

The statement now follows in whole generality from (4.30), (4.31), (4.17) for $q = \infty$ and (4.34) for $4d < q < +\infty$.

□

The following Lemma associated with Lemmas 4.3.2 and Lemma 4.3.3 allows to complete the proof of Theorem 4.3.1.

Lemma 4.3.4 (Difference of the iterated kernels)**.** *For all* $0 < t \le T$, $(x,y),(x',y') \in (\mathbb{R}^{2d})^2$ *and for all* $r \in \mathbb{N}$*:*

$$\begin{aligned} &|(\tilde p \otimes H^{(r)} - \tilde p_\varepsilon \otimes H_\varepsilon^{(r)})(t,(x,y),(x',y')| \\ &\le C^r r \Delta^d_{\varepsilon,\gamma,q} \left\{ \frac{t^{\frac{r\gamma}{2}}}{\Gamma\left(1 + \frac{r\gamma}{2}\right)} + \frac{t^{\frac{(r+2)\gamma}{2}}}{\Gamma\left(1 + \frac{(r+2)\gamma}{2}\right)} \right\} p_{c,K}(t,(x,y),(x',y')). \end{aligned} \tag{4.35}$$

Proof. Observe that Lemmas 4.3.2 and Lemma 4.3.3 respectively give (4.35) for $r = 0$ and $r = 1$. Let us assume that it holds for a given $r \in \mathbb{N}^*$ and let us prove it for $r+1$.

Let us denote for all $r \geq 1$,
$\eta_r(t,(x,y),(x',y')) := |(\tilde{p} \otimes H^{(r)} - \tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)})(t,(x,y),(x',y'))|$. Write

$$\begin{aligned}\eta_{r+1}(t,(x,y),(x',y')) &= \left|(\tilde{p} \otimes H^{(r)} - \tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)}) \otimes H(t,(x,y),(x',y'))\right| \\ &\quad + \left|\tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)} \otimes (H - H_\varepsilon)(t,(x,y),(x',y'))\right| \\ &\leq \eta_r \otimes |H|\,(t,(x,y),(x',y')) + \left|\tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)}\right| \otimes |(H - H_\varepsilon)|\,(t,(x,y),(x',y')).\end{aligned}$$

Now, η_r is controlled by the induction hypothesis, $|H|$ - through (4.20), Lemma 4.2.1 provides bounds for the convolution $\tilde{p}_\varepsilon \otimes H_\varepsilon^{(r)}$ and the difference $|(H - H_\varepsilon)|$ is controlled in (4.33). Thus, the induction hypothesis we get the result. □

Theorem 4.3.1 now simply follows from the controls of Lemma 4.3.4, the parametrix expansions (4.9) and (4.27) of the densities p, p_ε and the asymptotic of the Gamma function.

□

Stability for perturbed Euler schemes

Let us describe precisely the analogue of the scheme (4.23) with perturbed coefficients as in (4.27) which approximates the process (4.27) with perturbed coefficients $b_\varepsilon, \sigma_\varepsilon$:

$$\begin{cases} X_t^{\varepsilon,h} = x + \int_0^t b_\varepsilon(X_{\phi(s)}^{\varepsilon,h}, Y_{\phi(s)}^{\varepsilon,h})ds + \int_0^t \sigma_\varepsilon(X_{\phi(s)}^{\varepsilon,h}, Y_{\phi(s)}^{\varepsilon,h})dW_s, \\ Y_t^{\varepsilon,h} = y + \int_0^t X_s^{\varepsilon,h} ds. \end{cases} \tag{4.36}$$

for $t \in [0, t_j), 0 < j \leq N$, where $\phi(t) = t_i \ \forall t \in [t_i, t_{i+1})$.

Recall first from the Section 4.2.2 that we have the following representations for the densities p^h and p_ε^h :

$$p^h(t_j,(x,y),(x',y')) = \sum_{r=0}^{j} \tilde{p}^h \otimes_h H_h^{(r)}(t_j,(x,y),(x',y')),$$

$$p_\varepsilon^h(t_j,(x,y),(x',y')) = \sum_{r=0}^{j} \tilde{p}_\varepsilon^h \otimes_h H_{\varepsilon,h}^{(r)}(t_j,(x,y),(x',y')),$$

where $H_h^{\varepsilon,(r)}$ is defined analogously to $H_h^{(r)}$ in (4.25) with $L^h, \tilde{L}^h, \tilde{p}^h$ changed respectively by their perturbed counterparts in ε.

Theorem 4.3.5. *Fix $T > 0$ and let us define a time-grid $\Lambda_h := \{(t_i)_{i \in [\![1,N]\!]}\}, N \in \mathbb{N}^*$. Under (**A**), there exists $C \geq 1, c \in (0,1]$ s.t. for all $0 < t_j \leq T, ((x,y),(x',y')) \in (\mathbb{R}^{2d})^2$, $q \in (4d, +\infty]$:*

$$|p_h^\varepsilon - p_h|(t_j,(x,y),(x',y')) \leq C\Delta_{\varepsilon,\gamma,q}^d p_{c,K}(t_j,(x,y),(x',y')),$$

where $p_h^\varepsilon(t,(x,y),(.,.)), p_h(t,(x,y),(.,.))$ *respectively stand for the transition densities at time* t *of equations* (4.23), (4.36) *starting from* (x,y) *at time* 0.

The closeness of "main" parts $\tilde{p}^h$ and $\tilde{p}_\varepsilon^h$ in the above expansions can be derived analogously to Lemma 1 [KKM17] as the difference between two Gaussian densities with the small differences in means and covariances. The only point we would like to emphasize - the Kolmogorov-like density $p_{c,K}$ which stands in the bounds due to the control for the scheme transition density in the degenerate case. The complete proof could be found in [LM10], Theorem 2.1, (b).

Lemma 4.3.6 (Control and Comparison of the densities and their derivatives)**.** *There exist* $c_1 \geq 1,\ c \in (0,1]$ *s.t. for all* $0 < t_j \leq T,\ (x,y),(x',y') \in \mathbb{R}^{2d}$ *and all multi-index* $\alpha,\ |\alpha| \leq 4$*,*

$$|D_x^\alpha \tilde{p}^h(t_j,(x,y),(x',y')) - D_x^\alpha \tilde{p}_\varepsilon^h(t_j,(x,y),(x',y'))| \leq \frac{c_1 \Delta_{\varepsilon,\sigma,\gamma}^d p_{c,K}(t_j,(x,y),(x',y'))}{t_j^{|\alpha|/2}}.$$

where the last inequality holds for all $\eta \in (0,\gamma)$ *due to the mollification procedure.*

Proof. According to the definition

$$\tilde{p}_\varepsilon^h(t_j,(x,y),(x',y')) = \frac{1}{t_j^{2d}\det(V_j^\varepsilon)^{1/2}} G\left((V_j^\varepsilon)^{-1/2}\begin{pmatrix} \frac{x'-x}{\sqrt{t_j}} \\ \frac{y'-y-xt_j}{t_j^{3/2}} \end{pmatrix}\right), \qquad (4.37)$$

where

$$V_j^\varepsilon = \begin{pmatrix} \frac{1}{t_j}\int_0^{t_j} \tilde{a}_{\phi(s)}^\varepsilon ds & \frac{1}{t_j^2}\int_0^{t_j} \tilde{a}_{\phi(s)}^\varepsilon (t_j - s) ds \\ \frac{1}{t_j^2}\int_0^{t_j} \tilde{a}_{\phi(s)}^\varepsilon (t_j - s) ds & \frac{1}{t_j^3}\int_0^{t_j} \tilde{a}_{\phi(s)}^\varepsilon (t_j - s)^2 ds \end{pmatrix}$$

and $\forall z \in \mathbb{R}^{2d}, G(z) = \exp(-|z|^2/2)(2\pi)^{-d}$ stands for the density of the standard Gaussian vector of $\mathbb{R}^{2d}$. We emphasize that, in (4.37) we introduced the matrix V_j^ε which is non-degenerate and has order *one*, i.e. there exists $c := c(\mathbf{A}D) \geq 1$ s.t. $c^{-1}I_{2d} \leq V_j \leq cI_{2d}$. The matrix V_j^ε then acts on the components renormalized at their intrinsic scales, namely $\begin{pmatrix} \frac{x'-x}{\sqrt{t_j}} \\ \frac{y'-y-xt_j}{t_j^{3/2}} \end{pmatrix}$.

Taking the result from Lemma 4.3.6, the control for the difference $|\tilde{p}^h - \tilde{p}_\varepsilon^h|(t_j,(x,y),(x',y'))$ comes from the closeness of two Gaussian densities with the same mean and slightly different covariance matrices V_j and V_j^ε, recalling that by the definition V_j^0 is equal to V_j.

As $|\mathrm{det}V_j - \mathrm{det}V_j^\varepsilon| \le C(T,d)\Delta^d_{\varepsilon,\sigma,\gamma}$ for any $j \le N$, where $C(T,d)$ stands for the constant which depends only on the fixed time T and the dimension d. Also due to the definition of $\mathrm{det}V_j$, it has the first order in time.

Thus,

$$\begin{aligned}
&|\tilde p^h - \tilde p^h_\varepsilon|(t_j,(x,y),(x',y'))\\
&\le \frac{1}{(2\pi)^d t_j^{2d}}\left(\frac{1}{\det(V_j)^{1/2}} - \frac{1}{\det(V_j^\varepsilon)^{1/2}}\right)\exp\left(-\frac12\langle V_j^{-1}\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix},\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix}\rangle\right)\\
&\quad+\frac{1}{(2\pi)^d t_j^{2d}\det(V_j^\varepsilon)^{1/2}}\left(\exp\left(-\frac12\langle V_j^{-1}\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix},\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix}\rangle\right)\right.\\
&\quad\left.-\exp\left(-\frac12\langle (V_j^\varepsilon)^{-1}\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix},\begin{pmatrix}\frac{x'-x}{\sqrt{t_j}}\\ \frac{y'-y-xt_j}{t_j^{3/2}}\end{pmatrix}\rangle\right)\right)\\
&\le C(T,d)\Delta_{\varepsilon,\sigma,\gamma}p_{c,K}(t_j,(x,y),(x',y')),
\end{aligned}$$

where the difference between two exponents of scalar products can be control as usual - using the first order Taylor expansion. Dealing with $\alpha : |\alpha| > 0$ brings us additional polynomials multiplied with each exponents - the same as for the frozen densities for the diffusions. □

Lemma 4.3.7 (Control of the One-Step Convolution for the Chain)**.** *For all $\beta \in (0,\gamma)$, $0 < t_i \le T$, $(x,y),(x',y') \in \mathbb{R}^{2d}$ there exists C_β such that:*

$$|H_h(t_i,(u,v),(x',y')) - H_{h,\varepsilon}(t_i,(u,v),(x',y'))| \le \frac{C_\beta \Delta^d_{\varepsilon,\gamma,\infty}}{t_i^{1-\gamma/2}} p_{c,K}(t_i,(u,v),(x',y'))$$

Proof. 1. One step transition.
Note that if $t_i = h$, the transition probability $\tilde p^h(t_i - h,(\cdot,\cdot),(x',y'))$ is the Dirac measure $\delta_{x',y'}$ so that

$$\begin{aligned}
&H_h(h,(x,y),(x',y'))\\
&= h^{-1}\Big(\mathbb{E}[\delta_{x',y'}(X_h^h,Y_h^h)|X_0^h = x, Y_0^h = y] - \mathbb{E}[\delta_{x',y'}(\tilde X_h^h,\tilde Y_h^h)|\tilde X_0^h = x,\tilde Y_0^h = y]\Big),\\
&= h^{-1}\Big(p^h(h,(x,y),(x',y')) - \tilde p^h(h,(x,y),(x',y'))\Big).
\end{aligned}$$

As $p^h(h,(x,y),(x',y')), \tilde{p}^h(h,(x,y),(x',y'))$ are Gaussian densities, one can get

$$H_h(h,(x,y),(x',y')) = h^{-1}(2\sqrt{3})^d \times \left(\frac{G\begin{pmatrix} (h^{1/2}\sigma(x,y))^{-1}(x'-x-b(x',y')h) \\ 2\sqrt{3}(h^{3/2}\sigma(x,y))^{-1}(y'-y-\frac{x+x'}{2}h) \end{pmatrix}}{h^{2d}(\det(a(x,y)))^{1/2}} - \frac{G\begin{pmatrix} (h^{1/2}\sigma(x^{h'},y^{h'}))^{-1}(x'-x) \\ 2\sqrt{3}(h^{3/2}\sigma(x^{h'},y^{h'}))^{-1}(y'-y-\frac{x+x'}{2}h) \end{pmatrix}}{h^{2d}(\det(a(x^{h'},y^{h'})))^{1/2}} \right),$$

where $(x^{h'},y^{h'}) := (x', y'-x'h)$.

Applying the same technique for the perturbed version of the kernel function we get

$$H_{h,\varepsilon}(h,(x,y),(x',y')) = h^{-1}(2\sqrt{3})^d \times \left(\frac{G\begin{pmatrix} (h^{1/2}\sigma_\varepsilon(x,y))^{-1}(x'-x-b_\varepsilon(x,y)h) \\ 2\sqrt{3}(h^{3/2}\sigma_\varepsilon(x,y))^{-1}(y'-y-\frac{x+x'}{2}h) \end{pmatrix}}{h^{2d}\det(a_\varepsilon(x,y))^{1/2}} - \frac{G\begin{pmatrix} (h^{1/2}\sigma_\varepsilon(x^{h'},y^{h'}))^{-1}(x'-x) \\ 2\sqrt{3}(h^{3/2}\sigma_\varepsilon(x^{h'},y^{h'}))^{-1}(y'-y-\frac{x+x'}{2}h) \end{pmatrix}}{h^{2d}(\det(a_\varepsilon(x^{h'},y^{h'})))^{1/2}} \right).$$

As a result, the difference between kernel functions in the case of one-step transition could be estimated as the difference between Gaussian densities with close coefficients as in the Chapter 4. Also due to the fact that $\left|\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x^{h'} \\ y^{h'} \end{pmatrix}\right| \le |x'-x|(1+\frac{h}{2}) + |y'-y-\frac{x+x'}{2}h|$ it follows that $\exists c>0, C\ge 1$ s.t

$$|H_h(h,(x,y),(x',y')) - H_{h,\varepsilon}(h,(x,y),(x',y')| \le Ch^{-1+\gamma/2}\Delta^d_{\varepsilon,\gamma,\infty} p_{c,K}(h,(x,y),(x',y')).$$

Case $t_i > h$. Recall that for all $0 < i \le N$

$$H_h(t_i,(u,v),(x',y')) = h^{-1}\int_{\mathbb{R}^{2d}} \left[p^h - \tilde{p}^h\right](h,(u,v),(w,z))\tilde{p}^h(t_i-h,(w,z),(x',y'))dwdz. \tag{4.38}$$

Set $(x_h,y_h) := (x, y+hx)$, $(x'_{h,i}, y'_{h,i}) := (x', y'-x't_i)$. Define $\forall (u,v)\in\mathbb{R}^{2d}$, $B^h(u,v) := \begin{pmatrix} b(u,v)h \\ b(u,v)h^2/2 \end{pmatrix}$, $\Sigma^h(u,v) = \begin{pmatrix} h^{1/2}\sigma(u,v) & 0 \\ h^{3/2}\sigma(u,v)/2 & h^{3/2}\sigma(u,v)/(2\sqrt{3}) \end{pmatrix}$.

