UNIVERSITÉ PARIS-EST

UNIVERSITÀ DEGLI STUDI DI PADOVA

École Doctorale Mathématiques et STIC. Discipline: Mathématiques
Thése en cotutelle pour l'obtention du

Diplôme de Docteur de l'Université Paris-Est

Università degli Studi di Padova
Scuola di Dottorato in Scienze matematiche
Indirizzo: Matematica Computazionale

Tube Estimates for Hypoelliptic Diffusions and Scaling Properties of Stochastic Volatility Models

---

Candidat / Candidato

Paolo PIGATO

Directeurs de thèse / Relatori di tesi: Vlad BALLY, Paolo DAI PRA

Rapporteurs / Referees: Lucia CARAMELLINO, Peter FRIZ, Arnaud GUILLIN

Soutenue le 16 Octobre 2015 devant le jury composé de:
Difesa il 16 Ottobre 2015 davanti alla commissione composta da:

Vlad BALLY	Directeur de thèse / Relatore di Tesi
Francesco CARAVENNA	Examinateur / Esaminatore
Paolo DAI PRA	Directeur de thèse / Relatore di Tesi
François DELARUE	Examinateur / Esaminatore

Acknowledgments

I want to express my most sincere thanks to my advisors, Vlad Bally and Paolo Dai Pra, who introduced me to the fascinating world of scientific research. They suggested me interesting and challenging problems to work on, sharing with me their deep mathematical knowledge, guiding me while at the same time giving me the freedom to follow my intuition and personal inclination. I thank them for their constant encouragement, support and dedication.

I thank Lucia Caramellino, Peter Friz and Arnaud Guillin for agreeing to act as referees. Their remarks and suggestions have been of great help.

I am grateful to Francesco Caravenna and François Delarue for accepting to be part of the defense committee.

I have had the luck of discussing the content of the thesis in several occasions, getting useful insights and interesting ideas. For this I feel particularly indebted to Francesco Caravenna, Gianmarco Chinello, Giovanni Conforti, Noufel Frikha, Arturo Kohatsu-Higa, Eva Locherbach, Stephane Menozzi. I would like to thank Arturo also for his hospitality during my visit to Ritsumeikan University in Kyoto.

This thesis was written at the Laboratoire d'Analyse Mathématiques Appliquées of Université Paris-Est Marne-la-Vallé and at the Dipartimento di Matematica of Università di Padova. In both places I have found an excellent working environment and I have met great people. I would like to express my special thanks to all my Ph.D. colleagues of both universities. Merci en particulier à Marie-Noemie, Marwa, Pierre, Sebastian et Xavier, avec lesquels j'ai partagé le plus. Je tiens aussi à remercier Sylvie Cach pour son aide et sa disponibilité dans la gestion de la partie administrative de la cotutelle.

Merci à tous les amis parisiens, grâce auxquels cette expérience en France était si intéressante et heureuse. Voglio anche ringraziare tutte le persone che mi sono state vicine in questi anni: i miei amici d'infanzia, gli amici di Bassano e di Padova. Grazie in particolare a Enrico e Giovanni, per le infinite discussioni sugli argomenti più diversi. Un grande ringraziamento va a tutta la mia famiglia, soprattutto ai miei genitori e a Chiara, per essere un punto di riferimento su cui poter contare sempre.

This work has been financially supported by a Ph.D scholarship of Università di Padova, by the "Bourse de soutien aux thèses en cotutelle" of Université Paris-Est, and by the INRIA team Mathrisk for the last months. Their support is gratefully acknowledged.

Abstract

In this thesis we address two problems. In the first part we consider hypoelliptic diffusions, under both strong and weak Hörmander condition. We find Gaussian estimates for the density of the law of the solution at a fixed, short time. A main tool to prove these estimates is Malliavin Calculus, in particular some techniques recently developed to deal with degenerate problems. We then use these short-time estimates to show exponential two-sided bounds for the probability that the diffusion remains in a small tube around a deterministic path up to a given time. In our hypoelliptic framework, the shape of the tube must reflect the fact the diffusion moves with a different speed in the direction of the diffusion coefficient and in the direction of the Lie brackets. For this reason we introduce a norm accounting of this anisotropic behavior, which can be adapted to both the strong and weak Hörmander framework. We establish an equivalence between this norm and the standard control distance in the strong Hörmander case. In the weak Hörmander case, we introduce a suitable equivalent control distance.

In the second part of the thesis we work with mean reverting stochastic volatility models, with a volatility driven by a jump process. We first suppose that the jumps follow a Poisson process, and consider the decay of cross asset correlations, both theoretically and empirically. This leads us to study an algorithm for the detection of jumps in the volatility profile. We then consider a more subtle phenomenon widely observed in financial indices: the multiscaling of moments, i.e. the fact that the q-moment of the log-increment of the price on a time lag of length h scales as h to a certain power of q, which is non-linear in q. We work with models where the volatility follows a mean reverting SDE driven by a Lévy subordinator. We show that multiscaling occurs if the characteristic measure of the Lévy has power law tails and the mean reversion is super-linear at infinity. In this case the scaling function is piecewise linear.

Keywords: hypoellipticity, Hörmander condition, tube estimates for Ito processes, density estimates, stochastic volatility, multi-scaling.

Résumé

Dans cette thèse on aborde deux problèmes. Dans la première partie on considère des diffusions hypoelliptiques, à la fois sur une condition d'Hörmander forte et faible. On trouve des estimations gaussiennes pour la densité de la loi de la solution à un temps court fixé. Un outil fondamental pour prouver ces estimations est le calcul de Malliavin. On utilise en particulier des techniques développées récemment pour faire face à des problèmes de dégénérescence. Ensuite, grâce à ces estimations en temps court, on trouve des bornes inférieures et supérieures exponentielles sur la probabilité que la diffusion reste dans un petit tube autour d'une trajectoire déterministe jusqu'à un moment fixé. Dans ce cadre hypoelliptique, la forme du tube doit tenir compte du fait que la diffusion se déplace avec une vitesse différente dans les directions du coefficient de diffusion et dans les directions des crochets de Lie. Pour cette raison, on introduit une norme qui prend en compte ce comportement anisotrope, qui peut être adaptée aux cas d'Hörmander fort et faible. Dans le cas Hörmander fort on établit une équivalence entre cette norme et la distance de contrôle classique. Dans le cas Hörmander faible on introduit une distance de contrôle équivalente appropriée.

Dans la deuxième partie de la thèse, on travaille avec des modèles à volatilité +stochastique avec retour à la moyenne, où la volatilité est dirigée par un processus de saut. On suppose d'abord que les sauts suivent un processus de Poisson, et on considère la décroissance des corrélations croisées, théoriquement et empiriquement. Ceci nous amène à étudier un algorithme pour la détection de sauts de la volatilité. On considère ensuite un phénomène plus subtil largement observé dans les indices financiers: le "multiscaling" des moments, c'est-à-dire le fait que les moments d'ordre q des log-incréments du prix sur un temps h, ont une amplitude d'ordre h à une certaine puissance, qui est non linéaire dans q. On travaille avec des modèles où la volatilité suit une EDS avec retour à la moyenne dirigée par un subordinateur de Lévy. On montre que le multiscaling se produit si la mesure caractéristique du Lévy a des queues de loi de puissance et le retour à la moyenne est superlinéaire à l'infini. Dans ce cas l'exposant de scaling est linéaire par morceaux.

Mots clés: hypoellipticité, condition d'Hörmander, estimations des tube pour les processus d'Ito, estimations de densité, volatilité stochastique, multi-échelle.

Sommario

In questa tesi ci occupiamo di due problemi. Nella prima parte consideriamo delle diffusioni ipoellittiche, sia sotto una condizione di Hörmander forte che debole. Troviamo delle stime gaussiane per la densità della legge della soluzione in tempo corto. Uno strumento fondamentale per dimostrare questo tipo di stime è il calcolo di Malliavin. In particolare, utilizziamo delle tecniche sviluppate negli ultimi anni per affrontare dei problemi degeneri. Poi, grazie a queste stime in tempo corto, troviamo dei bound inferiore e superiore esponenziali per la probabilità che la diffusione rimanga in un piccolo tubo attorno a una traiettoria deterministica, fino a un tempo fissato. In questo contesto ipoellittico, la forma del tubo deve riflettere il fatto che la diffusione si muove con una velocità diversa nella direzione dei coefficienti di diffusione e nella direzione delle parentesi di Lie. Per questo motivo introduciamo una norma che tenga conto di questo comportamento anisotropo, adattabile al caso di Hörmander forte e debole. Nel caso Hörmander forte stabiliamo un'equivalenza tra questa norma e la distanza di controllo classica. Nel caso Hörmander debole introduciamo una distanza di controllo equivalente adeguata.

Nella seconda parte della tesi lavoriamo con dei modelli a volatilità stocastica con ritorno alla media, in cui la volatilità è diretta da un processo di salto. Supponiamo inizialmente che i salti siano dati da un processo di Poisson, e consideriamo il decadimento delle correlazioni incrociate, sia teoricamente che empiricamente. Questo ci porta a studiare un algoritmo per identificare i picchi nel profilo della volatilità. Consideriamo successivamente un fenomeno più sottile largamente osservato negli indici finanziari: il "multiscaling" dei momenti, ovvero il fatto che i momenti d'ordine q dei log-incrementi del prezzo su un tempo h, hanno un'ampiezza di ordine h a una certa potenza, che è non lineare in q. Lavoriamo con dei modelli in cui la volatilità è data da un'equazione differenziale stocastica con ritorno alla media, diretta da un subordinatore di Lévy. Mostriamo che il multiscaling si produce se la misura caratteristica del Lévy ha delle code di legge di potenza e il ritorno alla media è superlineare all'infinito. In questo caso l'esponente di scaling è lineare a tratti.

Parole chiave: ipoellitticità, condizione di Hörmander, stime di tubo per i processi di Ito, stime di densità, volatilità stocastica, multiscaling.

Contents

Chapter 1

Introduction

1.1 Regularity of Stochastic Differential Equations

The investigation of regularity properties of solutions of Stochastic Differential Equations (SDEs) driven by the Brownian motion is one of the motivations and most important applications of the Malliavin calculus on the Wiener space. Consider the diffusion process in $\mathbb{R}^n$ solution of

$$dX_t = \sum_{j=1}^{d} \sigma_j(X_t)\, dW_t^j + b(X_t)dt, \qquad X_0 = x, \tag{1.1.1}$$

where $W = (W^1, ..., W^d)$ is a d-dimensional Brownian motion, $b, \sigma_j \in C_b^\infty(\mathbb{R}^n, \mathbb{R}^n)$ for $j = 1, \dots d$. Denoting $(\sigma)_{i,j} = \sigma_j^i$, we say that (1.1.1) is *uniformly elliptic* if there exist $c > 0$ such that $\sigma\sigma^T(y) \geq cId_n$, for all $y \in \mathbb{R}^n$, where we denote with σ^T the transpose of σ and Id_n the $n \times n$ identity matrix. It is well known that under these assumptions (1.1.1) admits a unique strong solution for every initial condition x, and that the following two-sided bound holds for the fixed-time marginals. Fix a time horizon $T > 0$.

Theorem 1.1. *For every initial condition $x \in \mathbb{R}^n$ and every $0 < t \leq T$, the law of X_t is absolutely continuous with respect to the Lebesgue measure, and its density in y, $p_t(x, y)$, is infinitely differentiable with respect to each variable. Moreover, there exist constants $c_{0,T}, C_{0,T} \in \mathbb{R}^+$ and functions $k_0, K_0 : \mathbb{R}^+ \to \mathbb{R}^+$ such that*

$$k_0(t) \exp\left(-\frac{c_{0,T}|y - x|^2}{t}\right) \leq p_t(x, y) \leq K_0(t) \exp\left(-\frac{C_{0,T}|y - x|^2}{t}\right), \tag{1.1.2}$$

where $|\cdot|$ denotes the Euclidean norm.

This result can be proved with PDE's methods (see for example [2]) or with Malliavin calculus techniques. Being able to investigate regularity properties of the law of solutions of SDEs was the original motivation for developing the theory of Malliavin calculus, and since then many other applications has been considered. The Malliavin derivative permits to quantify the sensitivity of the system with respect to the noise, and if the so-called "Malliavin covariance matrix" is non-degenerate, the sensibility is non-zero in any direction. Then X_t admits a density, and it is in some cases possible to obtain bounds for it.

Besides Malliavin, the foudations of this theory were laid by Stroock, Bismut, Watanabe, Ikeda, Shigekawa and others. We mention here [74] and the series of papers by Kusuoka and Stroock [70], [71], [72].

It is clear that these assumptions on the coefficients of (1.1.1) are quite demanding. Uniform ellipticity, for instance, implies that d, the dimension of the Brownian Motion, is greater or equal to n, the dimension of the diffusion. One way to relax the ellipticiy condition is to assume the *Hörmander condition*, a celebrated *hypoelliptic* condition which is one of the key points of our work. For $f, g : \mathbb{R}^n \to \mathbb{R}^n$ we recall the definition of the directional derivative of f in the direction g as

$$\partial_g f(x) = \nabla f\, g(x) = \sum_{i=1}^{n} g^i(x)\partial_{x_i} f(x).$$

The Lie bracket $[f, g]$ in x is then defined as

$$[f, g](x) = \partial_f g(x) - \partial_g f(x).$$

We denote now $\sigma_0 = b$, and say that the Hörmander condition holds at point $x \in \mathbb{R}^n$ if the vector fields

$$\sigma_1, \dots, \sigma_d, \quad [\sigma_i, \sigma_j],\, 0 \le i, j \le d, \quad [\sigma_i, [\sigma_j, \sigma_k]], 0 \le i, j, k \le d, \dots \tag{1.1.3}$$

span $\mathbb{R}^n$ at x. Under this condition the following theorem holds (see [79], Theorem 2.3.3).

Theorem 1.2. *Assume that* $(X_t)_{t>0}$ *is the solution to* (1.1.1)*. Suppose that the coefficients* $\sigma_1, \dots, \sigma_d, b$ *are infinitely differentiable with bounded partial derivatives of all orders and that Hörmander's condition* (1.1.3) *holds. Then for any* $t > 0$ *the random vector* X_t *has an infinitely differentiable density.*

This result can be considered as a probabilistic version of Hörmander's theorem on the hypoellipticity of the second-order differential operators. We refer to [98], [79] for details regarding this interpretation. For many related issues see [64], [92], [79], [90], [96] and the bibliography there. Similar results are also available for SDEs with coefficients with dependence on time, under very weak regularity assumptions ([34]), SDEs driven by a fractional Brownian Motion ([15]) and for rough differential equations ([33]). Besides the existence and differentiability of the density, the problem of finding bounds for the density of X_t in the spirit of (1.1.2) has been widely addressed in the literature, also in a hypoelliptic framework. One outstanding contribution is certainly the work of Kusuoka and Stroock. Denote with α a multi-index $(\alpha_1, ..., \alpha_k) \in$

$\{1, ..., n\}^k$, and with $|\alpha|$ the lenght of α. In [71], supposing (1.1.3), the following upper bounds for the density and its derivatives are proved:

$$\begin{aligned} p_t(x,y) &\leq K_0(t) \exp\left(-\frac{C_{0,T}|y-x|^2}{t}\right), \\ |\partial_y^\alpha p_t(x,y)| &\leq K_{|\alpha|}(t) \exp\left(-\frac{C_{|\alpha|,T}|y-x|^2}{t}\right), \end{aligned} \tag{1.1.4}$$

where the constant $C_{|\alpha|,T}$ and the function $K_{|\alpha|}(\cdot)$ depend on how many iterated Lie Brackets we need to take in (1.1.3) to span $\mathbb{R}^n$. Remark that in this case we have only the upper bound, whereas if the diffusion is elliptic an analogous lower bound holds.

A two-sided bound for the density of X_t is proved in [72], under a more demanding condition. Consider (1.1.3), but imagine that we are not allowed to take brackets involving the drift vector field $b(\cdot) = \sigma_0(\cdot)$. More formally, suppose that $\mathbb{R}^n$ is spanned by

$$\sigma_1, \dots, \sigma_d, \quad [\sigma_i, \sigma_j],\, 1 \leq i,j \leq d, \quad [\sigma_i, [\sigma_j, \sigma_k]], 1 \leq i,j,k \leq d, \dots \tag{1.1.5}$$

This hypothesis is usually referred to as *Strong* Hörmander condition, and in opposition to this (1.1.3) is often called *Weak* Hörmander condition. In [72] the strong Hörmander non-degeneracy condition is assumed, and moreover the drift b is supposed to be generated by the vector fields of the diffusive part, i.e.: $b(x) = \sum_{j=1}^d \alpha_i \sigma_i(x)$, with $\alpha_i \in C_b^\infty(\mathbb{R}^n)$.

Under this assumption, a Gaussian bound is proved in the control distance that we now define. For $x, y \in \mathbb{R}^n$ we denote by $C(x,y)$ the set of controls $\psi \in L^2([0,1]; \mathbb{R}^d)$ such that the corresponding skeleton solution of

$$du_t(\psi) = \sum_{j=1}^d \sigma_j(u_t(\psi))\psi_t^j dt, \qquad u_0(\psi) = x$$

satisfies $u_1(\psi) = y$. The control (Caratheodory) distance is defined as

$$d_c(x,y) = \inf\left\{ \Big(\int_0^1 |\psi_s|^2\, ds \Big)^{1/2} : \psi \in C(x,y) \right\}.$$

Geometrically speaking, this corresponds to take the geodesic (i.e. the length-minimizing curve) joining x and y on the sub-Riemannian manifold associated with the diffusion coefficient σ. The main result in [72] is the following: there exist a constant $M \geq 1$ such that

$$\begin{aligned} &\frac{1}{M|B_{d_c}(x, t^{1/2})|} \exp\left(-\frac{M d_c(x,y)^2}{t}\right) \\ &\qquad \leq p_t(x,y) \leq \frac{M}{|B_{d_c}(x, t^{1/2})|} \exp\left(-\frac{d_c(x,y)^2}{Mt}\right) \end{aligned} \tag{1.1.6}$$

for $(t, x, y) \in (0,1] \times \mathbb{R}^n \times \mathbb{R}^n$, where $B_d(x,r) = \{y \in \mathbb{R}^n : d(x,y) < r\}$.

Again under a strong Hörmander condition, Ben Arous and Léandre investigate the decay of the heat kernel of a hypoelliptic diffusion over the diagonal in their celebrated paper [18]. An important difference here is that the authors are interested in asymptotic results, whereas the works previously mentioned provide results holding for finite positive times.

Loosely speaking, strong Hörmander means that we take advantage of the noise propagating in the system through the vector fields of the diffusion and their Lie brackets, whereas when the non-degeneracy is of weak Hörmander type the drift gives an additional specific contribution which is usually difficult to handle when trying to estimate the desity of the solution. Some related works are [11] and [42], where bounds are provided for the density of the Asian type SDE and for a chain of SDEs, and [63], where a stable driven degenerate SDE is considered. An analytical approach to similar density estimates is given by Polidoro, Pascucci and Boscain in [85], [80], [24].

In this thesis we consider a problem which is closely related to the issues mentioned above: the so-called tube estimates, i.e. estimates on the probability that an Ito process remains around a deterministic path up to a given time. More precisely, we consider (1.1.1) and introduce the associated skeleton path solution of the following ODE:

$$dx_t(\phi) = \sum_{j=1}^{d} \sigma_j(x_t(\phi))\phi_t^j dt + b(x_t(\phi))dt, \qquad x_0(\phi) = x, \tag{1.1.7}$$

for a certain control $\phi \in L^2([0,T],\mathbb{R}^d)$. With tube estimate we mean that we are interested in $\mathbb{P}\left(\sup_{t\leq T} \|X_t - x_t(\phi)\| \leq R\right)$. Several works have considered this subject, starting from Stroock and Varadhan in [94], where such result is used to prove the support theorem fo diffusion processes. In their work $\|\cdot\|$ is the Euclidean norm, but later on different norms have been used to take into account the regularity of the trajectories (about this, see for example [16] and [53]). This kind of problems are also related to the Onsager-Machlup functional, large and moderate deviation theory: see e.g. [30], [64], [58].

There is a strong connection between tube and density estimates. In this work we will use a concatenation of short time density estimates to prove a tube estimate, but one may proceed in reverse order: tubes estimates, for instance, easily give lower bounds for the density. In [10] tube estimates for locally elliptic diffusions with time dependent radius are proved, and applied to find lower bounds for the probability to be in a ball at fixed time and bounds for the distribution function. In [9], this is applied to lognormal-like stochastic volatility models, finding estimates for the tails of the distribution, and estimates on the implied volatility.

1.2 Outline of results: Part I

In **Chapter 2** we work in a fairly abstract framework, applying general Malliavin Calculus techniques. We obtain some estimates for the density of a random variable, based on the fact that we can measure the distance between the densities of the laws of two random variables F and G in terms of the Malliavin-Sobolev norm of $F - G$, under some non-degeneracy conditions.

In **Chapter 3** we consider a two-dimensional diffusion X driven by a one-dimensional Brownian Motion W, in a (local) weak Hörmander framework. As we already mentioned, similar models have been considered in the literature ([11], [42], [85], [80], [24]). Here we allow for a more general coefficient for the Brownian Motion, since we suppose $\partial_\sigma \sigma(y) = \kappa_\sigma(y)\sigma(y)$, whereas the works mentioned above would apply for $\sigma = (\sigma_1, 0)$, which is a more restrictive condition. We prove a short time non-asymptotic density estimate, which allows us to find exponential lower and upper bounds for the probability that the diffusion remains in the tube. For this result we only need local hypoellipticity along the control, which is also a novelty of our work.

Since we work under Hörmander-type conditions, in order to give accurate estimates it is necessary to consider some norms accounting for the non-diffusive time scale of the process. Indeed, thanks to the Hörmander condition, the noise propagates in the whole $\mathbb{R}^2$, but with different speeds in the directions generated by the diffusion coefficient and the Lie brackets. With this purpose we define the following norm: let M be a $n \times m$ matrix with full row rank. We write M^T for the transposed matrix, and since MM^T is invertible we can set, for $y \in \mathbb{R}^n$,

$$|y|_M = \sqrt{\langle (MM^T)^{-1} y, y \rangle}. \tag{1.2.1}$$

For any $R > 0$, we denote with $A_R(x)$ the matrix $\left(R^{1/2}\sigma(x), R^{3/2}[\sigma, b](x)\right)$. We assume the following weak Hörmander condition: σ, $[\sigma, b]$ span $\mathbb{R}^2$ locally around $x(\phi)$ (given by (1.1.7)), and so $A_R(x_t(\phi))$ is invertible, and we can define $|\cdot|_{A_R}$ as in (1.2.1). We prove, for small δ and y in a neighborhood of $x + b(x)\delta$,

$$\frac{K_1}{\delta^2} \exp\left(-L_1 |y - x - b(x)\delta|^2_{A_\delta(x)}\right) \le p_{X_\delta}(y) \le \frac{K_2}{\delta^2} \exp\left(-L_2 |y - x - b(x)\delta|^2_{A_\delta(x)}\right).$$

We stress that our density estimates are in short time, but differently from [17], where asymptotics for the heat kernel are proved, our result holds for finite, positive time. This is crucial to prove the following tube estimate: for small R

$$\begin{aligned} &\exp\left(-C_T \int_0^T \left(\frac{1}{R} + |\phi_t|^2\right) dt\right) \le \\ &\qquad \mathbb{P}\left(\sup_{t \le T} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} \le 1\right) \le \\ &\qquad\qquad \exp\left(-\frac{1}{C_T} \int_0^T \left(\frac{1}{R} + |\phi_t|^2\right) dt\right). \end{aligned} \tag{1.2.2}$$

In this chapter we also prove a diagonal density estimate in short time for the chain of stochastic differential equations considered in [42], under local hypoellipticity conditions, which is consistent with the main result of the original paper.

In **Chapter 4** we consider a SDE in dimension n, assuming a (local) strong Hörmander condition of order one, meaning that we suppose that $\{\sigma_i, [\sigma_j, \sigma_p] : i, j, p = 1, ..., d\}$ span $\mathbb{R}^n$ locally around the control $x(\phi)$. Here we allow for a general drift, and we work with time-dependent coefficients, differently from [71], [72] (see [34] for the regularity of the law of X_t with time dependent-coefficients).