Introducing for all $(x,y),(w,z),(x',y')\in(\mathbb{R}^{2d})^3$ transitions: $P_h\Big((w,z),(x',y')\Big) := \Sigma^h(w,z)\begin{pmatrix} x' \\ y' \end{pmatrix}$, $T_h\Big((x,y),(w,z),(x',y')\Big) := B^h(x,y) + P_h\Big((w,z),(x',y')\Big)$, we can rewrite (4.38)

$$H_h(t_i,(x,y),(x',y')) = h^{-1}\int_{\mathbb{R}^{2d}} dudvG(u,v)$$
$$\times\Bigg\{\Bigg[\tilde p^h\Bigg(t_i-h,\begin{pmatrix} x_h\\ y_h\end{pmatrix}+T_h\Big((x,y),(x,y),(u,v)\Big),(x',y')\Bigg)-\tilde p^h\Bigg(t_i-h,\begin{pmatrix} x_h\\ y_h\end{pmatrix},(x',y')\Bigg)\Bigg]$$
$$-\Bigg[\tilde p^h\Bigg(t_i-h,\begin{pmatrix} x_h\\ y_h\end{pmatrix}+P_h\Big((x'_{h,i},y'_{h,i}),(u,v)\Big),(x',y')\Bigg)-\tilde p^h\Bigg(t_i-h,\begin{pmatrix} x_h\\ y_h\end{pmatrix},(x',y')\Bigg)\Bigg]\Bigg\}.$$

According to the Taylor expansion at order one:

$$H_h(t_i,(x,y),(x',y')) = h^{-1}\int_{\mathbb{R}^{2d}} dwdzG(w,z)\int_0^1 d\eta$$
$$\times\Bigg\{D_x\tilde p^h\Bigg(t_i-h,(x_h,y_h)+\eta T_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Bigg)\times\Bigg(T_h\Big((x,y),(x,y),(w,z)\Big)\Bigg)^{(x)}$$
$$-D_x\tilde p^h\Bigg(t_i-h,(x_h,y_h)+\eta P_h\Big((x'_{h,i},y'_{h,i}),(w,z)\Big),(x',y')\Bigg)\times\Bigg(P_h((x'_{h,i},y'_{h,i}),(w,z))\Bigg)^{(x)}\Bigg\}+$$
$$\Bigg\{h^{-1}\int_{\mathbb{R}^{2d}} dwdzG(w,z)\int_0^1 d\eta$$
$$\times\Bigg\{D_y\tilde p^h\Bigg(t_i-h,(x_h,y_h)+\eta T_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Bigg)\times\Bigg(T_h((x,y),(x,y),(w,x))\Bigg)^{(y)}-$$
$$D_y\tilde p^h\Bigg(t_i-h,(x_h,y_h)+\eta P_h\Big((x'_{h,i},y'_{h,i}\Big),(w,z)),(x',y')\Bigg)\times\Bigg(P_h\Big((x'_{h,i},y'_{h,i}),(w,x)\Big)\Bigg)^{(y)}\Bigg\}$$
$$:= (M_1^h+R_1^h)(t_i,(x,y),(x',y')),$$

where D_x, D_y denotes the differentiation w.r.t. the first and the second components respectively, (x) and (y) denote $(1:d)$ and $(d+1:2d)$

As we are interested in the difference $|H_h - H_{h,\varepsilon}|(t_i,(u,v),(x',y'))$ it is enough to estimate the closeness of $R_1^h, R_1^{h,\varepsilon}$ and $M_1^h, M_1^{h,\varepsilon}$.

We need to recall two following controls which has been mentioned before in (4.17). Let $\mu=(\mu_1,\dots,\mu_d)\in\mathbb{N}^d, \nu=(\nu_1,\dots,\nu_d)\in\mathbb{N}^d$ be multi-indices. We have, $\exists c>0, C\geq 1, \forall(\mu,\nu), |\mu|\leq 3, |\nu|\leq 4, \forall 0<i\leq N, (x,y),(x',y')\in\mathbb{R}^{2d}$,

$$|D_x^\nu D_y^\mu \tilde p^h(t_i,(x,y),(x',y'))| \leq C(t_i)^{-(|\nu|/2+3/2|\mu|)} p_{c,K}(t_i,(x,y),(x',y')).$$

Observe as well that there exists $C>0$ s.t.

$$|T_h\Big((x,y),(x,y),(w,z)\Big)^{(x)} - P_h\Big((x'_{h,i},y'_{h,i}),(w,z)\Big)^{(x)}| \leq C(h+|(x,y)-(x'_{h,i},y'_{h,i})|^\gamma h^{1/2}|(w,z)|),$$

$$|T_h\Big((x,y),(x,y),(w,z)\Big)^{(y)} - P_h\Big((x'_{h,i},y'_{h,i}),(w,z)\Big)^{(y)}| \le C(h^2 + |(x,y)-(x'_{h,i},y'_{h,i})|^\gamma h^{3/2}|(w,z)|)$$

since $b(\cdot,\cdot)$ is bounded and the difference between $\Sigma(x,y)$ and $\Sigma(x'_{h,i},y'_{h,i})$ can be controlled due to the Hölder continuity.

As in the article [LM10] expanding terms $D_x\tilde{p}^h\Bigg(t_i,(x_h,y_h)+\eta T_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Bigg)$ and $D_y\tilde{p}^h\Bigg(t_i,(x_h,y_h)+\eta T_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Bigg)$ at order 2 around (x_h,y_h) in M_1^h one can get:

$H_h(t_i,(x,y),(x',y')) = H(t_i,(x,y),(x',y')) + (R_1^h + R_2^h)(t_i,(x,y),(x',y'))$ where, we have denoted, with a slight abuse of notation $H(t_i,(x,y),(x',y')) = (L-\tilde{L})\tilde{p}^h(t_i,(x,y),(x',y'))$ whereas from the continuous case $H(t_i,(x,y),(x',y')) = (L-\tilde{L})\tilde{p}(t_i,(x,y),(x',y'))$. Pay attention that *a priori*, $\tilde{p}^h(t_i,(x,y),(x',y')) \neq \tilde{p}(t_i,(x,y),(x',y'))$. The only difference between those two objects is in the covariance matrices for which the backward transport of the final point is taken in continuous time in $\tilde{p}$ and in discrete time in δp^h.

$$R_2^h(t_i,(x,y),(x',y')) := \langle -b(x',y'), D_x\tilde{p}^h\Big(t_i-h,(x_h,y_h),(x',y')\Big)\rangle +$$
$$\frac{1}{2}\mathrm{Tr}\Bigg\{\Big(a(x,y)-a(x'_{h,i},y'_{h,i})\Big)D_x^2\tilde{p}^h\Big(t_i-h,(x_h,y_h),(x',y')\Big)\Bigg\}$$

The difference between $|H-H_\varepsilon|(t_i,(x,y),(x',y'))$ can be controlled as in (4.33). There exist $c\in(0,1], c_3\in(0,1]$ such that for $0<t_i\le T$ and all $(x,y),(x',y')\in\mathbb{R}^{2d}$

$$|H-H_\varepsilon|(t_i,(u,v),(x',y'))| \le \Big(1\vee T^{\frac{1-\gamma}{2}}\Big)c_3\Delta^d_{\varepsilon,\gamma,\infty}\frac{p_{c,K}(t_i,(x,y),(x',y'))}{(t_i)^{1-\gamma/2}}.$$

Using the definition of $R_1^h(t_i,(x,y),(x',y'))$ and the telescoping sums combined with bounds for the derivatives of the frozen transition density one can get:

$$|\Big(R_1^h - R_1^{h,\varepsilon}\Big)(t_i,(x,y),(x',y'))| = h^{-1}\int_{\mathbb{R}^{2d}} dwdzG(w,z)\int_0^1 d\eta$$
$$\times\Bigg\{D_y\tilde{p}^h\Bigg(t_i-h,(x_h,y_h)+\eta T_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Bigg)\times\Bigg(T_h((x,y),(x,y),(w,x))\Bigg)^{(y)}$$
$$-D_y\tilde{p}^h\Bigg(t_i-h,(x_h,y_h)+\eta P_h\Big((x'_{h,i},y'_{h,i}),(w,z)\Big),(x',y')\Bigg)\times\Bigg(P_h((x'_{h,i},y'_{h,i}),(w,x))\Bigg)^{(y)}\Bigg\}$$

$$-h^{-1}\int_{\mathbb{R}^{2d}} dwdzG(w,z)\int_0^1 d\eta$$

$$\Big\{D_y\tilde{p}^h_\varepsilon\Big(t_i-h,(x_h,y_h)+\eta T^\varepsilon_h\Big((x,y),(x,y),(w,z)\Big),(x',y')\Big)\times\Big(T^\varepsilon_h\Big((x,y),(x,y),(w,x)\Big)\Big)^{(y)}$$

$$-D_y\tilde{p}^h_\varepsilon\Big(t_i-h,(x_h,y_h)+\eta P^\varepsilon_h\Big((x'_{h,i},y'_{h,i}),(w,z)\Big),(x',y')\Big)\times\Big(P^\varepsilon_h\Big((x'_{h,i},y'_{h,i}),(w,x)\Big)\Big)^{(y)}\Big\}$$

$$\leq\Big(1\vee T^{\frac{1-\gamma}{2}}\Big)c_2\Delta^d_{\varepsilon,\gamma,\infty}\frac{p_{c,K}(t_i,(x,y),(x',y'))}{t_i^{1-\gamma/2}}.$$

Also the difference $|R^h_2-R^{h,\varepsilon}_2|(t_i,(x,y),(x',y'))$ can be held according to the boundedness of $b(\cdot,\cdot), b_\varepsilon(\cdot,\cdot)$, Hölder properties of $a(\cdot,\cdot), a_\varepsilon(\cdot,\cdot)$ and bounds for the derivatives $|D^\alpha_x\tilde{p}^h\Big(t_i-h,(x_h,y_h),(x',y')\Big)|$.

$$|\Big(R^h_2-R^{h,\varepsilon}_2\Big)(t_i,(x,y),(x',y'))|$$

$$=\langle b_\varepsilon(x',y')-b(x',y'),D_x\tilde{p}^h\Big(t_i-h,(x_h,y_h),(x',y')\Big)\rangle$$

$$+\langle b_\varepsilon(x',y'),\Big(D_x\tilde{p}^h_\varepsilon-D_x\tilde{p}^h\Big)\Big(t_i-h,(x_h,y_h),(x',y')\Big)\rangle$$

$$-\frac{1}{2}\mathrm{Tr}\Big\{\Big(a(x,y)-a(x'_{h,i},y'_{h,i})-a_\varepsilon(x,y)+a_\varepsilon(x'_{h,i},y'_{h,i})\Big)D^2_x\tilde{p}^h\Big(t_i-h,(x_h,y_h),(x',y')\Big)$$

$$-\frac{1}{2}\mathrm{Tr}\Big\{\Big(a_\varepsilon(x,y)-a_\varepsilon(x'_{h,i},y'_{h,i})\Big)\Big(D^2_x\tilde{p}^h_\varepsilon-D^2_x\tilde{p}^h\Big)\Big(t_i-h,(x_h,y_h),(x',y')\Big)\Big\}$$

$$\leq\frac{\Delta^d_{\varepsilon\gamma\infty}p_{c,K}(t_i-h,(x_h,y_h),(x',y'))}{(t_i-h)^{1-\gamma/2}}.$$

Thus, we finally have proved the Lemma. □

Lemma 4.3.8 (Difference of the iterated kernels)**.** *For all t_i, $i\in(0,j]$,$t_j\leq T$, $(x,y),(x',y')\in\mathbb{R}^{2d}$ and $r\in\mathbb{N}$:*

$$|(\tilde{p}^h\times H^{(r)}_h-\tilde{p}^h_\varepsilon\times H^{(r)}_{h,\varepsilon})(t_i,(x,y),(x',y'))| \qquad (4.39)$$
$$\leq C^r\Delta^d_{\varepsilon,\gamma,\infty}\left\{\frac{t_i^{\frac{r\gamma}{2}}}{\Gamma\left(1+\frac{r\gamma}{2}\right)}+\frac{t_i^{\frac{(r+2)\gamma}{2}}}{\Gamma\left(1+\frac{(r+2)\gamma}{2}\right)}\right\}p_{c,K}(t_i,(x,y),(x',y')).$$

Proof. Observe that Lemmas 4.3.6 gives (4.39) for $r = 0$. Let us assume that it holds for a given $r \in \mathbb{N}^*$ and let us prove it for $r+1$.

Let us denote for all $r \geq 1$,
$\eta_r(t_i,(x,y),(x',y')) := |(\tilde{p}^h \otimes H^{(r)} - \tilde{p}^h_\varepsilon \otimes H^{(r)}_{h,\varepsilon})(t_i,(x,y),(x',y'))|$. Write

$$\begin{aligned}\eta_{r+1}(t_i,(x,y),(x',y')) \leq &\left|(\tilde{p}_h \otimes H^{(r)}_h - \tilde{p}^h_\varepsilon \otimes H^{(r)}_{h,\varepsilon}) \otimes H_h(t_i,(x,y),(x',y'))\right| \\ &+ \left|\tilde{p}^h_\varepsilon \otimes H^{(r)}_{h,\varepsilon} \otimes (H_h - H_{h,\varepsilon})(t_i,(x,y),(x',y'))\right| \\ \leq \eta_r \otimes |H_h|\,(t_i,(x,y),(x',y')) &+ \left|\tilde{p}^h_\varepsilon \otimes H^{(r)}_{h,\varepsilon}\right| \otimes |(H_h - H_{h,\varepsilon})|\,(t_i,(x,y),(x',y')).\end{aligned}$$

Thus,from the induction hypothesis, similarly to Lemma 4.3.4, we get the result. □

Through the Lemma 4.3.8 one can prove the Lemma 4.3.5.

4.4 Weak error

In the same manner as in the article [KM17] we would like to consider the analogue to the difference between the degenerate diffusion and it's Euler scheme in the case of non-smooth coefficients.

Remark 4.4.1. We would like to emphasize that for our error controls, we need to consider $\gamma/2$ for the Hölder index of the degenerate second variable. According to the existing literature, see e.g. Lunardi [Lun97] or Priola [Pri09], concerning Schauder estimates for PDEs associated with generators deriving from (4.9), one could expect this regularity to be $\gamma/3$ which corresponds to the homogeneity index of the degenerate variable (see again the above references or Bramanti *et al.* [MGEE10] or [Men18] for some related applications to harmonic analysis). The current index appears through our analysis because of some specific properties of the model, namely the increment over time step of the degenerate component needs to be handled (unbounded coefficient). This precisely leads to the indicated restriction (see Theorem 4.5.1 and its proof).

There are two kinds of quantities we would be interested in while studying approximations of the SDE's solution. First, we can focus on the analogue to (1.4):

$$\mathcal{E}_w(f,(x,y),T,h) := \mathbb{E}_{(x,y)}[f(X^h_T, Y^h_T)] - \mathbb{E}_{(x,y)}[f(X_T,Y_T)], \tag{4.40}$$

where f is a test function that lies in a suitable functional space. The second quantity we will be interested in concerns directly the difference of the densities. We have indicated above that the Euler scheme (4.23) has a density enjoying Gaussian bounds. We can refer to the Chapter 4 to justify that, under the current Assumptions **(AD)**, the diffusion in (4.9) itself has a density. The existence of the density also follows from the well-posedness of the martingale problem associated with the generator of (4.9) and the estimates in Theorem 4.4.2. We will try to quantify, for a given time $t \in \{(t_i)_{i\in[\![0,N]\!]}\}$, in terms of h the difference

$$\mathcal{E}_d((x,y),(x',y'),t,h) := (p-p_h)(t,(x,y),(x',y')), \tag{4.41}$$

where $p(t,(x,y),(x',y'))$ (resp. $p_h(t,(x,y),(x',y'))$ denotes the density of the unique weak solution of the SDE (4.9), at time t and point (x',y') when the starting point at time 0 is (x,y) (resp. X^h given by the Euler scheme (4.23) at time t and point (x',y') when the starting point at time 0 is (x,y)).

To perform the further analysis we have to assume more about Hölder properties of coefficients as it has been already mentioned in Remark 4.4.1. Namely, instead of (**AD3**), we assume for some $\gamma \in (0,1]$, κ,

$$|b(x,y)-b(x',y')| + |\sigma(x,y)-\sigma(x',y')| \le \kappa\left(|x-x'|^{\gamma} + |y-y'|^{\gamma/2}\right).$$

and denote that as (**$\hat{\mathbf{A}}$D3**). Thus, we say that assumption **($\hat{\mathbf{A}}$D)** holds when conditions **(AD1)**, **(AD2)**,**($\hat{\mathbf{A}}$D3)** are in force.

Remark 4.4.2. Due to the boundedness of coefficient (**$\hat{\mathbf{A}}$D3**) is included in (**AD3**), meaning that all previous results, achieved under (**AD3**) still hold under (**$\hat{\mathbf{A}}$D3**).

Our first main result, which we have already mentioned in Chapter 1, is the following theorem.

Theorem 4.4.1. *Assume **($\hat{A}$D)** holds and fix $T > 0$. For any test function $f \in C^{\beta,\beta/2}(\mathbb{R}^{2d})$ ($\beta-$Hölder in the first variable and $\beta/2-$Hölder in the second variable functions) for $\beta \in (0,1]$, there exists $C > 0$, such that for $\mathcal{E}^1$ as in* (4.40)*:*

$$|\mathcal{E}_w(f,(x,y),T,h)| \le Ch^{\gamma/2}(1+|x|^{\gamma/2}).$$

Proof. Denote, using Markovian notations, $v(t,x,y) := \mathbb{E}[f(X_T^{t,(x,y)}, Y_T^{t,(x,y)})] = \int_{\mathbb{R}^{2d}} p(T-t,(x,y),(x',y'))f(x',y')dx'dy'$. Now, well posedness of the martingale problem yields that v is actually a weak solution of the PDE:

$$\begin{cases}(\partial_t v + Lv)(t,x,y) = 0,\\ v(T,x,y) = f(x,y),\ (x,y)\in\mathbb{R}^{2d},\end{cases} \tag{4.42}$$

where L stands for the generator of (4.9) at time t, i.e. for all $\varphi \in C_0^2(\mathbb{R}^{2d},\mathbb{R}), (x,y) \in \mathbb{R}^{2d}$,

$$L\varphi(x,y) = b(x,y)\cdot\nabla_x\varphi(x,y) + x\nabla_y\varphi(x,y) + \frac{1}{2}\mathrm{Tr}(a(x,y)D_x^2\varphi(x,y)).$$

Pay attention that, even though we have good controls on the spatial gradients for the non-degenerate variables, see again Theorem 4.4.2 below, in the current degenerate setting this does not yield that v is a classical solution to (4.42). Indeed, it does not seem to be an easy task to directly control pointwise, under our mild Hölder assumption ($\hat{\mathbf{A}}\mathbf{D}$) [1], the derivatives w.r.t. degenerate variable of the density p expressed as a convergent parametrix sum (see once more the proof of Theorem 4.4.2 for the parametrix expansion of the density). We also mention that similar features appear in the papers who handle Schauder estimates for PDEs related with (4.9). In [Lun97] and [Pri09] the derivatives w.r.t. to the non-degenerate variable are controlled up to order 2, whereas for the degenerate variable(s) the bounds obtained are for Hölder moduli of continuity of v (w.r.t. to those variables).

To circumvent this difficulty we need to introduce a smoothing procedure of the coefficients.

Mollification procedure.
Let us specify the mollification procedure. Namely, for a small parameter ε, we smooth suitably the coefficients and the function f introducing:

$$\begin{aligned} b_\varepsilon(x,y) &:= b\star\rho_\varepsilon(x,y) = \textstyle\int_{\mathbb{R}^{2d}} b(u,v)\rho_\varepsilon(x-u,y-v)dudv,\\ \sigma_\varepsilon(x,y) &:= \sigma\star\rho_\varepsilon(x,y) = \textstyle\int_{\mathbb{R}^{2d}} \sigma(u,v)\rho_\varepsilon(x-u,y-v)dudv,\\ f_\varepsilon(x,y) &:= f\star\rho_\varepsilon(x,y) = \textstyle\int_{\mathbb{R}^{2d}} f(u,v)\rho_\varepsilon(x-u,y-v)dudv, \end{aligned} \tag{4.43}$$

where $\star$ stands for the spatial convolution and ρ_ε is a spatial mollifier, i.e.

$$\rho_\varepsilon(x,y) = \varepsilon^{-3d}\rho(x/\varepsilon, y/\varepsilon^2),\ \ \rho\in C^\infty(\mathbb{R}^{2d}), \int_{\mathbb{R}^{2d}}\rho(x,y)dxdy = 1, |\mathrm{supp}(\rho)|\subset K,$$

for some compact set $K\subset\mathbb{R}^{2d}$.

According to the notations and the Hölder and boundness properties of the coefficients one can prove:

$$|b-b_\varepsilon| \le \left|\int_{\mathbb{R}^{2d}}(b(x,y)-b(x-u\varepsilon, y-v\varepsilon^2)\rho(u,v)dudv\right|.$$

From the Hölder continuity of b:

$$\sup_{(x,y)\in\mathbb{R}^{2d}}|(b-b_\varepsilon)(x,y)| \le C_\rho\varepsilon^\gamma,\quad C_\rho := \kappa\int_{\mathbb{R}^{2d}}(|u|^\gamma + |v|^{\gamma/2})\rho(u,v)dudv.$$

[1]Observe that if the coefficients were smooth, the Konakov and Mammen trick would also give the pointwise controls on the derivatives w.r.t. y.

The same analysis can be performed for σ_ε and f_ε so that σ_ε and f_ε satisfies Hölder conditions. This gives

$$\begin{aligned} |b-b_\varepsilon|+|\sigma-\sigma_\varepsilon| &\leq C\varepsilon^\gamma, \\ |f-f_\varepsilon| &\leq C\varepsilon^\beta. \end{aligned} \tag{4.44}$$

As the result we get the following controls for closeness of coefficients:

$$\begin{aligned} \sup_{(x,y)\in\mathbb{R}^{2d}} |b(x,y)-b_\varepsilon(x,y)| &\leq C\varepsilon^\gamma, \\ \sup_{(x,y)\in\mathbb{R}^{2d}} |f(x,y)-f_\varepsilon(x,y)| &\leq C\varepsilon^\beta, \\ \forall\eta\in(0,\gamma),\ \sup_{(x,y)\in\mathbb{R}^{2d}} |\sigma(x,y)-\sigma_\varepsilon(x,y)| &+ |(\sigma-\sigma_\varepsilon)|_\eta \\ &\leq C_\eta(\varepsilon^\gamma+\varepsilon^{\gamma-\eta}) \leq C_\eta\varepsilon^{\hat{\mathbf{d}},\gamma-\eta}, \end{aligned}$$

where

$$\forall(x,y),\ (x',y')\in(\mathbb{R}^d)^2,\ \hat{\mathbf{d}}\big((x,y),(x',y')\big) := |x-x'|+|y'-y|^{1/2}. \tag{4.45}$$

Namely, a measurable function f is in $C^\gamma_{b,\mathbf{d}}(\mathbb{R}^d,\mathbb{R}^d \otimes \mathbb{R}^d)$ if

$$|f|_{\hat{\mathbf{d}},\gamma} := \sup_{x\in\mathbb{R}^d}|f(x)|+[f]_{\hat{\mathbf{d}},\gamma},\ [f]_{\hat{\mathbf{d}},\gamma} := \sup_{(x,y)\neq(x',y')\in\mathbb{R}^{2d}} \frac{|f(x,y)-f(x',y')|}{\hat{\mathbf{d}}\big((x,y),(x',y')\big)^\gamma} < +\infty.$$

Following the arguments of Chapter 4 since we can control the closeness between the transition densities correspond to SDEs with mollified and non-mollified coefficients, it is then possible to control the difference between transition densities of the corresponding diffusions. Namely, under **($\hat{\mathbf{A}}$D)**, the exist $C_\eta \geq 1, c \leq 1$, s.t. for all $0 \leq i < j \leq N, (x,y),(x',y') \in (\mathbb{R}^{2d})^2$:

$$|(p-p^\varepsilon)(t_j,(x,y),(x',y')| \leq C_\eta\varepsilon^\gamma p_{c,K}(t_j,(x,y),(x',y')), \tag{4.46}$$

and, similarly, it is established in the Chapter 4 that the same control holds for the scheme (4.23) and its associated perturbation:

$$|(p_h-p_h^\varepsilon)(t_j,(x,y),(x',y')| \leq C_\eta\varepsilon^\gamma p_{c,K}(t_j,(x,y),(x',y')), \tag{4.47}$$

where $p_{c,K}$ has been denoted in (4.15).

With these notations and controls at hand we rewrite our initial error as:

$$\begin{aligned}
\mathcal{E}^1(f,(x,y),T,h) &= \mathbb{E}[f(X_T^{h,0,(x,y)},Y_T^{h,0,(x,y)})] - \mathbb{E}[f(X_T^{0,x,y},Y_T^{0,x,y})] \\
&= \mathbb{E}[f(X_T^{h,0,(x,y)},Y_T^{h,0,(x,y)})] - \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,h,0,(x,y)},Y_T^{\varepsilon,h,0,(x,y)})] \\
&\quad +\mathbb{E}[f_\varepsilon(X_T^{\varepsilon,h,0,(x,y)},Y_T^{\varepsilon,h,0,(x,y)})] - \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,0,(x,y)},Y_T^{\varepsilon,0,(x,y)})] \\
&\quad +\mathbb{E}[f_\varepsilon(X_T^{\varepsilon,0,(x,y)},Y_T^{\varepsilon,0,(x,y)})] - \mathbb{E}[f(X_T^{0,x,y},Y_T^{0,x,y})] \\
&=: \sum_{k=1}^{3} \mathcal{E}^{1k,\varepsilon}(f,(x,y),T,h), \qquad (4.48)
\end{aligned}$$

where the notation $(X_t^{\varepsilon,t,(x,y)},Y_t^{\varepsilon,t,(x,y)}))$ stands for the solution of the SDE obtained replacing the coefficients in (4.9) with $b_\varepsilon,\sigma_\varepsilon$. The solution exists due to the additional smoothness we assumed w.r.t. (4.9). Let us first control $\mathcal{E}^{11,\varepsilon}(f,(x,y),T,h)$ which we again split into two parts:

$$\begin{aligned}
\mathcal{E}^{11,\varepsilon}(f,(x,y),T,h) &= \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,0,(x,y)},Y_T^{\varepsilon,0,(x,y)})] - \mathbb{E}[f_\varepsilon(X_T^{0,x,y},Y_T^{0,x,y})] \\
&\quad +\mathbb{E}[f_\varepsilon(X_T^{0,x,y},Y_T^{0,x,y})] - \mathbb{E}[f(X_T^{0,x,y},Y_T^{0,x,y})] \\
&=: \Big(\mathcal{E}^{111,\varepsilon}+\mathcal{E}^{112,\varepsilon}\Big)(f,(x,y),T,h).
\end{aligned}$$

Now, from the Gaussian upper-bound for the density, deriving from Theorem 4.4.2 above, and similarly to inequality (4.44) (which does not exploit the boundedness of the considered function), we get:

$$\mathcal{E}^{112,\varepsilon} \le C\int_{\mathbb{R}^{2d}} p_{c,K}(T-t,(x,y),(x',y'))|(f_\varepsilon - f)(x',y')|dx'dy' \le C\varepsilon^\beta.$$

On the other hand, the stability result (4.46) yields:

$$\begin{aligned}
\mathcal{E}^{111,\varepsilon} &\le C_\eta\varepsilon^\gamma\int_{\mathbb{R}^{2d}} p_{c,K}(T-t,(x,y),(x',y'))|f_\varepsilon(x',y')|dx'dy' \\
&\le C_\eta\int_{\mathbb{R}^{2d}}\Bigg(p_{c,K}(T-t,(x,y),(x',y'))|f_\varepsilon(x',y') - f_\varepsilon(x'-x,y'-y-x(T-t))| \\
&+ p_{c,K}(T-t,(x,y),(x',y'))||f_\varepsilon(x'-x,y'-y-x(T-t))|dx'dy' \\
&\le C_\eta\varepsilon^\gamma(1+|R_{T-t}\begin{pmatrix} x \\ y\end{pmatrix}|^\beta),
\end{aligned}$$

where $R_u := \begin{pmatrix} I_d & 0_d \\ (T-t)I_d & I_d\end{pmatrix}$, exploiting as well the Young inequality for the last control. This finally gives:

$$\mathcal{E}^{11,\varepsilon} \underset{\varepsilon\to 0}{\longrightarrow} 0. \qquad (4.49)$$

The sensitivity result (4.47) for the scheme would yield similarly:

$$\mathcal{E}^{13,\varepsilon} \underset{\varepsilon\to 0}{\longrightarrow} 0. \tag{4.50}$$

We now focus on the contribution $\mathcal{E}^{12,\varepsilon}$ for which we can rely on a PDE type analysis technique. Such an approach had first been used in the context of non-degenerate Hölder continuous Euler schemes by Mikulevičius and Platen [MP91] through Schauder estimates. This in particularly required the final test function to be smooth (specifically $f \in C^{2+\gamma}$ for γ-Hölder continuous coefficients b, σ). This approach was extended in [KM17] using direct control bounds on the heat-kernel allowing that way to consider only β-Hölder continuous test functions $\beta \in (0, 1]$.

The point here is that, through the regularization we are able to use pointwise bounds of the derivatives of the function

$$v_\varepsilon(t, x, y) := \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,t,(x,y)}, Y_T^{\varepsilon,t,(x,y)})] = \int_{\mathbb{R}^{2d}} p_\varepsilon(T - t, (x, y), (x', y'))f_\varepsilon(x', y')dx'dy'$$

and to control as well pointwise and **uniformly in ε small enough**, under **($\hat{\mathbf{A}}$)**, the spatial derivatives of v_ε w.r.t. the non-degenerate component.

Observe that, since $f_\varepsilon, b_\varepsilon, \sigma_\varepsilon$ are smooth, it is readily seen, from the smoothness of v_ε and the Markov property (see e.g. [TT90]), that v_ε satisfies the PDE

$$\begin{cases} (\partial_t v_\varepsilon + L_\varepsilon v_\varepsilon)(t, x, y) = 0, \\ v_\varepsilon(T, x, y) = f_\varepsilon(x, y),\ (x, y) \in \mathbb{R}^{2d}, \end{cases}$$

where L_ε stands for the generator associated with SDE obtained replacing the coefficients in (4.9) with $b_\varepsilon, \sigma_\varepsilon$, i.e. for all $\varphi \in C_0^2(\mathbb{R}^{2d}, \mathbb{R}), (x, y) \in \mathbb{R}^{2d}$,

$$L_\varepsilon\varphi(x, y) = b_\varepsilon(x, y) \cdot \nabla_x\varphi(x, y) + x\nabla_y\varphi(x, y) + \frac{1}{2}\mathrm{Tr}(a_\varepsilon(x, y)D_x^2\varphi(x, y)).$$

For the further analysis we have to apply the Ito formula directly to the scheme

(4.23) viewed as an Ito process exploiting Hölder continuity of coefficients.

$$
\begin{aligned}
\mathcal{E}^{12,\varepsilon} &= \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,h,0,(x,y)}, Y_T^{\varepsilon,h,0,(x,y)})] - \mathbb{E}[f_\varepsilon(X_T^{\varepsilon,0,(x,y)}, Y_T^{\varepsilon,0,(x,y)})] \\
&= \sum_{i=0}^{N-1} \mathbb{E}[v_\varepsilon(t_{i+1}, X_{t_{i+1}}^{\varepsilon,h,0,(x,y)}, Y_{t_{i+1}}^{\varepsilon,h,0,(x,y)}) - v_\varepsilon(t_i, X_{t_i}^{\varepsilon,h,0,(x,y)}, Y_{t_i}^{\varepsilon,h,0,(x,y)})] \quad (4.51) \\
&= \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} \Big\{\partial_s v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) + \nabla_x v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) \\
&\times\ b_\varepsilon(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) + X_s^{\varepsilon,h,0,x,y}\nabla_y v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) \\
&+\ \frac{1}{2}\mathrm{Tr}(D_x^2 v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) a(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}))\Big\} ds\Big] \\
&= \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} \Big\{\partial_s v_\varepsilon + L_\varepsilon v_\varepsilon\Big\}(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) ds\Big] \\
&+\ \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} \Big\{\nabla_x v_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) \cdot (b_\varepsilon(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) \\
&-\ b_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})) \\
&+\ \frac{1}{2}\mathrm{Tr}(D_x^2 v_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})(a(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) \\
&-\ a(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})))\Big\} ds\Big] \\
&= \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} \Big\{\nabla_x v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) \cdot (b_\varepsilon(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) \\
&-\ b_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})) \\
&+\ \frac{1}{2}\mathrm{Tr}(D_x^2 v_\varepsilon(s, X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})(a(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) \\
&-\ a(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})))\Big\} ds\Big], \qquad (4.52)
\end{aligned}
$$

exploiting the PDE satisfied by v_ε for the last equality.

To complete the analysis we first need to control $\nabla_x v_\varepsilon(s,x,y)$ and $D_x^2 v_\varepsilon(s,x,y)$ uniformly in $\varepsilon \in (0,1]$. Moreover, we will also exploit the Hölder properties of b_ε and a_ε in order to control differences $b_\varepsilon(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) - b_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})$ and $a_\varepsilon(X_{t_i}^{\varepsilon,h,0,x,y}, Y_{t_i}^{\varepsilon,h,0,x,y}) - a_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y})$.

To achieve bounds for the heat kernel derivatives we refer the reader to Theorem 4.4.2 below.

Theorem 4.4.2. *Under **($\hat{\mathbf{A}}$D)**, for any $t \in [0,T]$ and $|\alpha| \le 2$ there exist $C \ge 1, c \in (0,1]$ such that, for $\varepsilon \in [0,\varepsilon_0]$, for $\varepsilon_0 > 0$ small enough,*

$$|D_x^\alpha p_\varepsilon(t,(x,y),(x',y'))| \le \frac{C}{t^{|\alpha|/2}} p_{c,K}(t,(x,y),(x',y')).$$

Let us postpone the proof to Appendix, Section 4.6.

Remark 4.4.3. The result of the Theorem 4.4.2 is of interest by itself. Up to the best of our knowledge, these are the first pointwise bounds obtained on the derivatives w.r.t. the non-degenerate variables under the sole Hölder continuity assumption for the coefficients in (4.9). They extend to Kolmogorov diffusions the well-known controls derived by Il'in *et al.* [IKO62]. Investigating the quantitative behaviour of the derivatives w.r.t. the degenerate under *minimal* smoothness assumptions remains a very interesting and open problem. Provided the coefficients are Lipschitz, the Konakov and Mammen trick should apply to get the expected control, namely the normalized Kolmogorov density multiplied by an additional singularity of the characteristic order in time, here $(T-t)^{-3/2}$. Finding out the minimal assumption yielding such a bound is rather challenging.