We define $A_R(t,x) = (\sqrt{R}\sigma_i(t,x), R[\sigma_j, \sigma_p](t,x))_{i,j,p=1,\ldots,d,j\neq p}$ and find, for small R, a tube estimates analogous to (1.2.2) in the $|\cdot|_{A_R(t,x_t(\phi))}$ norm (remark that here we have also the dependence on t in the matrix). Also in this case the proof is based on a diagonal density estimate in short time. In general, one application of tube estimates is to prove lower bound for the density. This looks particularly interesting in this case since the classical lower bound in [72] holds under much more demanding hypothesis on the drift coefficient (b is generated by the vector fields of the diffusive part).

We also establish a link between the norm $|\cdot|_M$ and the control-Caratheodory distance (see [78] for classical results). In the strong Hörmander case we establish a local equivalence that allows us to write the tube estimate in the control distance. This is interesting again in comparison with [72], where Gaussain bounds are proved, but only for a very specific form of the drift. In the weak Hörmander case this equivalence cannot hold, since the definition of the standard control distance involves only σ, which is not enough to span the whole space even if we consider its Lie brackets. We nevertheless introduce a suitable control distance and prove a local equivalence with the matrix norm $|\cdot|_M$.

1.3 Modeling Stylized Facts of Financial Markets

In the last few decades a number of researchers has shown an increasing interest in the field of economics and finance and their links with statistical mechanics. Many interesting phenomena arise when looking at financial data with mathematical tools coming from statistical physics, this being motivated by the fact that a financial market is somehow analogous to a physical "complex system", being the reflexion of the interactions of a huge number of agents. What we are looking at is not the behavior of the single agent, but some macroscopic quantity that we consider important. This new viewpoint has led to the discovery of some striking empirical properties, detected in various types of financial markets, considered now as *stylized facts* of these markets. Examples of such facts are scaling properties, auto-similarity, properties of the volatility profile, such as peaks and clustering, and long range dependence (see [37]).

In the second part of this thesis we deal with such phenomena, but the point of view we adopt comes from mathematical finance more than statistical mechanics. We do not look at the microscopic behavior, but directly at the macroscopic quantities mentioned above. We can suppose that the large-scale phenomena under study have their origin in some small-scale interactions, but what we try to model here is just the macroscopic world. For this purpose we put ourselves in the framework of continuous-time stochastic volatility models. More precisely, the market models that will be used in this part are mean reverting stochastic volatility models, with a volatility driven by a jump process. This means that the process for the detrended log-price evolves through $dX_t = \sigma_t dB_t$, where B is a Brownian motion and the volatility $\sigma_t = \sqrt{V_t}$ is the square root of the stationary solution of a SDE of the following form:

$$dV_t = -f(V_t)dt + dL_t. \tag{1.3.1}$$

The function f, what we call "mean reversion", has the role of pushing the volatility back to a certain equilibrium value, whereas $L = (L_t)_t$ is a process which models the noise in the volatility, and it is often taken as a Levy process (see [50], [67], [12]). If V is independent of B and has paths in $L^2_{loc}(\mathbb{R})$ a.s., the process X can be seen as a random time-change of a Brownian motion:

$$X_t = W_{I(t)}, \tag{1.3.2}$$

where $I(t) = \int_0^t V_s ds$ is sometimes called *trading time*. An example of such process is the model presented in [1], which accounts a number of the stylized facts mentioned above, namely: the crossover in the log-return distribution from a power-law to a Gaussian behavior, the slow decay in the volatility autocorrelation, the diffusive scaling and the multi-scaling of moments, while keeping a simple formulation and an explicit dependence on the parameters. In this thesis we analyze cross-correlations of log returns for a bivariate version of this model. The correlation of both increments and absolute increments of two assets at a certain time has been widely studied, especially because of its direct link with systemic risk and portfolio management (see for instance [48], [25]). It is also possible to consider the cross-correlation of absolute increments at different times, and compute how this correlation decays as the time difference increases. This issue has been addressed by Podobnik et al. in [83], where they analyze the Dow Jones industrial and the S&P500 indices, and in [84], [97], where long range cross-correlations between magnitudes are found in a number of studies including nanodevices, atmospheric geophysics, seismology and finance. This aspect, together with the analysis of an algorithm for the detection of peaks in the volatility profile, will be the focus of Chapter 5.

In Chapter 6 we consider a generalization of the model in [1], and focus on one of the stylized facts mentioned above, namely the *multi-scaling of moments*. Let $(X_t)_{t\geq 0}$ be some stochastic process representing the detrended *log-price* of an asset. We say that the multi-scaling of moments occurs if $\mathbb{E}\left(|X_{t+h} - X_t|^q\right)$ scales, in the limit as $h \downarrow 0$, as $h^{A(q)}$, with *scaling exponent* $A(q)$ non-linear. This pattern is rather systematically observed in time series of financial assets ([95, 56, 54, 44, 43]).

A class of stochastic processes that exhibit multi-scaling for a rather arbitrary scaling function $A(q)$ are the so-called *multifractal models* ([28, 27, 26]). In these models, the process X_t is given as the random time change of a Brownian motion (as in (1.3.2)), where $I(t)$ is a stochastic process, often taken to be independent of W, with continuous and increasing trajectories. Modeling financial series through a random time-change of Brownian motion is a classical topic, dating back to Clark ([35]), and reflects the natural idea that external information influences the speed at which exchanges take place in a market. In multifractal models, the trading time

$I(t)$ is a process with *non absolutely continuous* trajectories. As a consequence, X_t cannot be written as a stochastic volatility model, and this makes the analysis of multifractal models hard in many respects, as the standard tools of Ito's Calculus cannot be applied.

The model constructed in [1] exhibits a bi-scaling behavior, meaning that $A(q)$ is piecewise linear and the slope $A'(q)$ takes two different values, which suffices to fit most of the cases observed. This process, in opposition to general multi-fractal models, is a stochastic volatility model. In Chapter 6 we will prove that this behaviour is common to a much wider class of stochastic volatility models of the form (1.3.1).

1.4 Outline of results: Part II

Chapter 5 is based on [22]. We consider the model introduced in [1], addressing the two following issues: a study of a bivariate version of the model, and an algorithm for the detection of peaks in the volatility profile. These two aspects are linked by the fact that the cross-correlation between the magnitude of the increments of two indices is highly dependent on the jumps of the volatility process.

We find an explicit formula for the decay of cross-asset correlations between absolute returns depending on the time lag, analogous to the formula for the decay of autocorrelations. We then apply this result to the time series of the Dow Jones Industrial Average (DJIA) and the Financial Times Stock Exchange (FTSE) 100, from 1984 to 2013, finding an excellent agreement between predictions of the model and empirical findings. In particular we find that in both modeling and empirical data the decay of autocorrelations and cross-correlations is almost coincident, and it is slow over time, confirming that this is a long-memory processes.

In our model cross-correlation is highly dependent on the jumps of the volatility process, and for this reason we propose here an algorithm for the detection of jumps in the volatility, which shares some features with the commonly used ICSS-GARCH algorithm, but works well under less demanding assumptions. We prove formally some results justifying the convergence, and some heuristic considerations on the output of the algorithm confirm its validity in the detection of jumps.

The fact that our model displays a behavior that is completely analogous to real data in all of these aspects, even the most subtle, is an interesting validation of our model. We mention that the satistical analysis performed by Podobnik et al. in [83], [97] lead to results analogous to ours, concerning also the similarity in the decay of autocorrelations and cross-asset correlations.

In **Chapter 6**, following [41], we analyze multi-scaling of moments in the class of models $dX_t = \sigma_t dB_t$, with a volatility process σ_t independent of the Brownian motion B_t: these processes are exactly those that can be written in the form (1.3.2) with a trading time $I(t)$ independent of W, and with *absolutely continuous* trajectories. We say the multi-scaling of moments occurs if the limit

$$\limsup_{h \downarrow 0} \frac{\log \mathbb{E}\left(|X_{t+h} - X_t|^q\right)}{\log h} =: A(q)$$

is *non-linear* on the set $\{q \geq 1 : |A(q)| < +\infty\}$. Since it is reasonable to expect $A(q) = \frac{q}{2}$, multi-scaling of moments can be identified with deviations from this *diffusive* scaling. This is exactly what happens for the empirical time series of many financial indices, for q above a given threshold.

We devote special attention to models in which $V_t := \sigma_t^2$ is a *stationary* solution of a stochastic differential equation of the form(1.3.1), for a *Levy subordinator* L_t whose characteristic measure has *power law tails at infinity*. We first show multi-scaling is *not* possible if $f(\cdot)$ has *linear growth*, and therefore the heavy tails produced by the Levy process are not sufficient to produce multi-scaling. On the other hand, we show that, if $f(\cdot)$ behaves as Cx^γ as $x \to +\infty$, with $C > 0$ and $\gamma > 1$, then the stochastic volatility process whose volatility is a stationary solution of (1.3.1) exhibits multi-scaling (see Theorem 6.10 for the precise result). The class of processes considered in this paper are also of the form (1.3.2), where the time change $I(t) = \int_0^t V_s ds$ has absolutely continuous paths. This structure, although quite restrictive, provides considerable advantages when these model are used in Mathematical Finance, e.g. in option pricing. A further feature of these model is that the scaling function $A(q)$ is piecewise linear, with two values for the slope, providing a good fit for most of the observed cases, as remarked before.

1.5 Perspectives

We conclude this introduction pointing out some possible developments of the work presented in this thesis.

The theoretical starting point of the density estimates presented in *Part I* is the recent series of papers by Bally and Caramellino [5, 7, 8], where a variant of Malliavin calculus is developed, in order to obtain estimates of the distance between the density function of two random variables in terms of the inverse of their Malliavin covariance matrix and of Sobolev type norms. In our work we use this type of techniques to estimates the law of solutions of SDEs under some specific non-degeneracy conditions. An interesting example of application is the system of coupled oscillators analyzed in [42] (see also the related models 5.1, 5.1). In [42] Delarue and Menozzi provide Gaussian estimates for the density of the solution, using parametrix and control methods. In section 5.1, we find similar estimates with the Malliavin calculus techniques mentioned above, but it is reasonable to expect that our techniques may produce estimates for the derivatives of the density as well. We believe this would be an interesting continuation of our work, in particular for the specific model in [42].

Another aspect which is worth considering concerns the development of higher order schemes for the simulation of degenerate SDEs. The proofs of our density estimates require a stochastic Taylor development of the diffusion, which brings out a geometric structure associated with the equation, linked to the degeneracy of the covariance matrix of the process and to the different speeds of the dynamic in the different directions. We think it may be possible to use this structure to implement higher order simulation methods, which may apply in particular to the pricing of

Asian Options, a remarkable example of degenerate diffusion in the weak-Hörmander framework. Another model which would possibly be concerned, in neuroscience, is the "Stochastic Hodgkin-Huxley" (see [61, 62]). Some simulations presented in [61] suggest a possible connection between the weak non-degeneracy and the spiking behavior of the system, and this would certainly be an interesting topic to investigate, also in connection with numerical aspects.

A more distant but possible development could be the generalization of the abstract techniques presented in chapter 2. Indeed, it may be possible to develop analogous estimates for different kind of stochastic differential equations. The first reasonable attempt would be to try an extension to equations driven by a fractional Brownian Motion. This might produce some better understanding of the speed/scaling of the solution of such equations.

We look now at *Part II*. For the model presented in chapter 5, pricing and implied volatility issues have been considered in [31, 32], but for the generalization of chapter 6 many applied aspects are still open. In [31] the asymptotic behavior of the call and put option prices is explicitly linked to the tail probabilities $\mathbb{P}(|X_t| > \kappa)$, where X_t denotes the risk-neutral (or detrended) log-return. When this is applied to the stochastic volatility model in [1], an interesting property concerning the divergence of the implied volatility appears, which is usually obtained only in presence of jumps in the price. Being related to the tail probabilities, this fact looks naturally connected to the multiscaling of moments, and therefore we believe it would be worth considering implied volatility also for the generalization introduced in chapter 6. Besides this precise aspect, option pricing, calibration and many other applied financial issues are still to explore.