Coming back to the proof of Theorem 4.4.1, we derive from the control in Theorem 4.4.2 that for

$$D_x^\alpha v_\varepsilon(t,x,y) = \int_{\mathbb{R}^d} D_x^\alpha p_\varepsilon(T-t,(x,y),(x',y'))[f_\varepsilon(x',y') - f_\varepsilon(R_{T-t}\begin{pmatrix} x \\ y \end{pmatrix})]dx'dy',$$

the following inequality holds:

$$\begin{aligned} |D_x^\alpha v_\varepsilon(t,x,y)| \le C \int_{\mathbb{R}^{2d}} & \frac{[f_\varepsilon]_{C^{\beta,\beta/2}} p_{c,K}(T-t,(x,y),(x',y'))}{(T-t)^{\alpha/2}} \\ & \times \left\{|x'-x|^\beta + |y'-(y+(T-t)x|^{\beta/2}\right\} dx'dy', \end{aligned} \tag{4.53}$$

where $[f]_{C^{\beta,\beta/2}} := \sup_{(x,y)\neq(x',y')} \frac{|f(x,y)-f(x',y')|}{d^\beta((x,y),(x',y'))}$, $d^\beta((x,y),(x',y')) := |x-x'|^\beta + |y-y'|^{\beta/2}$.

To have the same scale as in the exponent in (4.15) let us rewrite (4.53) taking into account that:

$$\begin{aligned} |x'-x|^\beta + |y'-y-x(T-t)|^{\beta/2} &= (T-t)^{\beta/2}\left(\frac{|x'-x|}{(T-t)^{1/2}}\right)^\beta \\ &+ (T-t)^{\beta/2}\left(\frac{|y'-y-\frac{(x+x')(T-t)}{2} - \frac{(x-x')(T-t)}{2}|}{(T-t)}\right)^{\beta/2} \\ &\le C(T-t)^{\beta/2}\left[\left(\frac{|x'-x|}{(T-t)^{1/2}}\right)^\beta + \left(\frac{|y'-y-\frac{(x+x')(T-t)}{2}|}{(T-t)^{3/2}}\right)^{\beta/2}\right] \end{aligned} \tag{4.54}$$

From (4.15) and (4.54) one can get, up to a modification of the constant c to $\bar{c}$:

$$\begin{aligned} |D_x^\alpha v_\varepsilon(t,x,y)| &\le \int_{\mathbb{R}^{2d}} \frac{C[f_\varepsilon]_{C^{\beta,\beta/2}} p_{\bar{c},K}(T-t,(x,y),(x',y'))}{(T-t)^{\alpha/2-\beta/2}} dx'dy' \\ &\le \frac{C[f_\varepsilon]_{C^{\beta,\beta/2}}}{(T-t)^{|\alpha|/2-\beta/2}}. \end{aligned} \tag{4.55}$$

Plugging (4.55) into (4.52) we get:

$$|\mathcal{E}^{12,\varepsilon}| \leq C[f_\varepsilon]_{C^{\beta,\beta/2}} \int_0^T ds \Bigg\{ \mathbb{E}\left[|b_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) - b_\varepsilon(X_{\phi(s)}^{\varepsilon,h,0,x,y}, Y_{\phi(s)}^{\varepsilon,h,0,x,y})|\right] \frac{1}{(T-s)^{1/2-\beta/2}}$$

$$+ \mathbb{E}\left[|a_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}) - a_\varepsilon(X_{\phi(s)}^{\varepsilon,h,0,x,y}, Y_{\phi(s)}^{\varepsilon,h,0,x,y})|\right] \frac{1}{(T-s)^{1-\beta/2}} \Bigg\}, \tag{4.56}$$

where $\phi(s) = t_i$ for $s \in [t_i, t_{i+1}), i = 0, \ldots, N-1$.

Let us denote $\Psi_\varepsilon(x,y)$ for any of two functions $b_\varepsilon(x,y)$ or $a_\varepsilon(x,y)$ cause they are both satisfies the same Hölder continuity assumptions. Following introduced notations, we are tempted to bound

$$\mathbb{E}\left[|\Psi_\varepsilon(X_s^{\varepsilon,h,0,x,y}, Y_s^{\varepsilon,h,0,x,y}0) - \Psi_\varepsilon(X_{\phi(s)}^{\varepsilon,h,0,x,y}, Y_{\phi(s)}^{\varepsilon,h,0,x,y})|\right]$$

$$\leq [\Psi_\varepsilon]_{C^{\gamma,\gamma/2}} \left\{ \mathbb{E}\left[|(X_s^{\varepsilon,h,0,x,y} - X_{\phi(s)}^{\varepsilon,h,0,x,y}|^\gamma\right] + \mathbb{E}\left[|Y_s^{\varepsilon,h,0,x,y} - Y_{\phi(s)}^{\varepsilon,h,0,x,y}|^{\gamma/2}\right] \right\}$$

From the definition (4.23) of the approximation scheme:

$$\mathbb{E}\left[|(X_s^{\varepsilon,h,0,x,y} - X_{\phi(s)}^{\varepsilon,h,0,x,y}|^\gamma\right]$$

$$= \mathbb{E}\left[\left|b_\varepsilon(X_{\phi(s)}^{\varepsilon,h,0,x,y}, Y_{\phi(s)}^{\varepsilon,h,0,x,y})(s-\phi(s)) + \sigma(X_{\phi(s)}^{\varepsilon,h,0,x,y}, Y_{\phi(s)}^{\varepsilon,h,0,x,y})(W_s - W_{\phi(s)})\right|^\gamma\right]$$

$$\leq C(|b_\varepsilon|_\infty \vee |\sigma_\varepsilon|_\infty) h^{\gamma/2}. \tag{4.57}$$

To control the error in the second component we cannot compensate the transport so we have to keep the dependency on the starting point in the final bound:

$$\mathbb{E}\left[|(Y_s^{\varepsilon,h,0,x,y} - Y_{\phi(s)}^{\varepsilon,h,0,x,y}|^{\gamma/2}\right] = \mathbb{E}\left[\left|\int_{\phi(s)}^s X_u^{\varepsilon,h,0,x,y} du\right|^{\gamma/2}\right]$$

$$\leq \sup_{s\in[0,T]} \mathbb{E}\left[\left|X_s^{\varepsilon,h,0,x,y}\right|^{\gamma/2}\right] h^{\gamma/2} \leq C h^{\gamma/2} |x|^{\gamma/2}. \tag{4.58}$$

Applying controls in (4.57) and (4.58) for (4.56) we have a final control for the $\mathcal{E}^{12,\varepsilon}$:

$$|\mathcal{E}^{12,\varepsilon}| \leq C(a, b, f_\varepsilon, T) h^{\gamma/2}(1 + |x|^{\gamma/2}). \tag{4.59}$$

which proves the statement of the Theorem 4.4.2.

□

Remark 4.4.4 (About the convergence order). We want to stress that in the mild smoothness setting we consider, the convergence rate appearing in (4.59) is similar to the one appearing when considering the strong error. Indeed, since we cannot hope to go beyond the expansion of order 2 for the PDE satisfied by v_ε (at least with controls uniform in ε) we are led to compare the increments of the Euler scheme (4.23) at the power corresponding to the Hölder exponent.

Remark 4.4.5. The rate $h^{\gamma/2}(1+|x|^{\gamma/2})$ holds even for tests functions $f \in C^{\beta_1,\beta_2}(\mathbb{R}^{2d})$, $(\beta_1,\beta_2) \in (0,1]^2$. The idea is to suitably handle the last steps according to the β_1,β_2. We refer to [KM17] for the non-degenerate case.

Our second result, already mentioned in the Introduction (Chapter 1), provides bounds on the difference between the densities $\mathcal{E}_d$.

4.5 Global error

Theorem 4.5.1. *Fix a final time horizon $T>0$ and a time step $h = T/N, N \in \mathbb{N}^*$ for the Euler scheme. Under assumptions **($\hat{\mathbf{A}}$D)**, for $\gamma \in (1/2,1]$ and $\beta \in (0,\gamma-\frac{1}{2})$, for all t in the time grid $\Lambda_h := \{(t_i)_{i\in[\![1,N]\!]}\}$ and $(x,y),(x',y') \in \mathbb{R}^{2d}$ there exist $C := (T,b,a,\beta), c>0$ such that :*

$$\begin{aligned} &|p(t,(x,y),(x',y')) - p_h(t,(x,y),(x',y')| \\ &\le Ch^\beta(1+(|x|\wedge|x'|))^{1+\gamma}) \sup_{s\in[t-h,t]} p_{c,K}(s,(x,y),(x',y')), \end{aligned} \tag{4.60}$$

where as in (4.15) $p_{c,K}(s,(x,y),(x',y'))$ *stands for the Kolmogorov-type gaussian density at time s.*

The proof is given below in Section 4.5.1. The above result is in clear contrast with the one of Theorem 4.4.1 for the weak error, i.e. when additionally consider an integration of a Hölder function w.r.t. the final (or forward variable). The point is that such an integration allows to exploit directly the spatial bounds of Theorem 4.4.2 on the underlying heat-kernel (with possibly mollified coefficients). When handling directly the difference of the densities we cannot avoid to control sensitivities of the kernels w.r.t. to the degenerate variable. Such sensitivities lead to higher time singularities and make the unbounded transport term appear. The higher time-singularities yield the stated restriction on the Hölder index γ. The unbounded transport gives the term $|x|\wedge|x'|$ in the above bound. We finally can reach a global error of order $h^\beta, \beta < \gamma - 1/2$ which is *close* to the expected one in $h^{\gamma/2}$ when γ goes to 1.

To improve the above result, we fill that some new advanced approaches to error analysis should be considered. This means that either the scheme would have to be modified or the error decomposed very differently than in the current huge literature (from the seminal papers of [KM00] and [KM02] the same lines are considered for the error decomposition, see e.g. [KM10], [KM17], [Fri18]). Eventually, a specific difficulty of the current model consists in dealing the unbounded transport term.

4.5.1 Proof of Theorem 4.5.1

The basic idea to prove Theorem 4.5.1 consists in applying parametrix expansion to both densities $p(t,(x,y),(x',y')), p_h(t,(x,y),(x',y')$. The convergence of the parametrix series expansion for the solution of (4.9) and the scheme (4.23) follows from Section 4.2 above.

In order to derive bounds for the difference of densities in (4.60), let us introduce for $0 \le j < j' \le N$ $\forall (x,y),(x',y') \in \mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$,

$$p^d(t_j,(x,y),(x',y')) := \sum_{r\in\mathbb{N}} \tilde{p} \otimes_h H^{(r)}(t_j,(x,y),(x',y')). \tag{4.61}$$

From (4.20) and the semigroup property (4.16) it follows that $\tilde{p}\otimes_h H(t_j,(x,y),(x',y')) \le C(b,T,\gamma) t_j^{\gamma/2} B(1,\frac{\gamma}{2}) p_{c,K}(t_j,(x,y),(x',y'))$ which is by induction yields that for all $r \ge 1$, $\forall (x,y),(x',y') \in \mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$,

$$|\tilde{p} \otimes_h H^{(r)}(t_j,(x,y),(x',y'))| \le C^r t_j^{r\gamma/2} \prod_{i=1}^r B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right) p_{c,K}(t_j,(x,y),(x',y'))$$

with $C := C(\lambda,\gamma)(|b|_\infty T^{\frac{1-\gamma}{2}}+1)$.

From the last inequality we readily get that the series in (4.61) converges absolutely and uniformly on $\mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$ and that $\forall (x,y),(x',y') \in \mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$,

$$p^d(t_j,(x,y),(x',y')) \le E_{\gamma/2,1}(C(|b|_\infty T^{1/2}+T^{\gamma/2}) p_{c,K}(t_j,(x,y),(x',y'))). \tag{4.62}$$

As the result we decompose the total global error into two terms:

$$|(p-p_h)(t_i,(x,y),(x',y'))| \le |(p-p^d)(t_i,(x,y),(x',y'))| + |(p^d-p_h)(t_i,(x,y),(x',y'))|.$$

Error bound on $p^d - p_h$ (same discrete convolution)

Remark that for $r \ge 1$ as it can be decomposed with the classical approach from [KM02] (see also [Fri18] and [KM17] for connections with the current Hölder settings).

$$\begin{aligned} &\tilde{p}\otimes_h H^{(r)} - \tilde{p}\otimes_h H_h^{(r)} \\ &= \left(\left(\tilde{p}\otimes_N H^{(r-1)}\right)\otimes_h (H-H_h)\right) + \left(\left(\tilde{p}\otimes_N H^{(r-1)} - \tilde{p}\otimes_h H_h^{(r-1)}\right)\otimes_h H_h\right). \end{aligned}$$

For the sum from $r=1$ to $r=\infty$ it yields:

$$p^d - p_h = p^d \otimes_h (H-H_h) + (p^d - p_h)\otimes_h H_h.$$

By induction, for $0 \le j < j' \le N$ one gets for all $(x,y),(x',y') \in \mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$,

$$(p^d - p_h)(t_j,(x,y),(x',y')) = \sum_{r\geq 0}\{p^d \otimes (H-H_h)\} \otimes_h H^{(r)}(t_j,(x,y),(x',y')). \quad (4.63)$$

As the result, it is sufficient for us to establish the right control for each term in the sum (4.63).

Lemma 4.5.2. *Under assumptions* ***(ÂD)****, for all* $0 \leq j < j' \leq N$*, for all* $(x,y),(x',y') \in \mathbb{R}^{2d}\times\mathbb{R}^{2d,*}$*, one has*

$$|\{p^d \otimes (H-H_h)\} \otimes_h H^{(r)}(t_j,(x,y),(x',y'))| \leq Ch^{\gamma/2} p_{c,K}(t_i,(x,y) \quad (4.64)$$

for some constant $c := c(\lambda,\gamma) \geq 1$ *and a non decreasing positive fuction* $T \to C := C(T,b,\sigma)$.

Proof. First, let us consider the first step separately. For $j=1$ directly from the kernel function definition it follows that:

$$\begin{aligned}&(H-H_h)(t_j,(w,z),(\hat{w},\hat{z}))\\ &= \langle b(\hat{w},\hat{z}) D_x \tilde{p}(t_j,(w,z),(\hat{w},\hat{z}))\rangle + \frac{1}{2}\mathrm{Tr}\left\{(a(w,z)-a(\hat{w},\hat{z}-\hat{w}t_j))D^2_w \tilde{p}(t_j,(w,z)(\hat{w},\hat{z}))\right\}\\ &\quad -h^{-1}(p_h-\tilde{p}_h)(t_j,(w,z),(\hat{w},\hat{z})).\end{aligned}$$

From (4.20) it follows that:

$$\begin{aligned}&|\langle b(\hat{w},\hat{z}) D_x \tilde{p}(t_j,(w,z),(\hat{w},\hat{z}))\rangle + \frac{1}{2}\mathrm{Tr}\left\{(a(w,z)-a(\hat{w},\hat{z}-\hat{w}t_j))D^2_w \tilde{p}(t_j,(w,z)(\hat{w},\hat{z}))\right\}|\\ &\leq C(|b|_\infty t_j^{2-\gamma/2}+1)\frac{1}{t_j^{1-\gamma/2}} p_{c,K}(t_j,(w,z),(\hat{w},\hat{z})).\end{aligned}$$

In [LM10] authors achieved the following control to prove Lemma 4.1(see [LM10], Appendix, A1, proof of Lemma 4.1, the case (b) which absolutely covers our model and assumptions):

$$h^{-1}(p_h-\tilde{p}_h)(t_j,(w,z),(\hat{w},\hat{z})) \leq \frac{C(T,b,\sigma)}{t_j^{1-\gamma/2}} p_{c,K}(t_j,(w,z),(\hat{w},\hat{z})) \text{ for } t_j = h.$$

Combining the last two estimates together one cat get for $t_j = h$, $\forall (w,z)$, $(\hat{w},\hat{z}) \in \mathbb{R}^{2d}\times\mathbb{R}^{2d,*}$:

$$\begin{aligned}|(H-H_h)(t_j,(w,z),(\hat{w},\hat{z}))| &\leq (|H|+|H_h|)(t_j,(w,z),(\hat{w},\hat{z}))\\ &\leq \frac{C}{t_j^{1-\gamma/2}} p_{c,K}(t_j,(w,z),(\hat{w},\hat{z})), \quad (4.65)\end{aligned}$$

where $T \to C := C(T,b,\sigma)$ is a non-decreasing positive function.

Now we do a decomposition: $\forall i \in [2,N]$

$$
\begin{aligned}
& p^d \otimes (H - H_h)(t_i,(x,y),(w,z)) \\
= & \sum_{k=0}^{i-2} h \int_{\mathbb{R}^{2d}} dudv p^d(t_k,(x,y),(u,v))(H-H_h)(t_i - t_k,(u,v),(w,z)) \\
& + h \int_{\mathbb{R}^{2d}} p^d(t_{i-1},(x,y),(u,v))(H-H_h)(h,(u,v),(z,w))dudv.
\end{aligned} \tag{4.66}
$$

From (4.62), (4.65) and the semigroup property (4.16), we derive:

$$
\begin{aligned}
& h\left|\int_{\mathbb{R}^{2d}} p^d(t_{i-1},(x,y),(u,v))(H-H_h)(h,(u,v),(w,z))dudv\right| \\
& \leq \frac{Ch}{h^{1-\gamma/2}} p_{c,K}(t_i,(x,y),(w,z)),
\end{aligned} \tag{4.67}
$$

where $T \to C := C(T,b,\sigma)$ is a positive non-decreasing function.

Let us again mention the paper [LM10]. We would like to emphasize that under previous assumptions for coefficients in our model, according to the paper, there exist a constant $c := c(\lambda,\gamma) > 1$ such that for all $1 < j < j' \leq N$:

$$
|H(t_j,(x,y),(x'y')) - H_h(t_j,(x,y),(x'y'))| \leq \frac{C}{t_j^{1-\gamma/2}} p_{c,K}(t_j,(x,y),(x',y')) \tag{4.68}
$$

where $T \to C = C(T,b,\sigma)$ is a positive non-decreasing function. The case $j=1$ has been already proved in (4.65).

From (4.62), (4.68) and the semigroup property (4.16) one gets:

$$
\begin{aligned}
& \left|\sum_{k=0}^{i-2} h \int_{\mathbb{R}^{2d}} dudv p^d(t_k,(x,y),(u,v))(H-H_h)(t_i-t_k,(u,v),(w,z))\right| \\
& \leq C h^{\gamma/2} p_{c,K}(t_i,(x,y),(w,z)).
\end{aligned}
$$

Due to all the previous estimates, we derive

$$
\forall i \in [2,N], |p^d \otimes_h (H-H_h)(t_i,(x,y),(w,z))| \leq C h^{\gamma/2} p_{c,K}(t_i,(x,y),(w,z)),
$$

where $T \to C := C(T,b,\sigma)$ is a non-decreasing positive function and (4.64) follows by induction. □

From Lemma 4.5.2, we obtain $\forall (x,y),(x',y') \in \mathbb{R}^{2d} \times \mathbb{R}^{2d,*}$:

$$
|(p^d - p_h)(t_i,(x,y),(x',y'))| \leq C(T,b,\sigma) h^{\gamma/2} p_{c,K}(t_i,(x,y),(x',y')). \tag{4.69}
$$

Remark 4.5.1. We would like to emphasize that up to now the bound we have in (4.69) is much better which has been stated in the Theorem 4.5.1. This term can be controlled better since we do not feel the explosion which comes for $p - p^d$ when we basically have to investigate the difference between the time integral and the Riemann sums.

Error bound on $p - p^d$

It still remains to control the difference $p - p^d$. For $r \geq 1$, we write the decomposition according to the same iteration procedure as in [KM02],

$$\tilde{p} \otimes H^{(r)} - \tilde{p} \otimes_h H^{(r)} = \left[\left(\tilde{p} \otimes H^{(r-1)}\right) \otimes H - \left(\tilde{p} \otimes H^{(r-1)}\right) \otimes_h H\right] + \left[\left(\tilde{p} \otimes H^{(r-1)}\right) - \left(\tilde{p} \otimes H^{(r-1)}\right)\right] \otimes_h H.$$

Summing up from $r = 1$ to ∞ we get

$$p - p^d = p \otimes H - p \otimes_h H + (p - p^d) \otimes_h H.$$

As in the paper [KM17]:

$$\begin{aligned}(p - p^d)(t_j, (x, y), (x', y')) &= (p \otimes H - p \otimes_h H)(t_j, (x, y), (x', y')) \\ &\quad + (p - p^d) \otimes_h H(t_j, (x, y), (x', y')) \\ &= \sum_{r \geq 0} (p \otimes H - p \otimes_h H) \otimes_h H^{(r)}(t_j, (x, y), (x', y')),\end{aligned} \tag{4.70}$$

The key point is thus to control $|p \otimes H - p \otimes_h H|$.
For that purpose let us write:

$$\begin{aligned} & (p \otimes H - p \otimes_h H)(t_j, (x, y), (x', y')) \\ = \ & \sum_{k=0}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \{p(u, (x, y), (w, z))H(t_j - u, (w, z), (x', y')) \\ - \ & p(t_k, (x, y), (w, z))H(t_j - t_k, (w, z), (x', y'))\} dwdz \\ = \ & \sum_{k=0}^{j-1} \Big\{ \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \{[p(u, (x, y), (w, z)) - p(t_k, (x, y), (w, z))] \\ \times \ & H(t_j - u, (w, z), (x', y'))\} dwdz \Big\} \\ + \ & \sum_{k=0}^{j-1} \Big\{ \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \{p(t_k, (x, y), (w, z)) \end{aligned}$$

$$\times \quad [H(t_j-u,(w,z),(x',y'))-H(t_j-t_k,(w,z),(x',y'))]\}dwdz\Big\}$$
$$=: \quad (D^{d,1}+D^{d,2})(t_j,(x,y),(x',y')). \tag{4.71}$$

- Bounds for the term $D^{d,1}$.

\- For $k=0$, one readily gets:

$$\left|\int_0^h du \int_{\mathbb{R}^{2d}} \{[p(u,(x,y),(w,z))-p(0,(x,y),(w,z))]H(t_j-u,(w,z),(x',y'))\}dwdz\right|$$
$$\leq \quad Cp_{c,K}(t_j,(x,y),(x',y'))\int_0^h \frac{du}{(t_j-u)^{1-\gamma/2}} \leq \frac{Ch}{(t_j)^{1-\gamma/2}}p_{c,K}(t_j,(x,y),(x',y'))$$
$$\leq \quad Ch^{\gamma/2}p_{c,K}(t_j,(x,y),(x',y')). \tag{4.72}$$

\- For $k\in[1,j-1]$ we are interested to control the sum:

$$\sum_{k=1}^{j-1}\int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \{p(u,(x,y),(w,z))-p(t_k,(x,y),(w,z))\}H(t_j-u,(w,z),(x',y'))dwdz. \tag{4.73}$$

To proceed with the case for $k\geq 1$ one needs the following result:

Lemma 4.5.3. *Under **($\hat{\mathbf{A}}\mathbf{D}$)** there exist constants $C(\lambda,\gamma), c:=c(\lambda,\gamma)\geq 1$ such that for all $t\in[0,T]$ for all $r\geq 0$, for all $(x,y),(x',y')\in\mathbb{R}^{2d}\times\mathbb{R}^{2d,*}$ and $0<s\leq t\leq T$ one has*

$$|\tilde{p}\otimes H^{(r)}(t_k+(u-t_k),(x,y),(x',y'))-\tilde{p}\otimes H^{(r)}(t_k,(x,y),(x',y'))|$$
$$\leq E_{\gamma/2,1}(1+|x|^{1+\gamma/2})C^{r+1}t_k^{r\gamma/2}\left\{\frac{(u-t_k)^{\gamma/2}}{t_k}+\frac{(u-t_k)}{t_k^{3/2}}\right\}$$
$$\times\prod_{i=1}^{r-1}B\Big(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\Big)\int_0^1 d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(x',y')) \tag{4.74}$$

from which it follows that

$$|p(t_k+(u-t_k),(x,y),(x',y'))-p(t_k,(x,y),(x',y'))|$$
$$\leq \quad C(b,T)E_{\gamma/2,1}(1+|x|^{1+\gamma/2})\left\{\frac{(u-t_k)^{\gamma/2}}{t_k}+\frac{(u-t_k)}{t_k^{3/2}}\right\}$$
$$\times \quad \int_0^1 d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(x',y')). \tag{4.75}$$

Proof. Let us start with the base for the induction. To control the difference between frozen densities at the step $r=0$ one can write, applying the mean-value theorem and the Kolmogorov equation, takeing $s=u-t_k$ for a moment:

$$\begin{aligned}
& \tilde{p}(t_k+s,(x,y),(x',y'))-\tilde{p}(t_k,(x,y),(x',y')) \\
= & \ s\int_0^1 \partial_\tau \tilde{p}(t_k+\lambda s,(x,y),(x',y'))|_{\tau=t_k+\lambda s} d\lambda \\
= & \ s\int_0^1 \frac{1}{2}\mathrm{Tr}(a(x',y'-x'(t_k+\lambda s))D_x^2\tilde{p}(t_k+\lambda s,(x,y),(x',y'))) \\
+ & \ \langle x, \nabla_y \tilde{p}(t_k+\lambda s,(x,y),(x',y'))\rangle \\
\leq & \ Cs\int_0^1 \left[\frac{1}{(t_k+\lambda s)}+\frac{|x|}{(t_k+\lambda s)^{3/2}}\right]\tilde{p}(t_k+\lambda s,(x,y),(x',y')) \\
\leq & \ Cs\left(\frac{1}{t_k}+\frac{|x|}{t_k^{3/2}}\right)\int_0^1 d\lambda \tilde{p}(t_k+\lambda s,(x,y),(x',y')).
\end{aligned} \tag{4.76}$$

So (4.74) is valid for $r=0$. Now proceeding by induction we assume that (4.74) is valid for $r\geq 0$. By a change of variables one has:

$$\begin{aligned}
& \tilde{p}\otimes H^{(r+1)}(t_k+s,(x,y),(x',y'))-\tilde{p}\otimes H^{(r+1)}(t_k,(x,y),(x',y')) \\
&= \int_0^{t_k+s} d\tau \int_{\mathbb{R}^{2d}} \tilde{p}\otimes H^{(r)}(\tau,(x,y),(w,z))H(t+s-\tau,(w,z),(x',y'))dwdz \\
&\quad - \int_0^{t_k} d\tau \int_{\mathbb{R}^{2d}} \tilde{p}\otimes H^{(r)}(\tau,(x,y),(w,z))H(t_k-\tau,(w,z),(x',y'))dwdz \\
&= \int_{t_k}^{t_k+s} d\tau \int_{\mathbb{R}^{2d}} \tilde{p}\otimes H^{(r)}(t_k+s-\tau,(x,y),(w,z))H(\tau,(w,z),(x',y'))dwdz \\
&\quad + \int_0^{t_k} d\tau \int_{\mathbb{R}^{2d}} \{\tilde{p}\otimes H^{(r)}(t_k+s-\tau,(x,y),(w,z)) \\
&\quad -\tilde{p}\otimes H^{(r)}(t_k-\tau,(x,y),(w,z))\}H(\tau,(w,z),(x',y'))dwdz \\
&= I+J.
\end{aligned}$$

From (4.20) and Lemma 4.2.1 one can get

$$\begin{aligned}
|I| &\leq \frac{C^{r+2}}{t_k^{1-\gamma/2}}\left(\int_{t_k}^{t_k+s}(t_k+s-\tau)^{r\gamma/2}d\tau\right)\prod_{i=1}^{r} B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)p_{c,K}(t_k,(x,y),(x',y')) \\
&\leq \frac{s}{t_k}C^{r+2}t_k^{\frac{(r+1)\gamma/2}{2}}\prod_{i=1}^{r} B\left(1+\frac{(i-1)\gamma}{2},\frac{\eta}{2}\right))p_{c,K}(t_k,(x,y),(x',y')),
\end{aligned}$$

where we used that $s=u-t_k\in[0,h]$ for the last inequality.

For the second term to full fill the induction assumption in time let us decompose $J = J_1 + J_2$, where:

$$\begin{aligned} J_1 &= \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} \{\tilde{p} \otimes H^{(r)}(t_k + s - \tau, (x,y), (w,z)) \\ &\quad - \tilde{p} \otimes H^{(r)}(t_k - \tau, (x,y), (w,z))\} H(\tau, (w,z), (x',y')) dwdz \\ J_2 &= \int_{t_k/2}^{t_k} d\tau \int_{\mathbb{R}^{2d}} \{\tilde{p} \otimes H^{(r)}(t_k + s - \tau, (x,y), (w,z)) \\ &\quad - \tilde{p} \otimes H^{(r)}(t_k - \tau, (x,y), (w,z))\} H(\tau, (w,z), (x',y')) dwdz. \end{aligned}$$

First, assume that $s = u - t_k \in [0, t_k/2]$. That means $s \in [0, t_k - u]$ for all $u \in [0, t_k/2]$ so that we can apply the induction hypothesis and get:

$$\begin{aligned} |J_1| &\leq C^{r+2} s(1 + |x|^{1+\gamma/2}) \left[\frac{1}{t_k} + \frac{1}{t_k^{3/2}}\right] \left(\int_0^{t_k/2} (t_k - \tau)^{r\gamma/2} \tau^{-(1-\gamma/2)} d\tau\right) \\ &\quad \times \left(\prod_{i=1}^{r-1} B\left(1 + \frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right)\right) \int_0^1 d\lambda p_{c,K}(t_k + \lambda s, (x,y), (x',y')) \\ &\leq C^{r+2} s(1 + |x|^{1+\gamma/2}) \left[\frac{1}{t_k} + \frac{1}{t_k^{3/2}}\right] t_k^{(r+1)\gamma/2} \prod_{i=1}^{r} B(1 + \frac{(i-1)\gamma}{2}, \frac{\gamma}{2}) \\ &\quad \times \int_0^1 d\lambda p_{c,K}(t_k + \lambda s, (x,y), (x',y')). \end{aligned}$$

Now, if $s \in (t_k/2, t_k]$ one writes $J_1 = J_1^1 + J_1^2$ with:

$$\begin{aligned} J_1^1 &= \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} \left\{\tilde{p} \otimes H^{(r)}(t_k - \tau + \frac{t_k}{2} + (s - \frac{t_k}{2}), (x,y), (w,z))\right. \\ &- \left.\tilde{p} \otimes H^{(r)}(t - \tau + \frac{t_k}{2}, (x,y), (w,z))\right\} H(\tau, (w,z), (x',y')) dwdz, \\ J_1^2 &= \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} \left\{\tilde{p} \otimes H^{(r)}(t_k - \tau + \frac{t_k}{2}, (x,y), (w,z))\right. \\ &- \left.\tilde{p} \otimes H^{(r)}(t_k - \tau, (x,y), (w,z))\right\} H(\tau, (w,z), (x',y')) dwdz. \end{aligned}$$

From the induction hypothesis and (4.20) using that $t_k/2 \leq t_k - \tau$ for $\tau \in [0, t_k/2]$

one has:

$$
\begin{aligned}
|J_1^1| &\leq C^{r+2}(1+|x|^{1+\gamma/2})\int_0^{t_k/2}(s-\frac{t_k}{2})\left\{\frac{1}{\frac{t_k}{2}+(t_k-\tau)}+\frac{1}{(\frac{t_k}{2}+(t_k-\tau))^{3/2}}\right\}\\
&\quad\times(t_k-\tau+\frac{t_k}{2})^{r\gamma/2}\tau^{-(1-\gamma/2)}d\tau\left(\prod_{i=1}^{r}B(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2})\right)\\
&\quad\times\int_0^1 d\lambda p_{c,K}(t_k+\lambda s,(x,y),(x',y'))\\
&\leq C^{r+2}(1+|x|^{1+\gamma/2})s\left[\frac{1}{t_k}+\frac{1}{t_k^{3/2}}\right]t_k^{(r+1)\gamma/2}\prod_{i=1}^{r}B(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2})\\
&\quad\times\int_0^1 d\lambda p_{c,K}(t_k+\lambda s,(x,y),(x',y')).
\end{aligned}
$$

Using the similar argument with $s\geq t_k/2$:

$$
\begin{aligned}
|J_1^2| &\leq (1+|x|^{1+\gamma/2})s\left[\frac{1}{t_k}+\frac{1}{t_k^{3/2}}\right]C^{r+2}t_k^{(r+1)\gamma/2}\prod_{i=1}^{r}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)\\
&\quad\times\int_0^1 d\lambda p_{c,K}(t_k+\lambda s,(x,y),(x',y')),
\end{aligned}
$$

which yields

$$
\begin{aligned}
|J_1| &\leq (1+|x|^{1+\gamma/2})s\left[\frac{1}{t_k}+\frac{1}{t_k^{3/2}}\right]C^{r+2}t_k^{(r+1)\gamma/2}\prod_{i=1}^{r}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)\\
&\quad\times\int_0^1 d\lambda p_{c,K}(t_k+\lambda s,(x,y),(x',y'))
\end{aligned}
$$

The last term J_2 is given by the sum of three terms:

$$
\begin{aligned}
J_2^1 &= -\int_0^s d\tau\int_{\mathbb{R}^{2d}}\tilde{p}\otimes H^{(r)}(\tau,(x,y),(w,z))H(t_k-\tau+s,(w,z),(x',y'))dwdz,\\
J_2^2 &= \int_{t_k/2}^{t_k/2+s}du\int_{\mathbb{R}^{2d}}\tilde{p}\otimes H^{(r)}(\tau,(x,y),(w,z))H(t_k-\tau+s,(w,z),(x',y'))dwdz,\\
J_2^3 &= \int_0^{t_k/2}d\tau\int_{\mathbb{R}^{2d}}\tilde{p}\otimes H^{(r)}(\tau,(x,y),(w,z))\\
&\times\ \{H(t_k-\tau+s,(w,z),(x',y'))-H(t_k-\tau,(w,z),(x',y'))\}dwdz.
\end{aligned}
$$

Using (4.20) and (4.2.1) one can get as usual:

$$\begin{aligned}|J_2^1| &\le C^{r+2}\left(\int_0^s \tau^{r\gamma/2}\frac{1}{(t_k+s-\tau)^{1-\frac{\gamma}{2}}}d\tau\right)\prod_{i=1}^{r-1}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)p_{c,K}(t_k,(x,y),(x',y'))\\ &\le \frac{s}{t_k}C^{r+2}t_k^{\frac{(r+1)\gamma}{2}}\prod_{i=1}^{r}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)p_{c,K}(t_k,(x,y),(x',y'))\end{aligned}$$

and similarly

$$\begin{aligned}|J_2^2| &\le C^{r+2}\frac{1}{t_k^{1-\gamma/2}}\left(\int_{t_k/2}^{t_k/2+s}\tau^{r\gamma/2}d\tau\right)\prod_{i=1}^{r}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)p_{c,K}(t_k,(x,y),(x',y'))\\ &\le \frac{s}{t_k}C^{r+2}t_k^{\frac{(r+1)\gamma}{2}}\prod_{i=1}^{r}B\left(1+\frac{(i-1)\gamma}{2},\frac{\gamma}{2}\right)p_{c,K}(t_k,(x,y),(x',y')).\end{aligned}$$

To control $|J_2^3|$ we need to derive bounds for the kernel time sensitivity $H(t-\tau+s,(w,z),(x',y'))-H(t-\tau,(w,z),(x',y'))$.