Another possible direction to take would be to suppose that the driving Lévy is an α-stable process, and try to find more explicit theoretical results, for instance concerning the stationary distribution of the volatility, or the asymptotics of the distribution in short or long time, similarly to what is done in [49].

A much deeper problem to address would be to try to understand how the superlinear mean reversion, which is a key feature of these systems, could be produced. This phenomenon is observed for instance in aggregated indices, so a possible origin could lie in the fact that we are averaging on a large number of stocks, whose mutual correlations follow dynamics that are not clearly understood. In particular, the superlinear mean reversion that appears in our model is reminiscent of the asymptotic dynamics of critical fluctuations (see [36]) of various models with mean-field interaction, although this analogy is quite vague.

Part I

Tube and density estimates for hypoelliptic diffusion processes

Chapter 2

Malliavin calculus and density estimates for random variables

In this chapter we work in the framework of Malliavin calculus, for which our main reference is [79]. We present some techniques for obtaining quantitative estimates of the density of a random variable on the Wiener space, based on the recent work of Bally and Caramellino, developed in [7], [8]. For some computations we also refer to [81]. Such estimates can be used to study the behavior of a diffusion in short time, and we give various applications in this sense in Chapter 3 and 4. These techniques are based on the fact that we can measure the distance between the densities of the laws of two random variables F and G in terms of the Malliavin-Sobolev norm of $F - G$, under a non degeneracy condition for F and G. Using this fact, if we can approximate F with some proxy G with explicit law, we are able to recover an estimate for the density of F. We apply it in particular to diffusions, and in that case our G is some main Gaussian component that we put in evidence trough a stochastic Taylor development. The main result of this section (cf. Theorem 2.4) consists in a two sided bound for p_F in terms of a localization of p_G, where the lower bound involves the Malliavin covariance of G and the upper bound involves the Malliavin covariance of F. When we apply this result to diffusions in short time in Chapter 3 and 4, this leads to very different technical difficulties when dealing with the lower or the upper bound.

2.1 Elements of Malliavin Calculus

2.1.1 Notations

We recall some basic notions in Malliavin calculus. Our main reference is [79]. We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a Brownian motion $W = (W_t^1, ..., W_t^d)_{t\geq 0}$ and the filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by W. For fixed $T > 0$, we denote with $\mathcal{H}$ the Hilbert space $L^2([0, T], \mathbb{R}^d)$. For $h \in \mathcal{H}$ we introduce this notation for the Itô integral of h: $W(h) = \sum_{j=1}^d \int_0^T h^j(s) dW_s^j$.

We denote by $C_p^\infty(\mathbb{R}^n)$ the set of all infinitely continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f and all of its partial derivatives have polynomial growth. We also denote by $\mathcal{S}$ the class of simple random variables of the form

$$F = f(W(h_1), ..., W(h_n)),$$

for some $f \in C_p^\infty(\mathbb{R}^n)$, $h_1, ..., h_n$ in $\mathcal{H}$, $n \geq 1$. The Malliavin derivative of $F \in \mathcal{S}$ is the $\mathcal{H}$ valued random variable given by

$$DF = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(W(h_1), ..., W(h_n))h_i.$$

We introduce the Sobolev norm of F:

$$\|F\|_{1,p} = [\mathbb{E}|F|^p + \|DF\|_{\mathcal{H}}^p]^{\frac{1}{p}}$$

where

$$\|DF\|_{\mathcal{H}} = \left(\int_0^T |D_s F|^2 ds\right)^{\frac{1}{2}}.$$

It is possible to prove that D is a closable operator an take the extension of D in the standard way. We can now define in the obvious way DF for any F in the closure of $\mathcal{S}$ with respect to this norm. Therefore, the domain of D will be the closure of $\mathcal{S}$.

The higher order derivative of F is obtained by iteration. For any $k \in \mathbb{N}$, for a multi-index $\alpha = (\alpha_1, ..., \alpha_k) \in \{1, ..., d\}^k$ and $(s_1, ..., s_k) \in [0, T]^k$, we can define

$$D^\alpha_{s_1,...,s_k} F := D^{\alpha_1}_{s_1} ... D^{\alpha_k}_{s_k} F.$$

We denote with $|\alpha| = k$ the length of the multi-index. Remark that $D^\alpha_{s_1,...,s_k} F$, is a random variable with values in $\mathcal{H}^{\otimes k}$, and so we define its Sobolev norm as

$$\|F\|_{k,p} = [\mathbb{E}|F|^p + \sum_{j=1}^{k} |D^{(j)} F|^p]^{\frac{1}{p}}$$

where

$$|D^{(j)} F|^2 = \left(\sum_{|\alpha|=j} \int_{[0,T]^k} |D^\alpha_{s_1,...,s_k} F|^2 ds_1 ... ds_k\right)^{1/2}.$$

The extension to the closure of $\mathcal{S}$ with respect to this norm is analogous to the first order derivative. We denote by $\mathbb{D}^{k,p}$ the space of the random variables which are k times differentiable in the Malliavin sense in L^p, and $\mathbb{D}^{k,\infty} = \bigcap_{p=1}^\infty \mathbb{D}^{k,p}$. As usual, we also denote with L the Ornstein-Uhlenbeck operator, i.e. $L = -\delta \circ D$, where δ is the adjoint operator of D.

2.1.2 Non-degeneracy

We consider random vector $F = (F_1, ..., F_n)$ in the domain of D. We define its *Malliavin covariance matrix* as follows:

$$\gamma_F^{i,j} = \langle DF_i, DF_j\rangle_{\mathcal{H}} = \sum_{k=1}^{d}\int_0^T D_s^k F_i \times D_s^k F_j ds.$$

We say that F is *non-degenerate* if its Malliavin covariance matrix is invertible and

$$\mathbb{E}(|\det \gamma_F|^{-p}) < \infty, \quad \forall p \in \mathbb{N}. \tag{2.1.1}$$

We denote with $\hat{\gamma}_F$ the inverse of γ_F.

2.1.3 Conditional Expectation

The following representation theorem for the conditional expectation has been proved in [5]. The same representation formula, under slightly more demanding hypothesis on the regularity, is a result by Malliavin and Thalmaier (see [75] and [92]).

Proposition 2.1. *Let $F = (F^1, \cdots, F^n)$ be such that $F^1, \cdots, F^n \in \bigcap_{p\in\mathbb{N}} \mathbb{D}^{2,p}$ and let $G \in \bigcap_{p\in\mathbb{N}} \mathbb{D}^{1,p}$. Assume* (2.1.1). *Then*

$$\mathbb{E}(G|F = x) = 1_{\{p_F>0\}}\frac{p_{F,G}(x)}{p_F(x)}$$

with

$$p_{F,G}(x) = \mathbb{E}\left[\sum_{i=1}^{n}\partial_i \mathcal{Q}_n(F - x)H_i(F; G)\right],$$

where $\mathcal{Q}_n$ denotes the Poisson kernel on $\mathbb{R}^n$, i.e. the fundamental solution of the Laplace operator $\Delta\mathcal{Q}_n = \delta_0$. This is given by

$$\mathcal{Q}_1(x) = \max(x, 0); \quad \mathcal{Q}_2(x) = \mathcal{A}_2^{-1}\ln|x|; \quad \mathcal{Q}_n(x) = -\mathcal{A}_n^{-1}|x|^{2-n},\ n > 2,$$

where $\mathcal{A}_n$ is the area of the unit sphere in $\mathbb{R}^n$. The Malliavin weights are given by

$$H(F; G) = G\hat{\gamma}_F \times LF - \langle D(\hat{\gamma}_F G), DF\rangle$$

Remark 2.2. With $G = 1$ this result gives a representation formula for the density:

$$p_F(x) = -\sum_{i=1}^{n}\mathbb{E}\left[\partial_i\mathcal{Q}_n(F - x)H_i(F; 1)\right] \tag{2.1.2}$$

2.1.4 Localization

The following notion of localization is introduced in [7]. Consider a random variable $U \in [0,1]$ and denote

$$d\mathbb{P}_U = U d\mathbb{P}.$$

We also denote

$$\|F\|_{p,U} = (\mathbb{E}_U(|F|^p))^{1/p} = (\mathbb{E}(|F|^p U))^{1/p}$$
$$\|F\|_{k,p,U} = \|F\|_{p,U} + \sum_{1\leq|\alpha|\leq k} (\mathbb{E}_U(|D^{(\alpha)}F|^p))^{1/p}.$$

We assume that $U \in \mathbb{D}^{1,\infty}$ and for every $p \geq 1$

$$m_U(k,p) := 1 + \sum_{i=1}^{k+1} (\mathbb{E}_U|D^i \ln U|^p)^{1/p} < \infty.$$

The specific localizing function we will use is the following. Consider the function depending on a parameter $a > 0$:

$$\psi_a(x) = 1_{|x|\leq a} + \exp\left(1 - \frac{a^2}{a^2-(x-a)^2}\right) 1_{a<|x|<2a}.$$

For $\Theta_i \in \mathbb{D}^{2,\infty}$ and $a_i > 0$, $i = 1\ldots,n$ we define the localization variable:

$$U = \prod_{i=1}^{n} \psi_{a_i}(\Theta_i) \tag{2.1.3}$$

For this choice of U we have that for any $p, k \in \mathbb{N}_0$

$$m_U(k,p) \leq C_{p,k}\left(1 + \sum_{i=1}^{k+1} \frac{\|D^i\Theta\|_p}{|a|}\right)^{k+1} \leq C_{p,k}\left(1 + \frac{\|\Theta\|_{k+1,p}}{|a|}\right)^{k+1} \tag{2.1.4}$$

and, for $k \geq 1$,

$$\begin{aligned} \|1-U\|_{k,p} &\leq C_{k,p}\left(\sum_i \mathbb{P}(|\Theta_i|^p > \alpha_i^p)^{1/(2p)}\right)\left(1 + \sum_{i=1}^{k} \frac{\|D^{(i)}\Theta\|_{2p}}{|a|^i}\right) \\ &\leq \frac{C_{k,p}}{|a|^k}\|\Theta\|_{k,2p} \end{aligned} \tag{2.1.5}$$

The proof of (2.1.4) follows from standard computations and inequality

$$\sup_x |(\ln \psi_a)^{(k)}(x)|^p \psi_a(x) \leq \frac{C_{k,p}}{a^{kp}} < \infty, \quad k = 1, 2, \ldots \tag{2.1.6}$$

To prove (2.1.5) we use again (2.1.6) and Markov inequality. For $F = (F^1, \cdots, F^n)$ such that $F^1, \cdots, F^n \in \bigcap_{p\in\mathbb{N}} \mathbb{D}^{2,p}$ and $V \in \bigcap_{p\in\mathbb{N}} \mathbb{D}^{1,p}$, for any localization function U we introduce the localized Malliavin weights

$$H_{i,U}(F,V) = \sum_{j=1}^{n} V\hat{\gamma}_F^{i,j} LF^j - \langle D(V\hat{\gamma}_F^{i,j}), DF^j\rangle - V\hat{\gamma}_F^{i,j}\langle D\ln U, DF^j\rangle$$

and the vector

$$H_U(F,V) = (H_{i,U}(F,V))_{i=1,\dots n}\,.$$

The following *representation formula for the density*, which is the localized version of (2.1.2), has been proved in [5].

Theorem 2.3. *Let U be a localizing r.v. such that under $\mathbb{P}_U$ (2.1.1) holds, i.e.*

$$\mathbb{E}_U[|\det\gamma_F|^{-p}] < \infty, \quad \forall p \in \mathbb{N}.$$

Then, under $\mathbb{P}_U$ the law of F is absolutely continuous and has a continuous density $p_{F,U}$ which may be represented as

$$p_{F,U}(x) = -\sum_{i=1}^{n} \mathbb{E}_U[\partial_i \mathcal{Q}_n(F-x)H_{i,U}(F,1)] \tag{2.1.7}$$

2.2 Density estimates

We discuss some techniques, based on Malliavin calculus, for estimating the density of a random variable. These ideas are based on the recent work of Bally and Caramellino, see for instance [7], [8].