Lemma 4.5.4.

$$\begin{aligned}&|H(t_k-\tau+s,(w,z),(x',y'))-H(t_k-\tau,(w,z),(x',y'))|\\ &\le C\left\{\frac{s}{(t-\tau)^{2-\gamma/2}}+\left(\frac{s^{\gamma/2}|x'|^{\gamma/2}}{(t_k-\tau)}+\frac{s|x'|^{\gamma/2}}{(t_k-\tau)^{2-\gamma/2}}+\frac{s|w|(1+|x'|^{\gamma/2})}{(t_k-\tau)^{5/2-\gamma/2}}\right)\right\}\\ &\times\int_0^1 d\lambda p_{c,K}(t_k-\tau+s\lambda,(w,z),(x',y')).\end{aligned} \tag{4.77}$$

Proof. According to the definition (4.18):

$$\begin{aligned}&H(t_k-\tau+s,(w,z),(x',y'))-H(t_k-\tau,(w,z),(x',y'))\\ =\ & b(w,z)D_w\tilde{p}(t_k-\tau+s,(w,z),(x',y'))-b(w,z)D_w\tilde{p}(t_k-\tau,(w,z),(x',y'))\\ +\ & \frac{1}{2}\mathrm{Tr}\left(\left(a(w,z)-a(x',y'-(t_k-\tau+s)x')\right)D_w^2\tilde{p}(t_k-\tau+s,(w,z),(x',y'))\right)\\ -\ & \frac{1}{2}\mathrm{Tr}\left(\left(a(w,z)-a(x',y'-(t_k-\tau)x')\right)D_w^2\tilde{p}(t_k-\tau,(w,z),(x',y'))\right).\end{aligned}$$

Let us estimate the most singular term. Others can be handled similarly.

$$\begin{aligned}
\Big(a(w,z) &- a(x',y'-(t_k-\tau+s)x')\Big) D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))\\
-\Big(a(w,z) &- a(x',y'-(t_k-\tau)x')\Big) D_w^2\tilde p(t_k-\tau,(w,z),(x',y'))\\
=\Big(a(x',y' &- (t_k-\tau)x')-a(x',y'-(t_k-\tau+s)x')\Big) D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))\\
+\Big(a(w,z) &- a(x',y'-(t_k-\tau)x')\Big)\Big(D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))\\
-D_w^2\tilde p(t_k &- \tau+s,(w,z),(x',y'))\Big) := \Delta H_1+\Delta H_2.
\end{aligned}$$

The term ΔH_1 is easier to control using just Hölder property of a and the standard estimation for the derivative of the frozen density $D_w^2\tilde p$, see (4.17).

$$\begin{aligned}
&|\Delta H_1| = |a(x',y'-(t_k-\tau)x')-a(x',y'-(t_k-\tau+s)x')|\\
&\times|D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))| \le |s|^{\gamma/2}|x'|^{\gamma/2}\frac{Cp_{c,K}(t_k-\tau+s,(w,z),(x',y')}{t_k-\tau+s} \qquad (4.78)
\end{aligned}$$

The control for $|\Delta H_2|$ is more involved. We have to derive the sensitivity of frozen density derivatives with respect to the time-variable, actually, to bound $|D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))-D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))|$, namely:

$$\begin{aligned}
&|D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))-D_w^2\tilde p(t_k-\tau+s,(w,z),(x',y'))| =\\
&\qquad \le C|s|\left|\int_0^1 d\lambda \partial_v D_w^2\tilde p(\tau,(w,z),(x',y')|_{v=t_k-\tau+\lambda s}\right|\\
&\le C|s|\int_0^1\Big\{d\lambda|a(x',y'-x'(t_k-\tau+\lambda s))||D_w^4\tilde p(t_k-\tau+\lambda s,(w,z),(x',y'))|\\
&\qquad +2|D_wD_z\tilde p(t_k-\tau+\lambda s,(w,z),(x',y'))|\\
&\qquad +|w||D_w^2\tilde p(t_k-\tau+\lambda s,(w,z),(x',y'))|\Big\} \qquad (4.79)
\end{aligned}$$

using the Kolmogorov equation, applied to the frozen density for the last step.

Thus, due to the control in (4.17), we get the following bounds for $|\Delta H_2|$:

$$\begin{aligned}
|\Delta H_2| \le C|s|&\left\{\frac{1}{(t_k-\tau)^{2-\gamma/2}}+\frac{|w|}{(t_k-\tau)^{5/2-\gamma/2}}\right\}(1+|x'|^{\gamma/2})\\
&\times\int_0^1 d\lambda p_{c,K}(t_k-\tau+\lambda s,(w,z),(x',y')). \qquad (4.80)
\end{aligned}$$

Together (4.78) and (4.80) provide the statement of the Lemma 4.5.4. □

Thus, from Lemma 4.5.4 it yields the final control:

$$
\begin{aligned}
|J_2^3| \quad \leq \quad & \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} C^{r+1}\tau^{r\gamma/2} \prod_{i=1}^{r-1} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) p_{c,K}(\tau,(x,y),(w,z)) \\
\times \quad & (1+|x'|^{1+\gamma/2}) \left\{ \frac{s}{(t_k-\tau)^{2-\gamma/2}} + \frac{s^{\gamma/2}}{(t_k-\tau)} + \frac{s}{(t_k-\tau)^{5/2-\gamma/2}} \right\} \\
\times \quad & \int_0^1 d\lambda p_{c,K}(t_k-\tau+s\lambda,(w,z),(x',y'))dwdz. \qquad (4.81)
\end{aligned}
$$

The term $\frac{s}{(t_k-\tau)^{2-\gamma/2}}$ in the convolution directly leads to prove the induction hypothesis :

$$
\begin{aligned}
& \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} C^{r+1}\tau^{r\gamma/2} \prod_{i=1}^{r-1} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) p_{c,K}(\tau,(x,y),(w,z)) \\
\times \quad & \left\{ \frac{s}{(t_k-\tau)^{2-\gamma/2}} \right\} \int_0^1 d\lambda p_{c,K}(t_k-\tau+s\lambda,(w,z),(x',y'))dwdz \\
\leq \quad & \frac{s}{t_k} C^{r+2} t_k^{\frac{(r+1)\gamma}{2}} \prod_{i=1}^{r} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) \int_0^1 d\lambda p_{c,K}(t_k+s\lambda,(x,y),(x',y')), \qquad (4.82)
\end{aligned}
$$

thus, we have to concentrate on the other two terms $\frac{s^{\gamma/2}}{(t_k-\tau)} + \frac{s}{(t_k-\tau)^{5/2-\gamma/2}}$ in (4.81) and find out which term dominates on the current interval.

Since on the interval we are considering $t_k-\tau > t_k/2 > s$ it is true that: $\frac{s^{\gamma/2}}{(t_k-\tau)} > \frac{s}{(t_k-\tau)^{2-\gamma/2}}$.

The first term again lead us to the standard computations as in (4.82) and the second term can be bounded with $\frac{s^{\gamma/2}}{(t_k-\tau)}$.

Finally,

$$
\begin{aligned}
& \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} C^{r+1}(1+|x'|^{1+\gamma/2})\tau^{r\gamma/2} \prod_{i=1}^{r-1} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) p_{c,K}(\tau,(x,y),(w,z)) \\
\times \quad & \left(\frac{s^{\gamma/2}}{(t_k-\tau)} + \frac{s}{(t_k-\tau)^{2-\gamma/2}} + \frac{s}{(t_k-\tau)^{5/2-\gamma/2}} \right) \int_0^1 d\lambda p_{c,K}(t_k-\tau+s\lambda,(w,z),(x',y'))dwdz \\
\leq \quad & \int_0^{t_k/2} d\tau \int_{\mathbb{R}^{2d}} C^{r+1}\tau^{r\gamma/2} \prod_{i=1}^{r-1} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) p_{c,K}(\tau,(x,y),(w,z)) \frac{s^{\gamma/2}(1+|x'|^{1+\gamma/2})}{(t_k-\tau)} \\
\times \quad & \int_0^1 d\lambda p_{c,K}(t_k-\tau+s\lambda,(w,z),(x',y')) \qquad (4.83)
\end{aligned}
$$

which together with (4.82) gives us the final control

$$|J_2^3| \leq \frac{s^{\gamma/2}(1+|x'|^{1+\gamma/2})}{t_k} C^{r+2} t_k^{\frac{(r+1)\gamma}{2}} \prod_{i=1}^{r} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) \int_0^1 d\lambda p_{c,K}(t_k+s\lambda,(x,y),(x',y')),$$

Taking together:

$$|J_1| \leq \left[\frac{s}{t_k}+\frac{s|x|}{t_k^{3/2}}\right] C^{r+2} t_k^{(r+1)\gamma/2} \prod_{i=1}^{r} B(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}) \int_0^1 d\lambda p_{c,K}(t_k+\lambda s,(x,y),(x',y')),$$

$$|J_2| \leq \frac{s^{\gamma/2}(1+|x'|^{\gamma/2})}{t_k} C^{r+2} t_k^{\frac{(r+1)\gamma}{2}} \prod_{i=1}^{r} B\left(1+\frac{(i-1)\gamma}{2}, \frac{\gamma}{2}\right) \int_0^1 d\lambda p_{c,K}(t_k+s\lambda,(x,y),(x',y')),$$

and equilibrating with $|x|$ and $|x'|$:

$$\frac{s|x|}{(t_k)^{3/2}} \quad \leq \quad \frac{s[|x-x'|+|x'|]}{t_k^{3/2}} = \left[\frac{|x-x'|}{t_k^{1/2}} t_k^{1/2} + |x'|\right] \frac{s}{t_k^{3/2}} = \frac{s}{t_k^{2-\gamma/2}} + \frac{s|x'|}{t_k^{3/2}}.$$

yields to the proof of the induction hypothesis (4.74) . □

Directly from Lemma 4.75, (4.20), (4.2.1) and the inequality $(t_k)^{-1} \leq 2(t_j)^{-1}$ for $1 \leq k \leq j/2$, one can get:

$$\begin{aligned}
|(4.73)| \leq C(b,T)E_{\gamma/2,1} &\sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} (1+|x|^{1+\gamma/2}) \left\{ \frac{(u-t_k)^{\gamma/2}}{t_k} + \frac{(u-t_k)}{t_k^{3/2}} \right\} \\
\times \int_0^1 & d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(w,z)) \frac{p_{c,K}(t_j-u,(w,z),(x',y'))}{(t_j-u)^{1-\gamma/2}} dwdz \\
&= C(b,T)E_{\gamma/2,1} \sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \left\{ \frac{(u-t_k)^{\gamma/2}(1+|x|^{1+\gamma/2})}{t_k} \right\} \\
\times \int_0^1 & d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(w,z)) \frac{p_{c,K}(t_j-u,(w,z),(x',y'))}{(t_j-u)^{1-\gamma/2}} dwdz \\
&+C(b,T)E_{\gamma/2,1} \sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \left\{ \frac{(u-t_k)(1+|x|^{1+\gamma/2})}{t_k^{3/2}} \right\} \\
\times \int_0^1 & d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(w,z)) \frac{p_{c,K}(t_j-u,(w,z),(x',y'))}{(t_j-u)^{1-\gamma/2}} dwdz \\
&=: \Delta_1 D^{d,1} + \Delta_2 D^{d,1}
\end{aligned}$$

For all $\eta \in (0,\gamma)$ it holds:

$$\Delta_1 D^{d,1} \leq C(b,T)E_{\gamma/2,1}\sum_{k=1}^{j-1}\int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \left\{\frac{(u-t_k)^{\frac{\gamma-\eta}{2}}(1+|x|^{1+\gamma/2})}{t_k^{1-\eta/2}}\right\}$$
$$\times \int_0^1 d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(w,z))\frac{p_{c,K}(t_j-u,(w,z),(x',y'))}{(t_j-u)^{1-\gamma/2}}dwdz$$
$$\leq C(b,T)E_{\gamma/2,1}(1+|x|^{1+\gamma/2})h^{\frac{\gamma-\eta}{2}}\sup_{\bar{u}\in[t_j-h,t_j]} p_{c,K}(\bar{u},(x,y),(x',y')), \quad (4.84)$$

$$\Delta_1 D^{d,2} \leq C(b,T)E_{\gamma/2,1}\sum_{k=1}^{j-1}\int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} \left\{\frac{(u-t_k)^{\gamma-\eta}(1+|x|^{1+\gamma/2})}{t_k^{1/2+\gamma/2}}\right\}$$
$$\times \int_0^1 d\lambda p_{c,K}(t_k+\lambda(u-t_k),(x,y),(w,z))\frac{p_{c,K}(t_j-u,(w,z),(x',y'))}{(t_j-u)^{1-\gamma/2}}dwdz$$
$$\leq C(b,T)E_{\gamma/2,1}(1+|x|^{1+\gamma/2})h^{\frac{\gamma-\eta}{2}}\sup_{\bar{u}\in[t_j-h,t_j]} p_{c,K}(\bar{u},(x,y),(x',y')). \quad (4.85)$$

Thus,

$$\begin{aligned}|D^{d,1}| &\leq |(4.72)|+|(4.84)|+|(4.85)|\\ &\leq C(b,T)E_{\gamma/2,1}h^{\frac{\gamma-\eta}{2}}(1+|x|^{1+\gamma/2})\sup_{\bar{u}\in[t_j-h,t_j]} p_{c,K}(\bar{u},(x,y),(x',y')), \quad (4.86)\end{aligned}$$

for $\eta \in (0,\gamma)$.

- Bounds for the term $D^{d,2}$.

$$\sum_{k=0}^{j-1}\int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} p(t_k,(x,y),(w,z))[H(t_j-u,(w,z),(x',y'))-H(t_j-t_k,(w,z),(x',y'))]dwdz \quad (4.87)$$

As usual, consider the case $k=0$ separately:

$$\begin{aligned}\int_0^h & du \int_{\mathbb{R}^{2d}} p(0,(x,y),(w,z))[H(t_j-u,(w,z),(x',y'))-H(t_j,(w,z),(x',y'))]dwdz\\ &= \int_0^h du \int_{\mathbb{R}^{2d}} H(t_j-u,(x,y),(x',y'))-H(t_j,(x,y),(x',y'))]dwdz\\ &\leq C\int_0^h du \int_{\mathbb{R}^{2d}} |H(t_j-u,(x,y),(x',y'))|+|H(t_j,(x,y),(x',y'))]|dwdz\\ &\leq \frac{Ch}{t_j^{1-\gamma/2}}p_{c,K}(t_j,(x,y),(x',y')) \leq Ch^{\gamma/2}p_{c,K}(t_j,(x,y),(x',y')). \quad (4.88)\end{aligned}$$

The result in Lemma 4.5.4 yields for $k > 1$:

$$
\begin{aligned}
& \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} p(t_k,(x,y),(w,z))[H(t_j-u,(w,z),(x',y')) - H(t_j-t_k,(w,z),(x',y'))]dwdz \\
\leq & \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} C(1+|x'|^{1+\gamma/2}) p_{c,K}(t_k,(x,y),(w,z)) \left\{ \frac{(u-t_k)}{(t-u)^{2-\gamma/2}} + \frac{(u-t_k)^{\gamma/2}}{(t-u)} + \frac{(u-t_k)}{(t-u)^{5/2-\gamma/2}} \right\} \\
\times & \int_0^1 d\lambda p_{c,K}(t-u+(u-t_k)\lambda,(w,z),(x',y'))dwdz \qquad (4.89)
\end{aligned}
$$

To equilibrate with the most singular term $\frac{(u-t_k)|x'|(1+|x'|^{\gamma/2})}{(t-u)^{5/2-\gamma/2}}$ we have to balance with the parameter β:

$$
\frac{(u-t_k)^{\beta}}{(t-u)^{5/2-\gamma/2-(1-\beta)}} \frac{(u-t_k)^{(1-\beta)}}{(t-u)^{(1-\beta)}} \leq C \frac{(u-t_k)^{\beta}}{(t-u)^{5/2-\gamma/2-(1-\beta)}},
$$

To have the integrable singularity one has to impose the following conditions on γ and β: $5/2-\gamma/2-(1-\beta) < 1$, which is only possible if $\gamma/2 > 1/2+\beta$. This is a key point - from now we have to assume that the Hölder index γ is at least bigger than 1/2. And the parameter *beta* have to be chosen as following: $0 < \beta < \gamma - 1/2$. Basically, we can just rewrite $\gamma := 1/2 + \beta$ for some $\beta \in (0, 1/2]$. Under mentioned assumptions we can achieve the total rate of convergence $h^{\beta}, \beta \in (0, 1/2]$ which, according to the restrictions we imposed before, means in case γ is close to 1 (getting closer to Lipschitz assumptions on coefficients) one can get the "standard" convergence rate of $h^{\gamma-1/2}$.