2.2.1 The distance between two local densities

In what follows for a given matrix A we consider its Frobenius norm, given as

$$\|A\|_F = \sqrt{\sum_{i,j}|A_{i,j}^2|} = \sqrt{Tr(A^TA)}.$$

We will employ the fact that the Frobenius norm is sub-multiplicative. Take a square $d\times d$ matrix γ, symmetric and positive definite. Denote with $\lambda^*(\gamma)$ and $\lambda_*(\gamma)$ the largest and the smallest eigenvalues of γ. From the equivalence between Frobenius and spectral norm we have

$$\lambda^*(\gamma) \le \|\gamma\|_F \le \sqrt{d}\lambda^*(\gamma).$$

Denoting $\hat{\gamma} = \gamma^{-1}$, it holds $\lambda^*(\hat{\gamma}) = 1/\lambda_*(\gamma)$. So

$$\frac{1}{\lambda_*(\gamma)} \le \|\hat{\gamma}\|_F \le \frac{\sqrt{d}}{\lambda_*(\gamma)}.$$

For two time dependent matrices A_s, B_s, we have the following "Cauchy-Schwartz" inequality:

$$|\int A_s B_s ds|_F^2 \leq \int |A_s|_F^2 ds \int |B_s|_F^2 ds.$$

In particular, if $B_s = v_s$ is a vector,

$$|\int A_s v_s ds|^2 \leq \int |A_s|_F^2 ds \int |v_s|^2 ds.$$

We fix some notation. Let W be a Brownian Motion in $\mathbb{R}^d$. For two random variables $F = (F_1, \dots F_n)$, $G = (G_1, \dots G_n)$ and a localization function U, we denote

$$\begin{aligned}
\Gamma_{F,U}(p) &= 1 + \left(\mathbb{E}_U \lambda_*(\gamma_F)^{-p}\right)^{1/p} \\
\Gamma_{F,G,U}(p) &= 1 + \sup_{0\leq\varepsilon\leq 1} \left(\mathbb{E}_U \lambda_*(\gamma_{G+\varepsilon(F-G)})^{-p}\right)^{1/p} \\
n_{F,G,U}(k,p) &= 1 + \|F\|_{k+2,p,U} + \|G\|_{k+2,p,U} + \|LF\|_{k,p,U} + \|LG\|_{k,p,U} \\
\Delta_j(F,G) &= \sum_{i=1}^{j} |D^{(i)}(F-G)| + \sum_{i=0}^{j-2} |D^{(i)} L(F-G)|, \quad j \geq 2.
\end{aligned}$$

When $U = 1$, i.e. the localization is "trivial", we omit it in the notation. We also write $n_{F,U}(k,p)$ for $n_{F,0,U}(k,p)$. Since we are differentiating with respect to a Brownian Motion, as a direct consequence of Meyer's inequality (see for instance [79]), we have

$$n_{F,G,U}(k,p) \leq 1 + C\left(\|F\|_{k+2,p} + \|G\|_{k+2,p}\right). \tag{2.2.1}$$

Also notice that

$$n_{F,G,U}(0,p) \leq 2n_{F,U}(0,p) + \|\Delta_2(F,G)\|_{2,p}.$$

We are now able to give the main result of this section, which consists in a lower bound for p_F which does not involve the Malliavin covariance of F, and an upper bound for p_F which does not involve the Malliavin covariance of G.

Theorem 2.4. *Take $F, G \in \mathbb{D}^{3,32n}$. Suppose $m_U(1,32n) < \infty$ and $n_{F,G}(1,32n) < \infty$. If $\Gamma_{G,U}(32n) < \infty$, there exists a constant C_1 such that*

$$p_{G,U}(y) - C_1 \|\Delta_2(F,G)\|_{32n,U} \leq p_{F,U}(y) \leq p_F(y)$$

If $\Gamma_F(32n) < \infty$, there exists a constant C_2 such that

$$p_F(y) \leq p_{G,U}(y) + C_2(\|\Delta_2(F,G)\|_{32n,U} + \|1-U\|_{1,14n})$$

Remark 2.5. We can take

$$\begin{aligned}
C_1 &= C\left[m_U(1,32n)\Gamma_{G,U}(32n) n_{F,G,U}(1,32n)\right]^{24n^2} \\
C_2 &= C\left[m_U(1,32n)\Gamma_F(32n) n_{F,G}(1,32n)\right]^{24n^2}
\end{aligned}$$

where C is a universal constant depending only on the dimension n.

The lower bound for $p_{F,U}$ is a version of Proposition 2.5. in [7], where here we have specified as possible choice for the exponent $p = 32n$. Similar estimates can be found also in [8].

Proof. We first need an estimate for the localized Malliavin weights and for the difference of weights:

Lemma 2.6. *Let k and p be given. There exists a constant C depending just on p and the dimension n such that*

$$\|H_U(F,V)\|_{k,p,U} \le C\|V\|_{k+1,p_1} m_U(k,p_2)\Gamma_{F,U}(p_3)^{k+2} n_{F,U}(k,p_4)^{2k+3} \tag{2.2.2}$$

for every p_i, $i = 1,\dots,4$ with $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{k+2}{p_3} + \frac{2k+3}{p_4}$. Moreover

$$\begin{aligned} &\|H_U(F,V) - H_U(G,V)\|_{k,p,U} \\ &\le C\|V\|_{k+1,p_1} m_U(k,p_2)\Gamma_{F,G,U}(p_3)^{k+3} n_{F,G,U}(k,p_4)^{2k+4}\|\Delta_{k+2}(F,G)\|_{p_5,U}. \end{aligned} \tag{2.2.3}$$

for every p_i, $i = 1,\dots,5$ with $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{k+3}{p_3} + \frac{2k+4}{p_4} + \frac{1}{p_5}$.

Proof. Consider the weight:

$$H_U(F,V) = V[\hat\gamma_F \times LF - \langle D\hat\gamma_F, DF\rangle] - \langle \hat\gamma_F(DV + VD\ln U), DF\rangle \tag{2.2.4}$$

(1) We first consider $D\gamma_F$ and have the following estimate:

$$\begin{aligned} &\sum_{l=1}^{d}\int_0^t \|D_s^l\gamma_F\|_F^2 ds \\ &= \sum_{l=1}^{d}\int_0^t \left\| \left(\sum_{k=1}^{d}\int_0^t D_s^l D_u^k F_i \times D_u^k F_j + D_u^k F_i \times D_s^l D_u^k F_j du \right)_{i,j} \right\|_F^2 ds \\ &\le 4|D^{(2)}F|^2|DF|^2 \end{aligned}$$

(2) We now consider $D\hat\gamma_F$. From the chain rule and the derivative of the inversion of matrices,

$$D^k\hat\gamma_F = -\hat\gamma_F(D^k\gamma_F)\hat\gamma_F. \tag{2.2.5}$$

So, applying also the previous estimate

$$\sum_k \int \|D_s^k\hat\gamma_F\|_F^2 ds \le \|\hat\gamma_F\|_F^4 \sum_k \int \|D_s^k\gamma_F\|_F^2 ds \le 4\|\hat\gamma_F\|_F^4 |DF|^2|D^{(2)}F|^2.$$

Now the estimate of $\|H_U(F,U)\|_{p,U}$ follows easily applying Minkowski and Holder inequalities for L_p norms to (2.2.4).

The estimate of $\|H_U(F,U)\|_{k,p,U}$ follows using very similar techniques. The part giving the "main" contribution is $D^{(k+1)}\hat{\gamma}_F$, for which, iterating (2.2.5), it is not difficult to see

$$|D^{(k+1)}\hat{\gamma}_F| \le C(|DF| + \cdots + |D^{(k+2)}F|)^{2k+2} \, \|\hat{\gamma}_F\|_F^{k+2}$$

(Recall that $D^{(l)}$ means "derivative of order l" and D^l means "derivative with respect to W^l"). This term is also multiplied by $|DF|$, so we have the estimate of the term giving the main contribution. We leave out the similar estimate of the other terms.

When considering the difference $\|H_U(F,V) - H_U(G,V)\|_{k,p,U}$, we use similar arguments and the following property of norms: $|ab - cd| \le |a-c||b| + |c||b-d|$. As before the main contribution comes from $D^{k+1}(\hat{\gamma}_F - \hat{\gamma}_G)$, so we consider this and leave out the estimates of the other terms. We remark that

$$\hat{\gamma}_F - \hat{\gamma}_G = \hat{\gamma}_F(\gamma_G - \gamma_F)\hat{\gamma}_G$$

and differentiate this product, finding

$$\begin{aligned} &|D^{k+1}(\hat{\gamma}_F - \hat{\gamma}_G)| \le C(1 + \|\gamma_F\|_F \vee \|\gamma_G\|_F)^{k+3} \\ &\left(1 + \sum_{i=1}^{k+1} |D^{(i)}\gamma_F| \vee |D^{(i)}\gamma_G|\right)^{k+1} \sum_{i=0}^{k+1} |D^{(i)}(\gamma_F - \gamma_G)| \end{aligned}$$

where

$$\left(1 + \sum_{i=1}^{k+1} |D^{(i)}\gamma_F| \vee |D^{(i)}\gamma_G|\right)^{k+1} \le C\left(1 + \sum_{i=1}^{k+2} |D^{(i)}F| \vee |D^{(i)}G|\right)^{2k+2}$$

and

$$\|D^{(i)}(\gamma_F - \gamma_G)\|_F \le C \sum_{l=1}^{i+1} |D^{(l)}(F-G)| \sum_{l=1}^{i+1} |D^{(l)}(F+G)|.$$

Multiplying with $|DF|$, and applying Holder inequality, we prove the statement. $\square$

Lemma 2.7. *There exists a constant C depending just on the dimension n such that*

$$\begin{aligned} &|p_{F,U}(y) - p_{G,U}(y)| \\ &\quad \le C\left[m_U(1,32n)\Gamma_{F,G,U}(32n)n_{F,G,U}(1,32n)\right]^{12n^2} \|\Delta_2(F,G)\|_{32n,U} \end{aligned}$$

Proof. We write the densities using (2.1.7):

$$\begin{aligned} p_{F,U}(y) - p_{G,U}(y) &= E_U(\langle \nabla \mathcal{Q}_d(F-y), H_U(F,1)\rangle - \langle \nabla \mathcal{Q}_d(G-y), H_U(G,1)\rangle) \\ &= E_U\langle \nabla \mathcal{Q}_d(F-y), H_U(G,1) - H_U(F,1)\rangle \\ &+ E_U\langle \nabla \mathcal{Q}_d(G-y) - \nabla \mathcal{Q}_d(F-y), H_U(G,1)\rangle \\ &= I + J \end{aligned}$$

We recall the following inequality proved in [5]. For $p > n$, with $p' = p/(p-1)$,

$$(\mathbb{E}_U|\nabla \mathcal{Q}_d(F-y)|^{p'})^{1/p'} \le C_{p,n}(\mathbb{E}_U|H_U(F,1)|^p)^{p\frac{n-1}{p-n}}.$$

In particular, for $p = 2n$ (fixed from now on), applying (2.2.2) with $k = 0, p_1 = p_2 = p_3 = p_4 = 7p = 14n$,

$$\begin{aligned}
&(\mathbb{E}_U|\nabla \mathcal{Q}_d(F-y)|^{2n/(2n-1)})^{(2n-1)/(2n)} \\
&\quad \le C(\mathbb{E}_U|H_U(F,1)|^{2n})^{2(n-1)} \\
&\quad \le C\left[m_U(0,14n)\Gamma_{F,U}(14n)^2 n_{F,U}(0,14n)^3\right]^{4n(n-1)}.
\end{aligned} \tag{2.2.6}$$

We use now Lemma 2.6 to estimate I and J:

$$\begin{aligned}
I =&\mathbb{E}_U|\langle \nabla \mathcal{Q}_d(F-y), H_U(G,1) - H_U(F,1)\rangle| \\
&\le \|\nabla \mathcal{Q}_d(F-y)\|_{\frac{2n}{2n-1},U}\|H_U(G,1) - H_U(F,1)\|_{2n,U}
\end{aligned}$$

and we have just provided the estimate for the first factor. For the second we apply (2.2.3):

$$\begin{aligned}
&\|H_U(F,1) - H_U(G,1)\|_{2n,U} \\
&\quad \le Cm_U(0,18n)\Gamma_{F,G,U}(18n)^3 n_{F,G,U}(0,18n)^4\|\Delta_2(F,G)\|_{18n,U},
\end{aligned}$$