As the result, for any $\beta \in (0, 1/2]$, summing (4.89), one has:

$$
\begin{aligned}
& \sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} p(t_k,(x,y),(w,z)) \\
\times & [H(t_j-u,(w,z),(x',y')) - H(t_j-t_k,(w,z),(x',y'))]dwdz \\
\leq & C \sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} du \int_{\mathbb{R}^{2d}} p_{c,K}(t_k,(x,y),(w,z)) \frac{(u-t_k)^{\beta}(1+|x'|^{1+\gamma/2})}{(t-u)^{5/2-\gamma/2-(1-\beta)}} \\
\times & \int_0^1 d\lambda p_{c,K}(t-u+(u-t_k)\lambda,(w,z),(x',y'))dwdz \\
\leq & C(b,T)(1+|x'|^{1+\gamma/2}) h^{\beta} \sup_{\bar{u} \in [t_j-h,t_j]} p_{c,K}(\bar{u},(x,y),(x',y')) \qquad (4.90)
\end{aligned}
$$

$$
|D^{d,2}| \leq (4.88) + (4.90) \leq C(b,T)(1+|x'|^{1+\gamma/2}) h^{\beta} \sup_{\bar{u} \in [t_j-h,t_j]} p_{c,K}(\bar{u},(x,y),(x',y')) \qquad (4.91)
$$

From (4.86) and (4.91) we finally get:

$$\begin{aligned}(p\otimes H-p\otimes_h H)(t_j,(x,y),(x',y')) &= (D^{d,1}+D^{d,2})(t_j,(x,y),(x',y'))\\ \le C(b,T)E_{\gamma/2,1}h^{\frac{\gamma-\eta}{2}}|x|(1+|x|^{\gamma/2}) &\sup_{\bar u\in[t_j-h,t_j]} p_{c,K}(\bar u,(x,y),(x',y'))\\ +C(b,T)(1+|x'|^{1+\gamma/2})h^{\beta} &\sup_{\bar u\in[t_j-h,t_j]} p_{c,K}(\bar u,(x,y),(x',y')) \qquad (4.92)\end{aligned}$$

Consequently, we also obtain

$$\begin{aligned}&|(p\otimes H-p\otimes_h H)\otimes_h H(t_j,(x,y),(x',y'))| \qquad (4.93)\\ &\le C^2((1+|x|^{1+\gamma/2})\vee(1+|x'|^{1+\gamma/2}))t_j^{\gamma/2}h^{\beta}B(1,\frac{\gamma}{2})\sup_{\bar u\in[t_j-h,t_j]} p_{c,K}(\bar u,(x,y),(x',y')),\end{aligned}$$

where $T\to C(T,b,\sigma)$ is non decreasing function.
and by induction, for $r\ge 0$:

$$\begin{aligned}&|(p\otimes H-p\otimes_h H)\otimes_h H^{(r)}(t_j,(x,y),(x',y'))| \qquad (4.94)\\ &\le C^{r+1}((1+|x|^{1+\gamma/2})\vee(1+|x'|^{1+\gamma/2}))h^{\beta}t_j^{r\frac{\gamma}{2}}\prod_{i=1}^{r}B\left(1+(i-1)\frac{\gamma}{2},\frac{\gamma}{2}\right)\\ &\times\sup_{\bar u\in[t_j-h,t_j]} p_{c,K}(\bar u,(x,y),(x',y')). \qquad (4.95)\end{aligned}$$

Plugging this in (4.70), due to the asymptotic of the Gamma function, one gets:

$$\begin{aligned}&|(p-p^d)(t_i,(x,y),(x',y'))|\\ &\le C(T,b,\sigma,\gamma,\beta)((1+|x|^{1+\gamma/2})\vee(1+|x'|^{1+\gamma/2}))h^{\beta}\sup_{\bar u\in[t_j-h,t_j]} p_{c,K}(\bar u,(x,y),(x',y')).\end{aligned}$$

Combining with (4.69) we complete the proof of the theorem 4.5.1.

4.6 Appendix

Proof of Theorem (4.4.2).

Proof. For the sake of simplicity, we consider the most singular case of $\varepsilon = 1$.

We start from the parametrix representation as usual. The basic strategy is to consider derivatives for the main part at first and then - for the reminder term. Having bounds for the main term from (4.17) we turn to the rest of the parametrix sum.

$$D_x^\alpha p(t,(x,y),(x',y')) = D_x^\alpha \tilde{p}(t,(x,y),(x',y')) + \sum_{r=1}^{\infty} D_x^\alpha \left[\tilde{p} \otimes H^{\varepsilon,(r)}(t,(x,y),(x',y'))\right],$$

for $|\alpha| = 1, 2$.

Since the first order direvation gives an integrable singularity in the time, we don't have any problems for $|\alpha| = 1$.

The case $|\alpha| = 2$ we have to discuss precisely. Let us denote

$$\begin{aligned} R(t,(x,y),(x',y')) &:= \sum_{r=1}^{\infty} \tilde{p} \otimes H^{(r)}(t,(x,y),(x',y')) = \tilde{p} \otimes \Phi(t,(x,y),(x',y')), \\ \Phi(t,(x,y),(x',y')) &:= \sum_{r=1}^{\infty} H^{(r)}(t,(x,y),(x',y')). \end{aligned}$$

Inequality (4.20) for H then yields for all $r \in \mathbb{N}^*, 0 < t \le T, (x,y),(x',y') \in (\mathbb{R}^{2d})^2$:

$$|H^{(r)}(t,(x,y),(x',y'))| \le ((1 \vee T^{(1-\gamma)/2})c_1)^r \prod_{i=1}^{r-1} B(\frac{\gamma}{2}, 1+(i-1)\frac{\gamma}{2}) p_{c,K}(t,(x,y),(x',y')) t^{-1+\frac{r\gamma}{2}}, \tag{4.96}$$

with the convention $\prod_{i=1}^{0} = 1$. We thus derive that for all $0 < t \le T, (x,y) \in (\mathbb{R}^d)^2$:

$$|\Phi(t,(x,y),(x',y'))| \le \frac{C}{t^{1-\gamma/2}} p_{c,K}(t,(x,y),(x',y')). \tag{4.97}$$

Then,

$$\begin{aligned} D_x^\alpha R(t,(x,y),(x',y')) &= \lim_{\tau \to 0} \int_\tau^{t/2} du \int_{\mathbb{R}^{2d}} D_x^\alpha \tilde{p}(u,(x,y),(w,z)) \Phi(t-u,(w,z),(x',y')dwdz \\ &+ \int_{t/2}^{t} du \int_{\mathbb{R}^{2d}} D_x^\alpha \tilde{p}(u,(x,y),(w,z)) \Phi(t-u,(w,z),(x',y'))dwdz \\ &=: \lim_{\tau \to 0} D_x^\alpha R^\tau(t,(x,y),(x',y')) + D_x^\alpha R^f(t,(x,y),(x',y')). \end{aligned} \tag{4.98}$$

The contribution $D_x^\alpha R^f(t,(x,y),(x',y'))$ does not exhibit time singularities in the integral, since on the considered integration set $u \geq \frac{1}{2}t$.

Thus, from inequalities (4.17) and (4.97):

$$|D_x^\alpha R^f(t,(x,y),(x',y'))| \leq \frac{C}{(t-s)^{(|\alpha|-\gamma)/2}} p_{c,K}(t,(x,y),(x',y')). \tag{4.99}$$

We should put more effort in the estimation of the rest part:$D_x^\alpha R^\tau(t,(x,y),(x',y'))$. For $|\alpha|=2$ we apply some kind of the cancellation properties of the Gaussian kernels as in [KM17].

Introduce for an arbitrary $\kappa^1,\kappa^2 \in \mathbb{R}^{2d}$: $\hat{C}_t := \int_0^t R_{t-u} B a(\kappa^1,\kappa^2) B^* R_{t-u}^* du$, $R_s = \begin{pmatrix} I_d & 0_d \\ sI_d & I_d \end{pmatrix}$, $B = \begin{pmatrix} I_{d\times d} \\ 0_{d\times d} \end{pmatrix}$ and

$$\begin{aligned} \tilde{p}^{\kappa^1,\kappa^2}(u,(x,y),(w,z)) &= \frac{\exp\Bigg(-\frac{1}{2}\langle \hat{C}_u^{\varepsilon,-1} Z, Z\rangle\Bigg)}{(2\pi)^d \det(\hat{C}_u^\varepsilon(\kappa^1,\kappa^2))^{1/2}}, \\ Z &:= \begin{pmatrix} w-x \\ z-y-xu \end{pmatrix}. \end{aligned} \tag{4.100}$$

Hence, for all multi-index α, $|\alpha|=2$:

$$\int_{\mathbb{R}^{2d}} D_x^\alpha \tilde{p}^{\kappa^1,\kappa^2}(u,(x,y),(w,z)) dwdz = 0. \tag{4.101}$$

Introducing the centering function $c^\alpha(u,(x,y),(w,z)) := (D_x^\alpha \tilde{p}^{\kappa^1,\kappa^2}(u,(x,y),(w,z)))|_{(\kappa^1,\kappa^2)=(x,y)}$, we derive that:

$$\begin{aligned} D_x^\alpha R^\tau(t,(x,y),(w,z)) &= \int_\tau^{t/2} du \int_{\mathbb{R}^{2d}} (D_x^\alpha \tilde{p} - c^\alpha)(u,(x,y),(w,z)) \Phi(t-u,(w,z),(x',y')) dwdz \\ &+ \int_\tau^{t/2} du \int_{\mathbb{R}^{2d}} c^\alpha(u,(x,y),(w,z)) (\Phi(t-u,(w,z),(x',y')) - \Phi(t-u,(x,y+xu),(x',y')) dwdz \\ &:= (R^{\tau,1} + R^{\tau,2})(t,(x,y),(x',y')), \end{aligned} \tag{4.102}$$

exploiting the centering condition (4.101) to introduce the last term of the first equality.

Since $c^\alpha(u,(x,y),(w,z))$ contains $\tilde{p}^{\kappa^1,\kappa^2}(u,(x,y),(w,z)))|_{(\kappa^1,\kappa^2)=(x,y)}$ as a 'true' density w.r.t. (w,z) we can cancel with $\Phi(t-u,(x,y+xu),(x',y'))$.

We recall that

$$|c^\alpha(u,(x,y),(w,z))| \le \frac{C}{u} p_{c,K}(u,(x,y),(w,z)).$$

On the one hand, the terms $D_x^\alpha \tilde p(u,(x,y),(w,z)), c^\alpha(u,(x,y),(w,z))$ differ in their frozen coefficients (respectively at point w,z and x,y). Moreover there is no back-flows w.r.t. the second variable in $c^\alpha(u,(x,y),(w,z))$. Exploiting the Hölder property in space of the mollified coefficients, it is then seen that:

$$\begin{aligned}|(D_x^\alpha \tilde p - c^\alpha)(u,(x,y),(w,z))| \ &\le \ C\left[\frac{|w-x|^\gamma}{u} + \frac{|z-y-xu|^\gamma}{u}\right] p_{c,K}(u,(x,y),(w,z)) \\ \le \frac{C}{u^{1-\gamma/2}} p_{c,K}(u,(x,y),(w,z)),\end{aligned}$$

where the last inequality comes from the standard absorbation of the additional time singularity with the Gaussian density $p_{c,K}(u,(x,y),(w,z))$. Thus, from (4.97):

$$|R^{\tau,1}(t,(x,y),(x',y'))| \le \frac{C}{t^{|\alpha|-\gamma}} p_{c,K}(t,(x,y),(x',y')). \tag{4.103}$$

The key idea to control the contribution of the rest part is to use the smoothing effect comes from the kernel Φ.

Lemma 4.6.1. *For* $A_u := \{(w,z) \in \mathbb{R}^{2d} : |w-x| + \frac{|z-y-xu|}{u} \le ct^{1/2}\}$ *(recall as well that* $u \in [0, \frac{t}{2}]$*) one has:*

$$\begin{aligned}&|\Phi(t-u,(x,y+xu),(x',y')) - \Phi(t-u,(w,z),(x',y'))| \\ &\le \frac{C}{(t-u)^{1-\gamma/4}}\left(|x-w|^{\gamma/2} + |z-y-ux|^{\gamma/2}\right) p_{c,K}(t-u,(w,z),(x',y')).\end{aligned} \tag{4.104}$$

Proof. From the definition of Φ and the smoothing effect of the kernel H in (4.96), it suffices to prove that on the set $\bar A_u := \{z,w \in \mathbb{R}^{2d} : |x-w| + \frac{|z-y-xu|}{u'-u} \le c(u'-u)^{1/2}\}$:

$$\begin{aligned}&|H(u'-u,(x,y+xu),(x'',y'')) - H(u'-u,(w,z),(x'',y''))| \\ &\le C\frac{|x-w|^{\gamma/2} + |z-y-xu'|^{\gamma/2}}{(u'-u)^{1-\gamma/4}} p_{c,K}(u'-u,(w,z),(x'',y'')),\end{aligned} \tag{4.105}$$

for $u' \in (u,t],\ u \in [0,t/2]$. Observe that $\bar A_u \subset A_u$.

Let us first prove (4.105). We concentrate on the second derivatives in H which yield the most singular contributions:

$$
\begin{aligned}
&\mathrm{Tr}((a(x,y+xu)-a(x'',y''-x''(u'-u)))D_x^2\tilde p(u'-u,(x,y+xu),(x'',y''))\\
&\quad -\mathrm{Tr}((a(w,z)-a(x'',y''-x''(u'-u)))D_x^2\tilde p(u'-u,(w,z),(x'',y''))\\
&\quad = \mathrm{Tr}((a(x,y+xu)-a(w,z))D_x^2\tilde p(u'-u,(x,y+xu),(x'',y'')))\\
&\quad -\mathrm{Tr}((a(w,z)-a(x'',y''-x''(u'-u)))\\
&\times(D_x^2\tilde p(u'-u,(w,z),(x'',y''))-D_x^2\tilde p(u'-u,(x,y+xu),(x'',y''))) =: I+II.
\end{aligned}
\tag{4.106}
$$

Then, from (4.17),

$$
\begin{aligned}
|I| &\le C\frac{|x-w|^\gamma+|z-y-xu|^\gamma}{(u-u')}p_{c,K}(u'-u,(w,z),(x'',y''))\\
&\le \frac{C|x-w|^{\gamma/2}+|z-y-xu|^{\gamma/2}}{(u-u')^{1-\gamma/4}}p_{c,K}(u'-u,(w,z),(x'',y'')))
\end{aligned}
\tag{4.107}
$$

using that $(w,z)\in\bar A_u$ for the second inequality. Now, from the explicit expression of the second order derivatives in (4.100), (**AD2**) and usual computations we also derive:

$$
\begin{aligned}
|II| \le &\left(|w-x''|^\gamma+|z-(y''-(u'-u)x'')|^\gamma\right)\int_0^1 \frac{d\lambda}{(u'-u)^{2d}}\exp\left(-\left\{\frac{|w-x''+\lambda(x-w)|^2}{c(u'-u)}\right.\right.\\
&\left.\left.+\frac{|y''-z-(u-u')w+\lambda(y+xu-z)|^2}{c(u'-u)^3}\right\}\right)\times\left(\frac{|w-x|}{(u'-u)^{3/2}}+\frac{|y+xu-z|}{(u'-u)^{5/2}}\right).
\end{aligned}
\tag{4.108}
$$

Due to $\bar A_u$ definition the term $\frac{|w-x|}{(u'-u)^{3/2}}+\frac{|y+xu-z|}{(u'-u)^{5/2}}$ brings the singularity of order $\frac{1}{u'-u}$. Moreover

$$
\begin{aligned}
&-\frac{|w-x''+\lambda(x-w)|^2}{c(u'-u)}-\frac{|y''-z-(u-u')w+\lambda(y+xu-z)|^2}{c(u'-u)^3}\\
&\le -\frac{|w-x''|^2}{c(u'-u)}-\frac{|y''-z-(u-u')w|^2}{c(u'-u)^3}-\left(\frac{|x-w|^2}{c(u'-u)}+\frac{|y+xu-z|^2}{c(u'-u)^3}\right),\\
&\left(\frac{|x-w|^2}{c(u'-u)}+\frac{|y+xu-z|^2}{c(u'-u)^3}\right)\le C \text{ for } (w,z)\in\bar A_u.
\end{aligned}
$$

Finally,

$$(4.108) \leq \frac{C(|w-x''|^\gamma + |z-(y''-(u'-u)x'')|^\gamma)}{(u'-u)(u'-u)^{2d}} \exp\left\{ -\frac{|w-x''|^2}{c(u'-u)} - \frac{|y''-z-(u-u')w|^2}{c(u'-u)^3} \right\}$$
$$\leq \frac{C|x-w|^{\gamma/2} + |z-y-xu|^{\gamma/2}}{(u-u')^{1-\gamma/4}} p_{c,K}(u'-u,(w,z),(x'',y''))).$$

using the usual convexity argument for the last inequality. Thus, we have proved (4.105) on $\bar{A}_u \subset A_u$.