We now study J. For $\lambda \in [0,1]$ we denote $F_\lambda = G + \lambda(F-G)$. With a Taylor expansion, applying Holder inequality, integrating again by parts and denoting $V_{j,k} = H_{j,U}(G,1)(F-G)_k$.

$$\begin{aligned}
&\mathbb{E}_U\langle \nabla \mathcal{Q}_d(F-y) - \nabla \mathcal{Q}_d(G-y), H_U(G,1)\rangle \\
&\quad = \sum_{k,j=1}^d \int_0^1 \mathbb{E}_U(\partial_k\partial_j \mathcal{Q}_d(F_\lambda - y)H_{j,U}(G,1)(F-G)_k)d\lambda \\
&\quad = \sum_{k,j=1}^d \int_0^1 \mathbb{E}_U(\partial_j \mathcal{Q}_d(F_\lambda - y)H_{k,U}(F_\lambda, H_{j,U}(G,1)(F-G)_k))d\lambda \\
&\quad = \sum_{k,j=1}^d \int_0^1 \mathbb{E}_U(\partial_j \mathcal{Q}_d(F_\lambda - y)H_{k,U}(F_\lambda, V_{j,k}))d\lambda
\end{aligned}$$

Now, applying twice (2.2.2), first with $k = 0$ and then with $k = 1$, with some computations in the same fashion as before, it is possible to show

$$\begin{aligned}
&\|(H_{k,U}(F_\lambda, V_{j,k}))_{j=1,\ldots,n}\|_{2n,U} \\
&\quad \le Cm_U(1,32n)^2\Gamma_{F,G,U}(32n)^5 n_{F,G,U}(1,32n)^8\|F-G\|_{1,32n,U}.
\end{aligned}$$

From (2.2.6) and Holder as before,

$$|J| \le C_n\left[m_U(1,32n)\Gamma_{F,G,U}(32n)^2 n_{F,G,U}(1,32n)^3\right]^{4n^2}\|F-G\|_{1,32n,U}.$$

The statement follows. □

Lemma 2.8. *There exists a universal constant C depending just on the dimension n such that*

$$\begin{aligned}&|p_{F,U}(y)-p_{G,U}(y)|\\&\le C\left[m_U(1,32n)(\Gamma_{F,U}\vee\Gamma_{G,U})(32n)n_{F,G,U}(1,32n)\right]^{24n^2}\|\Delta_2(F,G)\|_{32n,U}\end{aligned}$$

Proof. We denote in this proof $M=\hat{\gamma}_G(\gamma_{F_\lambda}-\gamma_G)$, and define, as in (2.1.3),

$$V=\prod_{1\le i,j\le n}\psi_{1/(8n^2)}(M_{i,j}).$$

We have from Lemma 2.7

$$\begin{aligned}&|p_{F,UV}(y)-p_{G,UV}(y)|\\&\le C\left[m_{UV}(1,32n)\Gamma_{F,G,UV}(32n)n_{F,G,UV}(1,32n)\right]^{12n^2}\|\Delta_2(F,G)\|_{32n,UV}\end{aligned}\tag{2.2.7}$$

Remark

$$\hat{\gamma}_G-\hat{\gamma}_{F_\lambda}=\hat{\gamma}_G(\gamma_{F_\lambda}-\gamma_G)\hat{\gamma}_{F_\lambda},$$

so

$$\|\hat{\gamma}_{F_\lambda}-\hat{\gamma}_G\|_F\le\|\hat{\gamma}_G(\gamma_{F_\lambda}-\gamma_G)\|_F\|\hat{\gamma}_{F_\lambda}\|_F$$

On $V\ne 0$ we have $\|\hat{\gamma}_G(\gamma_{F_\lambda}-\gamma_G)\|_F\le 1/2$ so

$$\|\hat{\gamma}_{F_\lambda}\|_F\le 2\|\hat{\gamma}_G\|_F$$

and therefore

$$\Gamma_{F,G,UV}(32n)\le 2\Gamma_{G,UV}(32n)\le 2\Gamma_{G,U}(32n).\tag{2.2.8}$$

Now,

$$p_{F,U(1-V)}(y)=\mathbb{E}_U(\langle\nabla\mathcal{Q}_n(F-y),H_U(F,1-V)\rangle)$$

which implies, using as before (2.2.2) and (2.2.6)

$$\begin{aligned}p_{F,U(1-V)}(y)&=E_{U(1-V)}\langle\nabla\mathcal{Q}_d(F-y),H_U(F,1-V)\rangle\\&\le C\left[m_U(0,14n)\Gamma_{F,U}(14n)^2n_{F,U}(0,14n)^3\right]^{4n(n-1)}\|H_U(F,1-V)\|_{2n,U}\\&\le C\left[m_U(0,24n)\Gamma_{F,U}(24n)^2n_{F,U}(0,24n)^3\right]^{8n(n-1)+1}\|1-V\|_{4n,U}\end{aligned}$$

and, using (2.1.5) and some standard computations

$$\begin{aligned}\|1-V\|_{1,4n,U}&\le C\|\hat{\gamma}_G(\gamma_{F_\lambda}-\gamma_G)\|_{1,4n,U}\\&\le C\Gamma_{G,U}^2(20n)n_{F,G,U}^2(0,20n)\|F-G\|_{2,20n,U}\end{aligned}$$

so

$$\begin{aligned}&p_{F,U(1-V)}(y)\\&\le C\left[m_U(0,24n)(\Gamma_{F,U}\vee\Gamma_{G,U})(24n)^2n_{F,G,U}(0,24n)^3\right]^{8n^2}\|\Delta_2(F,G)\|_{32n,U}\end{aligned}\tag{2.2.9}$$

We conclude writing

$$|p_{F,U}(y)-p_{G,U}(y)|\le|p_{F,UV}(y)-p_{G,UV}(y)|+p_{F,U(1-V)}(y)+p_{G,U(1-V)}(y)$$

and the statement follows easily. □

We can now prove the theorem. Let V as in the last proof. We can write

$$\begin{aligned} p_{F,U}(y) \geq p_{F,UV}(y) &\geq p_{G,UV}(y) - |p_{F,UV}(y) - p_{G,UV}(y)| \\ &= p_{G,U}(y) - p_{G,U(1-V)}(y) - |p_{F,UV}(y) - p_{G,UV}(y)|. \end{aligned}$$

From (2.2.2) and (2.2.6) as before

$$p_{G,U(1-V)}(y) \leq C\left[m_U(0,14n)\Gamma_{G,U}(14n)^2 n_{F,G,U}(0,14n)^3\right]^{8n^2} \|\Delta_2(F,G)\|_{32n,U}.$$

Using also (2.2.7) and (2.2.8) we obtain the desired lower bound for p_F.

For the upper bound we apply Proposition 2.1 with $G = 1 - U$. We have

$$p_{F,1-U}(x) = \mathbb{E}\left(\sum_{i=1}^{n} \partial_i \mathcal{Q}_n(F-x) H_i(F;1-U)\right), \tag{2.2.10}$$

We use (2.2.6) with $U = 1$:

$$(\mathbb{E}|\nabla \mathcal{Q}_d(F-y)|^{2n/(2n-1)})^{(2n-1)/(2n)} \leq C\left(\Gamma_F(14n)^2 n_F(14n)^3\right)^{4n(n-1)}.$$

Now we apply Holder to (2.2.10), and using also (2.2.2) as before and (2.2.6) we find

$$p_{F,1-U} \leq C\|1-U\|_{1,14n}\left[\Gamma_F(14n)^2 n_F(14n)^3\right]^{4n^2}. \tag{2.2.11}$$

We apply now the lower bound result to $p_{G,U}$, interchanging the roles of F and G, and find

$$p_{F,U}(y) \leq p_{G,U}(y) + [m_U(1,32n)\Gamma_{F,U}(32n) n_{F,G}(1,32n)]^{24n^2} \|\Delta_2(F,G)\|_{32n,U}.$$

Putting together this inequality and (2.2.11), we have the upper bound. □

2.2.2 Estimates of the derivatives of the density

We derive now analogous local estimates for the derivatives of the density. For $\alpha \geq 1$ we set $q_{n,\alpha} = 8n(\alpha+1)(\alpha+3)$.

Theorem 2.9. *Suppose $m_U(\alpha+1, q_{n,\alpha}) < \infty$ and $n_{F,G,U}(\alpha+1, q_{n,\alpha}) < \infty$. If $\Gamma_{G,U}(q_{n,\alpha}) < \infty$, there exists a constant C_1 such that*

$$|\nabla^\alpha p_F(y)| \geq |\nabla^\alpha p_{F,U}(y)| \geq |\nabla^\alpha p_{G,U}(y)| - C_1\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha},U}$$

If $\Gamma_F(q_{n,\alpha}) < \infty$, there exists a constant C_2 such that

$$|\nabla^\alpha p_F(y)| \leq |\nabla^\alpha p_{G,U}(y)| + C_2(\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha},U} + \|1-U\|_{\alpha+1,q_{n,\alpha}})$$

If $\Gamma_{G,U}(q_{n,\alpha})$ and $\Gamma_{F,U}(q_{n,\alpha})$ are both finite,

$$|\nabla^\alpha p_{F,U}(y) - \nabla^\alpha p_{G,U}(y)| \leq C_3\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha}}$$

Remark 2.10. We can take

$$C_1 = C\,[m_U(\alpha+1,q_{n,\alpha})\Gamma_{G,U}(q_{n,\alpha})n_{F,G}(\alpha+1,q_{n,\alpha})]^{q_{n,\alpha}^2}$$
$$C_2 = C\,[m_U(\alpha+1,q_{n,\alpha})\Gamma_{F}(q_{n,\alpha})n_{F,G,U}(\alpha+1,q_{n,\alpha})]^{q_{n,\alpha}^2}$$
$$C_3 = C[m_U(\alpha+1,q_{n,\alpha})\Gamma_{F,U}\vee\Gamma_{G,U}(q_{n,\alpha})n_{F,G,U}(\alpha+1,q_{n,\alpha})]^{q_{n,\alpha}^2}$$

where C is a universal constant depending only on n and α.

Proof. We do not go through all the details here since the computations are analogous to the proof of theorem 2.4. We start proving

$$|\nabla^\alpha p_{F,U}(y)-\nabla^\alpha p_{G,U}(y)| \\ \leq C\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha}}[m_U(\alpha+1,q_{n,\alpha})\Gamma_{F,G,U}(q_{n,\alpha})n_{F,G,U}(\alpha+1,q_{n,\alpha})]^{q_{n,\alpha}^2}$$

Defining by induction $H_U^1(F,G)=H_U(F,G)$ and $H_U^{\alpha+1}(F,G)=H_U(F,H_U^\alpha(F,G))$, integrating by parts we have

$$\nabla^\alpha p_{F,U}(y)=\mathbb{E}_U[\nabla^{\alpha+1}\mathcal{Q}_n(F-y)H_U(F,1)]=\mathbb{E}_U[\nabla\mathcal{Q}_n(F-y)H_U^{\alpha+1}(F,1)].$$

So

$$\begin{aligned}&\nabla^\alpha p_{F,U}(y)-\nabla^\alpha p_{G,U}(y)\\ &=\mathbb{E}_U[\nabla\mathcal{Q}_n(F-y)H_U^{\alpha+1}(F,1)]-\mathbb{E}_U[\nabla\mathcal{Q}_n(G-y)H_U^{\alpha+1}(G,1)]\\ &=\mathbb{E}_U[\nabla\mathcal{Q}_n(F-y)(H_U^{\alpha+1}(F,1)-H_U^{\alpha+1}(G,1))]\\ &+\mathbb{E}_U[(\nabla\mathcal{Q}_n(F-y)-\nabla\mathcal{Q}_n(G-y))H_U^{\alpha+1}(G,1)]\\ &=I+J.\end{aligned}$$

Now,

$$\begin{aligned}J&=\mathbb{E}_U[(\nabla\mathcal{Q}_n(F-y)-\nabla\mathcal{Q}_n(G-y))H_U^{\alpha+1}(G,1)]\\ &=\mathbb{E}_U[\int_0^1(\nabla^2\mathcal{Q}_n(F_\lambda-y))H_U^{\alpha+1}(G,1)(F-G)d\lambda]\\ &=\mathbb{E}_U[\int_0^1(\nabla\mathcal{Q}_n(F_\lambda-y))H_U(F_\lambda,H_U^{\alpha+1}(G,1))(F-G)d\lambda]\end{aligned}$$

Apply Holder inequality and for $1/q_1+1/q_2+1/q_3=1$ and

$$J\leq\|\nabla\mathcal{Q}_n(F_\lambda-y)\|_{q_1,U}\|H_U(F_\lambda,H_U^{\alpha+1}(G,1))\|_{q_2,U}\|F-G\|_{q_3,U}.$$

Iterating (2.2.2),

$$\|H_U(F_\lambda,H_U^{\alpha+1}(G,1))\|_{p,U} \\ \leq C(m_U(\alpha+1,p_1))^{\alpha+1}\Gamma_{F,U}(p_2)^{(\alpha+1)(a+3)}n_{F,U}(\alpha+1,p_3)^{(2\alpha+3)^2}$$

with $\frac{\alpha+1}{p_1}+\frac{(\alpha+1)(\alpha+4)/2}{p_2}+\frac{(\alpha+1)(a+3)}{p_3}=\frac{1}{p}$. Now,

$$|I|=\|\nabla\mathcal{Q}_n(F-y)\|_{p',U}\|H_U^{\alpha+1}(F,1)-H_U^{\alpha+1}(G,1)\|_{p,U}.$$

For $\|\nabla\mathcal{Q}_n(F-y)\|_{p,U}$ we use (2.2.6). For the second factor,

$$\begin{aligned}\|H_U^{\alpha+1}(F,1)-H_U^{\alpha+1}(G,1)\|_{k,p,U}&\le\|H_U(F,H_U^{\alpha}(F,1))-H_U(G,H_U^{\alpha}(F,1))\|_{k,p,U}\\&\quad+\|H_U(G,H_U^{\alpha}(F,1))-H_U(G,H_U^{\alpha}(G,1))\|_{k,p,U}.\end{aligned}$$

(2.2.3) implies

$$\begin{aligned}&\|H_U(F,H_U^{\alpha}(F,1))-H_U(G,H_U^{\alpha}(F,1))\|_{k,p,U}\\&\le C\|H_U^{\alpha}(F,1)\|_{k+1,p_1,U}m_U(k,p_2)\Gamma_{F,G,U}^{k+3}(p_3)n_{F,G,U}(k,p_4)^{2k+4}\|\Delta_{k+2}(F,G)\|_{p_5,U}\end{aligned}$$

with $\frac{1}{p}=\frac{1}{p_1}+\frac{1}{p_2}+\frac{k+3}{p_3}+\frac{2k+4}{p_4}+\frac{1}{p_5}$. Linearity of H_U in the second variable and (2.2.2) lead to

$$\begin{aligned}&\|H_U(G,H_U^{\alpha}(F,1))-H_U(G,H_U^{\alpha}(G,1))\|_{k,p,U}\\&\le\|H_U\left(G,H_U^{\alpha}(F,1)-H_U^{\alpha}(G,1)\right)\|_{k,p,U}\\&\le C\|H_U^{\alpha}(F,1)-H_U^{\alpha}(G,1)\|_{k+1,p_1,U}m_U(k,p_2)\Gamma_{F,G,U}^{k+2}(p_3)n_{F,G,U}(k,p_4,U)^{2k+3}\end{aligned}$$

By induction on α

$$\begin{aligned}&\|H_U^{\alpha+1}(F,1)-H_U^{\alpha+1}(G,1)\|_{p,U}\\&\le m_U(\alpha,p_1)^{\alpha+1}\Gamma_{F,G,U}^{(\alpha+1)(\alpha+4)/2}(p_2)n_{F,G,U}(\alpha,p_3)^{(\alpha+1)(a+3)}\|\Delta_{\alpha+2}(F,G)\|_{p_4,U}\end{aligned}$$

with $\frac{1}{p}=\frac{\alpha+1}{p_1}+\frac{(\alpha+1)(\alpha+4)/2}{p_2}+\frac{(\alpha+1)(a+3)}{p_3}+\frac{1}{p_4}$. Therefore, taking $p=2n$,

$$\begin{aligned}&|\nabla^{\alpha}p_{F,U}(y)-\nabla^{\alpha}p_{G,U}(y)|\\&\quad\le\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha}}[m_U(\alpha+1,q_{n,\alpha})\Gamma_{F,G,U}(q_{n,\alpha})n_{F,G,U}(k,q_{n,\alpha})]^{q_{n,\alpha}^2}\end{aligned}$$

It is possible to prove the analogous of Lemma 2.7 and 2.8. One first shows

$$\begin{aligned}&|\nabla^{\alpha}p_{F,U}(y)-\nabla^{\alpha}p_{G,U}(y)|\\&\le C\|\Delta_{\alpha+2}(F,G)\|_{q_{n,\alpha}}[m_U(\alpha+1,q_{n,\alpha})(\Gamma_{F,U}\vee\Gamma_{G,U})(q_{n,\alpha})n_{F,G,U}(\alpha,q_{n,\alpha})]^{q_{n,\alpha}^2},\end{aligned}$$

following the proof of Lemma 2.7 and iterating (2.2.2). As a consequence we find a lower bound for $|\nabla^{\alpha}p_F|$ which does not involve the Malliavin covariance of F, and an upper bound for $|\nabla^{\alpha}p_F|$ which does not involve the Malliavin covariance of G. We follow the proof of Lemma 2.8 and apply (2.1.5). □

2.2.3 Density estimates via local inversion

In this section we recall some results from [6], Appendix 2. We see how to use the inverse function theorem to transfer a known estimate for a Gaussian random variable to its image via a certain function η. This provide us some estimates for the localized density, from which we obtain lower bounds for the global density. To deal with the global upper bound problem we will need some Malliavin calculus techniques. For a standard version of the inverse function theorem see [88].

We consider $\Phi(\theta) = \theta + \eta(\theta)$, for a three times differentiable function $\eta : \mathbb{R}^d \to \mathbb{R}^d$. Define

$$c_2(\eta) = \max_{i,j=1,..,d} \sup_{|x|\le h_\eta} |\partial^2_{ij}\eta(x)|, \quad c_3(\eta) = \max_{i,j,k=1,..,d} \sup_{|x|\le h_\eta} |\partial^3_{ijk}\eta(x)|,$$

and

$$h_\eta = \frac{1}{16d^2(c_2(\eta) + \sqrt{c_3(\eta)})} \tag{2.2.12}$$

Lemma 2.11. *Take h_η as above. If the function η is such that*

$$\eta \in C^3(\mathbb{R}^d, \mathbb{R}^d), \quad \eta(0) = 0, \quad \nabla\eta(0) \le \frac{1}{2},$$

then there exists a neighborhood of 0, that we denote with $V_{h_\eta} \subset B(0, 2h_\eta)$, such that $\Phi : V_{h_\eta} \to B\left(0, \frac{1}{2}h_\eta\right)$ is a diffeomorphism. In particular, if we denote with Φ^{-1} the local inverse of Φ, we have

$$\Phi^{-1} : B\left(0, \frac{1}{2}h_\eta\right) \to B\left(0, 2h_\eta\right),$$

and we have this quantitative estimate:

$$\forall y \in B\left(0, \frac{1}{2}h_\eta\right), \quad \frac{1}{4}|\Phi^{-1}(y)| \le |y| \le 4|\Phi^{-1}(y)|. \tag{2.2.13}$$

Remark 2.12. Here we write Φ^{-1} for the inverse of the restriction of Φ to V_{h_η}, what is called a *local* inverse.

Proof. We have

$$\nabla\Phi(0) = Id + \nabla\eta(0).$$

So

$$|\nabla\Phi(0)x|^2 \ge \frac{1}{2}|x|^2 - |\nabla\eta(0)x|^2 \ge \frac{1}{2}|x|^2 - \frac{1}{4}|x|^2 = \frac{1}{4}|x|^2.$$

and

$$|\nabla\Phi(0)x|^2 \le 2|x|^2 + 2|\nabla\eta(0)x|^2 \le 2|x|^2 + \frac{1}{2}|x|^2 \le \frac{5}{2}|x|^2.$$

Therefore

$$\frac{1}{2}|x| \le |\nabla\Phi(0)x| \le \sqrt{3}|x|$$

This implies $\Phi(0)$ is invertible locally around 0, and the local inverse differentiable, using the classical inverse function theorem. We now look now at the image of the inverse, and at the estimates (2.2.13). We develop η around 0, writing $\nabla^2\eta(x)[u,v]$ to denote $\nabla^2\eta(x)$ computed in u and v.

$$\eta(\theta) = \nabla\eta(0)\theta + \int_0^1 (1-t)\nabla^2\eta(t\theta)[\theta,\theta]dt.$$

Fix $y \in \mathbb{R}^d$. Suppose $\Phi(\theta) = y$. Then

$$\begin{aligned}\theta &= (\nabla\Phi(0))^{-1}\nabla\Phi(0)\theta \\ &= (\nabla\Phi(0))^{-1}(\theta + \nabla\eta(0)\theta) \\ &= (\nabla\Phi(0))^{-1}\left(\theta + \eta(\theta) - \int_0^1 (1-t)\nabla^2\eta(t\theta)[\theta,\theta]dt\right) \\ &= (\nabla\Phi(0))^{-1}\left(y - \int_0^1 (1-t)\nabla^2\eta(t\theta)[\theta,\theta]dt\right).\end{aligned}$$

We define

$$U_y(\theta) = \left(y - \int_0^1 (1-t)\nabla^2\eta(t\theta)[\theta,\theta]dt\right),$$

so that θ can be seen as a fixed point for U_y. Recall that $|\frac{1}{2}x| \le |\nabla\Phi(0)x|$.

$$\begin{aligned}|U_y(\theta_1) - U_y(\theta_2)| &= \left|(\nabla\Phi(0))^{-1}\left(\int_0^1 (1-t)(\nabla^2\eta(t\theta_2)[\theta_2,\theta_2] - \nabla^2\eta(t\theta_1)[\theta_1,\theta_1])dt\right)\right| \\ &\le 2\left|\int_0^1 (1-t)(\nabla^2\eta(t\theta_2)[\theta_2,\theta_2] - \nabla^2\eta(t\theta_1)[\theta_1,\theta_1])dt\right| \\ &\le 2\int_0^1 (1-t)(|\nabla^2\eta(t\theta_1)[\theta_1,\theta_1-\theta_2]| + |\nabla^2\eta(t\theta_1)[\theta_1-\theta_2,\theta_2]| \\ &+ |\nabla^2\eta(t\theta_1)[\theta_2,\theta_2] - \nabla^2\eta(t\theta_2)[\theta_2,\theta_2]|)dt.\end{aligned}$$

Now, form (2.2.12), for $\theta_1,\theta_2 \in B(0,h_\eta)$

$$|\nabla^2\eta(t\theta_1)[\theta_1,\theta_1-\theta_2]| \le d^2c_2(\eta)h_\eta|\theta_1-\theta_2| \le \frac{1}{16}|\theta_1-\theta_2|,$$

and

$$|\nabla^2\eta(t\theta_1)[\theta_2,\theta_2] - \nabla^2\eta(t\theta_2)[\theta_2,\theta_2]| \le d^3c_3(\eta)|\theta_1-\theta_2|h_\eta^2 \le \frac{1}{256}|\theta_1-\theta_2|,$$

and therefore

$$|U_y(\theta_1) - U_y(\theta_2)| \le \frac{1}{4}|\theta_1-\theta_2|. \tag{2.2.14}$$

For $y \in B(0,\frac{1}{2}h_\eta)$ and $\theta \in B(0,2h_\eta)$ this implies

$$|U_y(\theta)| \le |U_y(\theta) - U_y(0)| + |U_y(0)| \le \frac{1}{4}|\theta| + 2y \le 2h_\eta$$