Recalling that we want to establish (4.104) on A_u, we consider the rest case: if $(w,z) \notin \bar{A}_u$, we get from (4.96):

$$\int_u^t du' \int_{\bar{A}_u^c} |H(u'-u,(x,y-xu),(x'',y'')) - H(u'-u,(w,z),(x'',y''))|$$
$$\times |(\sum_{i\geq 2} H^{(i)})(t-u',(x'',y''),(x',y')|dx''dy''$$
$$\leq \int_u^t du' \int_{\bar{A}_u^c} \frac{C}{(u'-u)^{1-\gamma/2}} (p_{c,K}(u'-u,(x,y-xu),(x'',y'')) + p_{c,K}(u'-u,(w,z),(x'',y'')))$$
$$\times \frac{|x-z|^{\gamma/2} + |y-w-xu|^{\gamma/2}}{(u'-u)^{\gamma/4}} \frac{C}{(t-u')^{1-\gamma}} p_{c,K}(t-u',(x'',y''),(x',y'))dx''dy''$$
$$\leq C \int_u^t \int_{\bar{A}_u^c} \frac{du'}{(u'-u)^{1-\gamma/4}} \frac{|x-w|^{\gamma/2} + |y+xu-z|^{\gamma/2}}{(t-u')^{1-\gamma}}$$
$$\times \left\{ p_{c,K}(t-u,(w,z),(x',y')) + p_{c,K}(t-u,(x,y+ux),(x',y')) \right\} dx''dy''$$

exploiting that $(w,z) \in \bar{A}_u^c$.

Let us consider precisely the compatibility of $p_{c,K}(t-u,(x,y+xu),(x',y'))$ and $p_{c,K}(t-s,(x,y),(x',y')$. Observe that,

$$p_{c,K}(t-u,(x,y+xu),(x',y')) \leq \frac{C}{(t-u)^{2d}} \exp\left\{ -\left[\frac{|x-x'|^2}{t-u} + \frac{|y+ux+x(t-u)-y'|^2}{(t-u)^3} \right] \right\}$$
$$\leq C p_{c,K}(t,(x,y),(x',y')), \tag{4.109}$$

recalling that $t-u$ is of order t for the last inequality.

As the result, we have proved (4.105) and completed the proof of the Lemma. □

Then, we can derive from (4.17), (4.102) and (4.104):

$$|R^{\tau,2}(t,(x,y),(x',y'))| \le C^2 \int_\tau^{t/2} du \int_{A_u} \frac{|x-w|^{\gamma/2}+|y-z-xu|^{\gamma/2}}{u} p_{c,K}(u,(x,y),(w,z))$$

which now compatible to absorb the singularity $\times \dfrac{1}{(t-u)^{1-\gamma/4}} p_{c,K}(t-u,(w,z),(x',y'))dwdz$

$$+\frac{C}{t^{\gamma/4}} \int_\tau^{t/2} du \int_{A_u^C} \frac{|x-w|^{\gamma/2}+|y-z-xu|^{\gamma/2}}{u} p_{c,K}(u,(x,y),(w,z))$$
$$\times\{|\Phi(t-u,(w,z),(x',y'))|+|\Phi(t-u,(x,y-xu),(x',y'))|\}dwdz. \tag{4.110}$$

On the complementary set A_u^c it holds:

$$|\int_\tau^{t/2} du \int_{A_u^c} c^\alpha(u,(x,y),(w,z))(\Phi(t-u,(w,z),(x',y'))-\Phi(t-u,(x,y+xu),(x',y'))dwdz|$$
$$\le \frac{1}{t^{\gamma/4}} \left| \int_\tau^{t/2} \int_{A_u^c} \frac{|w-x|^{\gamma/2}+|z-(y+ux)|^{\gamma/2}}{u} p_{c,K}(u,(x,y),(w,z)) \right.$$
$$\left. \times \left[|\Phi(t-u,(w,z),(x',y'))|+|\Phi(t-u,(x,y+xu),(x',y'))| \right] dwdz \right|$$
$$\le \frac{1}{t^{\gamma/4}} \left| \int_\tau^{t/2} du \int_{A_u^c} \frac{1}{u^{1-\gamma/4}} \frac{1}{(t-u)^{1-\gamma/2}} p_{c,K}(u,(x,y),(w,z)) \right.$$
$$\left. \times \left[p_{c,K}(t-u,(w,z),(x',y'))+p_{c,K}(t-u,(x,y+xu),(x',y')) \right] dwdz \right| \tag{4.111}$$

Plugging (4.109) that into (4.111) one can get the bound on A_u^c:

$$|\int_\tau^{t/2} du \int_{A_u^c} c^\alpha(u,(x,y),(w,z))(\Phi(t-u,(w,z),(x',y'))-\Phi(t-u,(x,y+xu),(x',y'))dwdz|$$
$$\le \frac{1}{t^{\gamma/4}} p_{c,K}(t,(x,y),(x',y')) \left| \int_\tau^{t/2} \frac{du}{u^{1-\gamma/4}} \frac{1}{(t-u)^{1-\gamma/2}} \right| \le \frac{C}{t^{1-\gamma/2}} p_{c,K}(t,(x,y),(x',y')). \tag{4.112}$$

thus, taking Lemma 4.6.1 into account, we have:

$$
\begin{aligned}
\left|R^{\tau,2}(t,(x,y),(x',y'))\right| &\le C\int_\tau^{t/2} du \int_{A_u} \frac{|x-w|^{\gamma/2}+|y-z-xu|^{\gamma/2}}{u} p_{c,K}(u,(x,y),(w,z)) \\
&\times \frac{1}{(t-u)^{1-\gamma/4}} p_{c,K}(t-u,(w,z),(x',y'))dwdz + \frac{C}{t^{1-\gamma/2}} p_{c,K}(t,(x,y),(x',y')) \\
&\le \frac{C}{t^{1-\gamma/2}} p_{c,K}(t,(x,y),(x',y')),
\end{aligned}
$$

which together with (4.103), (4.102), (4.99) and (4.98) gives the statement of the Section. □

Bibliography

[AKH17] Patrik Andersson and Arturo Kohatsu-Higa. Unbiased simulation of stochastic differential equations using parametrix expansions. *Bernoulli*, 23:2028–2057, 2017.

[Aro59] D. G. Aronson. The fundamental solution of a linear parabolic equation containing a small parameter. *Ill. Journ. Math.*, 3:580–619, 1959.

[BA09] V. Bally and Kohatsu Higa A. Lower bounds for densities of Asian type-stochastic differential equations. *to appear in J. Funct. Anal.*, 2009.

[Bai17] N.T.G. Bailey. *The mathematical theory of infectious diseases and its applications.* 2nd ed. Hafner Press, New York, 2017.

[BGM10] E. Benhamou, E. Gobet, and M. Miri. Expansion formulas for european options in a local volatility model. *Inter. J. Theor. Appl. Fin.*, 13:602 – 634, 2010.

[BL08] T. Bodineau and L. Lefevere. Large deviations of lattice hamiltonian dynamics coupled to stochastic thermostats. *J. Stat. Phys.*, 133:1–27, 2008.

[BP09] R.F. Bass and E.A. Perkins. A new technique for proving uniqueness for martingale problems. *From Probability to Geometry (I): Volume in Honor of the 60th Birthday of Jean-Michel Bismut*, pages 47–53, 2009.

[BPV01] E. Barucci, S. Polidoro, and V. Vespri. Some results on partial differential equations and asian options. *Math. Models Methods Appl. Sci*, 3:475–497, 2001.

[BR76] R. Bhattacharya and R. Rao. *Normal approximations and asymptotic expansions.* Wiley and sons, 1976.

[BT96a] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Prob. Th. Rel. Fields*, 104-1:43–60, 1996.

[BT96b] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations, II. Convergence rate of the density. *Monte-Carlo methods and Appl.*, 2:93–128, 1996.

[BY89] N. Becker and P. Yip. Analysis of variations in an infection rate. *Australian Journal of Statistics*, 31(1), 1989.

[CFP10] F. Corielli, P. Foschi, and A. Pascucci. Parametrix approximation of diffusion transition densities. *SIAM J. Fin. Math*, 1:833 – 867, 2010.

[CM04] H. Crámer and Leadbetter M.R. *Statistical Inference for Ergodic Diffusion Processes*. Dover Publications, Inc., Mineola, NY, 2004.

[CMP15] C. Cinti, S. Menozzi, and S. Polidoro. Two-sided bounds for degenerate processes with densities supported in subsets of rn̂. *Potential Analysis*, 42(1), 2015.

[DM10] F. Delarue and S. Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis*, 259–6:1577–1630, 2010.

[Dyn65] E. B Dynkin. *Markov Processes*. Springer Verlag, 1965.

[EPRB99] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.

[FH15] N. Frikha and L. Huang. A multi-step richardson–romberg extrapolation method for stochastic approximation. *Stochastic Processes and their Applications*, 125(11):4066 – 4101, 2015.

[FKHL16] N. Frikha, A. Kohatsu-Higa, and L. Li. On the first hitting times of one dimensional elliptic diffusions. *https://arxiv.org/abs/1609.09327*, 2016.

[FP10] Corielli F. Foschi P., Pascucci A. Parametrix approximation of diffusion transition densities. *SIAM J. Financial Math.*, 1:833 – 867, 2010.

[Fri64] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, 1964.

[Fri18] N. Frikha. On the weak approximation of a skew diffusion by an Euler-type scheme. *Bernoulli*, 24(3):1653–1691, 2018.

[Gav77] B. Gaveau. Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. *Acta Math.*, 139(1-2):95–153, 1977.

[GCJ93] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'I.H.P. Probab. et stat.*, 29:119–151, 1993.

[GS67] I. Gihman and A. Skorohod. *Stochastic Differential Equations.* Naukova dumka, Kiev., 1967.

[GS82] I. Gihman and A. Skorohod. *Stochastic Differential Equations and Applications.* Naukova dumka, Kiev., 1982.

[H6̈3] L. Hörmander. *Linear partial di]]erential operators.* Springer-Verlag, Berlin-GSttingen-Heidelberg, 1963.

[H6̈7] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[Had23] J Hadamard. Lectures on cauchy's problem in linear partial differential equations, yale univ. *Press. New Haven*, 1923.

[HM16] L. Huang and S. Menozzi. A Parametrix Approach for some Degenerate Stable Driven SDEs. *Annales Instit. H. Poincaré*, 52(4):1925–1975, 2016.

[HN04] Frédéric Hérau and Francis Nier. Isotropic hypoellipticity and trend to equilibrium for the fokker-planck equation with a high-degree potential. *Archive for Rational Mechanics and Analysis*, 171(2):151–218, 2004.

[IKO62] A. M. Il'in, A. S. Kalashnikov, and O. A. Oleinik. Second-order linear equations of parabolic type. *Uspehi Mat. Nauk*, 17–3(105):3–146, 1962.

[JYC10] M. Jeanblanc, M. Yor, and M. Chesney. *Mathematical Methods for Financial Markets.* Springer Finance, London, 2010.

[KHTZ16] A. Kohatsu-Higa, D. Taguchi, and J. Zhong. The parametrix method for skew diffusions. *Potential Analysis*, 45:299–329, 2016.

[KKM17] V. Konakov, A. Kozhina, and S Menozzi. Stability of densities for perturbed Diffusions and Markov Chains. *ESAIM: Probability and Statistics*, 21:88–112, 2017.

[KM85] V.D. Konakov and S.A Molchanov. On the convergence of Markov chains to diffusion processes. *Teor. Veroyatn. Mat. Statist. (in Russian)*, pages 51–64, 1984, English translation in Theory Prob. Math. Stat., 31, 59-73, (1985).

[KM00] V. Konakov and E. Mammen. Local limit theorems for transition densities of Markov chains converging to diffusions. *Prob. Th. Rel. Fields*, 117:551–587, 2000.

[KM02] V. Konakov and E. Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods Appl.*, 8–3:271–285, 2002.

[KM05] V. Konakov and E. Mammen. Edgeworth-type expansions for transition densities of markov chains converging to diffusions. *Bernoulli*, 11:591–641, 2005.

[KM07] V. Konakov and E. Mammen. Small time edgeworth-type expansions for weakly convergent nonhomogeneous markov chains. *Probability Theory and Related Fields*, 143:137, 2007.

[KM10] V. Konakov and S. Menozzi. Weak error for stable driven stochastic differential equations: Expansion of the densities. *Journal of Theoretical Probability*, 24-2:554–578, 2010.

[KM17] V. Konakov and S Menozzi. Weak Error for the Euler Scheme Approximation of Diffusions with non-smooth coefficients. *Electronic Journal of Probability*, 22:1–47, 2017.

[KMM10] V. Konakov, S. Menozzi, and S. Molchanov. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Annales de l'Institut Henri Poincaré, Série B*, 46–4:908–923, 2010.

[Kol34] A. N. Kolmogorov. Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math.*, 2-35:116–117, 1934.

[Kol00] V. Kolokoltsov. Symmetric stable laws and stable-like jump diffusions. *Proc. London Math. Soc.*, 80:725–768, 2000.

[Koz16] A. Kozhina. Stability of densities for perturbed degenerate diffusions. *Teor. Veroyatnost. i Primenen*, 3:570–579, 2016.

[Kry96] N. V. Krylov. *Lectures on elliptic and parabolic equations in Hölder spaces.* Graduate Studies in Mathematics 12. AMS, 1996.

[KS84] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. I. *Stochastic analysis (Katata/Kyoto, 1982)*, North-Holland Math. Library, 32:271–306, 1984.

[KS85] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math*, 32:1–76, 1985.

[Lev07] E. E. Levi. Sulle equazioni lineari totalmente ellittiche alle derivate parziali. *Rend. Circ. Mat. Palermo*, 24:275–317, 1907.

[LM10] V. Lemaire and S. Menozzi. On some non asymptotic bounds for the Euler scheme. *Electronic Journal of Probability*, 15:1645–1681, 2010.

[Lun97] A. Lunardi. Schauder estimates for a class of degenerate elliptic and parabolic operators with unbounded coefficients in $\{R\}^n$. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, 24(1):133–164, 1997.

[Mar55] G. Maruyama. Continuous markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4:48, 1955.

[Men11] S. Menozzi. Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electronic Communications in Probability*, 17:234–250, 2011.

[Men18] S. Menozzi. Martingale problems for some degenerate Kolmogorov equations. *arXiv:1404.0461v3, to appear in Stoc. Proc. Appl., https://doi.org/10.1016/j.spa.2017.06.001*, 128-3:756–802, 2018.

[MGEE10] Bramanti M., Cupini G., Lanconelli E., and Priola E. Global L^p estimates for degenerate Ornstein-Uhlenbeck operators. *Math. Z.*, 266(4):789–816, 2010.

[MP91] Remigius Mikulevičius and Eckhard Platen. Rate of convergence of the Euler approximation for diffusion processes. *Math. Nachr.*, 151:233–239, 1991.

[MS67] H. P. McKean and I. M. Singer. Curvature and the eigenvalues of the Laplacian. *J. Differential Geometry*, 1:43–69, 1967.

[MSH02] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185 – 232, 2002.

[Pri09] E. Priola. Global Schauder estimates for a class of degenerate Kolmogorov equations. *Studia Math.*, 194(2):117–153, 2009.

[RBL00] L. Rey-Bellet and L.E. Lawrence. Asymptotic behavior of thermal nonequilibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.*, 215(1):1–24, 2000.

[Shi96] A.N. Shiryaev. *Probability, Second Edition.* Graduate Texts in Mathematics, 95. Springer-Verlag, New York., 1996.

[Son67] I. M. Sonin. A class of degenerate diffusion processes (russian). *Teor. Verojatnost. i Primenen*, 12:540–547, 1967.

[SV79] D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes.* Springer-Verlag Berlin Heidelberg New-York, 1979.

[Tal02] D. Talay. Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8–2:163–198, 2002.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. and App.*, 8-4:94–120, 1990.

[Web51] M. Weber. The fundamental solution of a degenerate partial differential equation of parabolic type. *Trans. Amer. Math. Soc.*, 71:24–37, 1951.