Define now the sequence

$$\theta_0 = 0, \quad \theta_{k+1} = U_y(\theta_k).$$

We know that $\theta_k \in B(0, 2h_\eta)$ for any $k \in \mathbb{N}$, and therefore inequality (2.2.14) implies

$$|U_y(\theta_k) - U_y(\theta_{k+1})| \leq \frac{1}{4}|\theta_k - \theta_{k+1}|.$$

The Banach fixed-point theorem tells us that θ_k converges to the unique solution of $\theta = U_y(\theta)$, which is $\theta = \Phi^{-1}(y)$, and $\theta \in B(0, 2h_\eta)$. So it is possible to define the local inverse Φ^{-1} on $B\left(0, \frac{1}{2}h_\eta\right)$, and

$$V_{h_\eta} := \Phi^{-1} B\left(0, \frac{1}{2}h_\eta\right) \subset B(0, 2h_\eta).$$

Now, for $y \in B(0, \frac{1}{2}h_\eta)$, let $\theta = \Phi^{-1}(y)$ and the following inequalities hold

$$|\theta| = |U_y(\theta)| \leq \frac{1}{2}\theta + 2|y| \qquad \Rightarrow |\theta| \leq 4|y|$$

$$|\theta| = U_y(\theta) \geq |U_y(0)| - |U_y(\theta) - U_y(0)| \geq \frac{1}{2}|y| - \frac{1}{2}|\theta| \qquad \Rightarrow |\theta| \geq \frac{1}{4}|y|.$$

□

Let now Θ be a d-dimensional centered Gaussian variable with covariance matrix Q. Denote by $\underline{\lambda}$ and $\overline{\lambda}$ the lower and the upper eigenvalues of Q. Keeping in mind the setting of the last subsection, we also introduce the notation

$$c_*(\eta, h) = \sup_{|x| \leq 2h} \max_{i,j} |\partial_i \eta^j(x)|$$

for $h > 0$. Recall we are supposing $\eta \in C^3(\mathbb{R}^d, \mathbb{R}^d)$ and $\eta(0) = 0$.

Suppose now Q non-degenerate and take $r > 0$ such that

$$c_*(\eta, 16r) \leq \frac{1}{2d}\sqrt{\frac{\underline{\lambda}}{\overline{\lambda}}}, \qquad r \leq h_\eta = \frac{1}{16d^2(c_2(\eta) + \sqrt{c_3(\eta)})}. \tag{2.2.15}$$

We take a localizing function as in (2.1.3), with $a_i = r \ \forall i = 1, \ldots d$, and $U = \prod_{i=1}^d \psi_r(\Theta_i)$.

Lemma 2.13. *The density* $p_{G,U}$ *of*

$$G := \Phi(\Theta) = \Theta + \eta(\Theta)$$

under $\mathbb{P}_U$ *has the following bounds on* $B(0, r)$*:*

$$\frac{1}{C \det Q^{1/2}} \exp\left(-\frac{C}{\underline{\lambda}}|z|^2\right) \leq p_{G,U}(z) \leq \frac{C}{\det Q^{1/2}} \exp\left(-\frac{1}{C\overline{\lambda}}|z|^2\right) \tag{2.2.16}$$

Proof. For a general nonnegative, measurable function $f : \mathbb{R}^d \to \mathbb{R}$ with support included in $B(0,4r)$, we compute $E(f(G)1_{\{\Theta\in\Phi^{-1}B(0,4r)\}})$. Here Φ^{-1} is the local diffeomorphism of the inverse function theorem. After the multiplication with the characteristic function, on the support of the random variable that we are averaging, Φ is a diffeomorphism and the first equality holds. The second follows from the change of variable suggested by Lemma 2.11 for $G = \Phi(\Theta)$

$$\begin{aligned} E\left(f(G)1_{\{\Theta\in\Phi^{-1}B(0,4r)\}}\right) &= \\ &= \int_{\Phi^{-1}(B(0,4r))} f(\Phi(\theta))\frac{1}{(2\pi)^{d/2}\det Q^{1/2}}\exp\left(-\frac{1}{2}\langle Q^{-1}\theta,\theta\rangle\right)d\theta \\ &= \int_{B(0,4r)} f(z)\bar{p}_G(z)dz, \end{aligned}$$

where for $z \in B(0,4r)$

$$\bar{p}_G(z) = \frac{1}{(2\pi)^{d/2}\det Q^{1/2}|\det\nabla\Phi(\Phi^{-1}(z))|}\exp\left(-\frac{1}{2}\langle Q^{-1}\Phi^{-1}(z),\Phi^{-1}(z)\rangle\right).$$

Again from Lemma 2.11, since $4r \le \frac{h_\eta}{2}$, we have $z \in B(0,4r) \Rightarrow \theta \in B(0,16r)$. Using $c_*(\eta,16r) \le \frac{1}{2d}\sqrt{\frac{\underline{\lambda}}{\overline{\lambda}}}$,

$$\frac{1}{2}|x|^2 \le (1-d\,c_*(\eta,h_\eta))|x|^2 \le |\langle\nabla\Phi(\theta)x,x\rangle| \le (1+d\,c_*(\eta,h_\eta))|x|^2 \le 2|x|^2.$$

Therefore if $z \in B(0,4r)$

$$2^{-d} \le |\det\Phi(\Phi^{-1}(z))| \le 2^d.$$

Moreover, using Lemma 2.11

$$\begin{aligned} \langle Q^{-1}\Phi^{-1}(z),\Phi^{-1}(z)\rangle &\le \frac{1}{\underline{\lambda}}|\Phi^{-1}(z)|^2 \le \frac{16}{\underline{\lambda}}|z|^2, \\ \langle Q^{-1}\Phi^{-1}(z),\Phi^{-1}(z)\rangle &\ge \frac{1}{\overline{\lambda}}|\Phi^{-1}(z)|^2 \ge \frac{1}{16\overline{\lambda}}|z|^2. \end{aligned}$$

Therefore

$$\frac{1}{(8\pi)^{d/2}\det Q^{1/2}}\exp\left(-\frac{8}{\underline{\lambda}}|z|^2\right) \le \bar{p}_G(z) \le \frac{2^{d/2}}{\pi^{d/2}\det Q^{1/2}}\exp\left(-\frac{1}{32\overline{\lambda}}|z|^2\right).$$

Now we define, as in (2.1.3) the localization variables

$$U_1 = \prod_{i=1}^d \psi_{16r}(\Theta_i), \quad U_2 = \prod_{i=1}^d \psi_r(\Theta_i).$$

Notice that

$$U_2 \le 1_{\{\Theta\in\Phi^{-1}B(0,4r)\}} \le U_1,$$

so that we have

$$E\left(f(G)U_2\right) \leq E\left(f(G)1_{\{\Theta\in\Phi^{-1}B(0,4r)\}}\right) \leq E\left(f(G)U_1\right).$$

The following bounds for the local densities follow:

$$\begin{aligned} p_{G,U_1}(z) &\geq \frac{1}{(8\pi)^{d/2}\det Q^{1/2}}\exp\left(-\frac{8}{\underline{\lambda}}|z|^2\right), \\ p_{G,U_2}(z) &\leq \frac{2^{d/2}}{\pi^{d/2}\det Q^{1/2}}\exp\left(-\frac{1}{32\overline{\lambda}}|z|^2\right). \end{aligned}$$

$U_1 \geq U = U_2$, so for the localization via U both bounds hold. □

Chapter 3

Tube and density estimates for diffusion processes under a weak Hörmander condition

3.1 Introduction

Following [81], we consider a stochastic differential equation on $[0, T]$:

$$X_t = x + \int_0^t \sigma(X_s) \circ dW_s + \int_0^t b(X_s) ds \tag{3.1.1}$$

where the diffusion X is two-dimensional and the Brownian Motion W is one-dimensional. $\circ dW_s$ denotes the Stratonovich integral and, as we said, we suppose $\partial_\sigma \sigma(y) = \kappa_\sigma(y)\sigma(y)$. For this system the ellipticity assumption fails at any point, and the strong Hörmander condition fails as well. The prototype of this kind of problems is a two dimensional system where the first component X^1 follows a stochastic dynamic, and the second component X^2 is a deterministic functional of X^1, so the randomness acts indirectly on X^2. Besides the natural application to the Asian option, there are others such as in [61], [62]. In these papers the functioning of a neuron is modeled: X^2 is the concentration of some chemicals resulting from a reaction involving the first component X^1. Differently from our setting, though, there are several measurements corresponding to the input X^1, so X^2 is multi-dimensional. The pattern, however, is similar.

Under a non-degeneracy assumption of weak Hörmander type we find diagonal Gaussian estimates for the density in short time (for density estimates in a weak Hörmander framework we refer to [11], [42], [63], [85], [80], [24], [13])). In this paper we consider a more general coefficient for the Brownian Motion, in the sense that we suppose $\partial_\sigma \sigma(y) = \kappa_\sigma(y)\sigma(y)$, whereas the works mentioned above would apply for $\sigma = (\sigma_1, 0)$ which is a more restrictive condition. Moreover, our coefficients are just locally hypoelliptic.

The other novelty is that thanks to our short time non-asymptotic result we find exponential lower and upper bounds for probability that the diffusion remains in a small tube around a deterministic trajectory (theorem 3.10). More precisely, we consider (3.1.1) and introduce the associated skeleton path solution of the following ODE:

$$x_t(\phi) = x + \int_0^t \sigma(x_s(\phi))\phi_s ds + \int_0^t b(x_s(\phi))ds. \tag{3.1.2}$$

for a certain control $\phi \in L^2[0,T]$. A tube estimate for (3.1.1), is an estimate of $\mathbb{P}\left(\sup_{t\leq T} \|X_t - x_t(\phi)\| \leq R\right)$. We obtain this result in a norm which reflects the fact that the diffusion moves with speed $t^{1/2}$ in the direction σ and $t^{3/2}$ in the direction $[\sigma, b]$, and establish a connection between this norm and the standard control distance (cf. [78]).

We also consider, in section 3.5, the system of stochastic differential equations studied in [42]. We take a Brownian Motion in $W \in \mathbb{R}^d$, and a chain of n differential equations in dimension d:

$$\begin{aligned}
dX_t^1 &= B_1(X_t^1, \ldots, X_t^n)dt + \sigma(X_t^1, \ldots, X_t^n) \circ dW_t \\
dX_t^2 &= B_2(X_t^1, \ldots, X_t^n)dt \\
dX_t^3 &= B_3(X_t^2, \ldots, X_t^n)dt \\
&\ldots \\
dX_t^n &= B_n(X_t^{n-1}, X_t^n)dt
\end{aligned}$$

each X_t^i being $\mathbb{R}^d$ valued as well. An example of physical system satisfying (3.5.1) is a chain of n coupled oscillators, each of them connected to the nearest neighbors, with the first oscillator forced by a random noise. For $n = 2$, $d = 1$, this equation corresponds again to the dynamics used in mathematical finance to price Asian options. We apply the Malliavin calculus techniques introduced in chapter 2, finding a Gaussian density estimate in short time. This is coherent with the result of [42], proved by the authors using both parametrix and optimal stochastic control techniques. Our estimate holds just in short time, in a neighborhood of the initial condition of the diffusion, and we consider coefficients that do not depend on time, whereas in [42] the estimate is global, holds for any $t \in (0, T]$, and the coefficients may depend on t. On the other hand, we require only local non-degeneracy, in contrast to what is done in [42], and our result is more precise for small enough times.

3.2 Notation and results

3.2.1 Notations

We introduce some notations. For any function $\eta : D \subset \mathbb{R}^d \to \mathbb{R}^m$ we denote with $\nabla\eta(x)$ the differential of η in x, which is a linear function from $\mathbb{R}^d$ in $\mathbb{R}^m$, given by the Jacobian matrix. For $f, g : \mathbb{R}^n \to \mathbb{R}^n$ we recall the definition of the directional derivative of f in the direction g as

$$\partial_g f(x) = (\nabla f)\, g(x) = \sum_{i=1}^n g^i(x)\partial_{x_i} f(x).$$

The Lie bracket $[f, g]$ in x is defined as

$$[f, g](x) = \partial_f g(x) - \partial_g f(x).$$

We denote with M^T the transpose of any matrix M. For a squared matrix M we also use the notation $\lambda_*(M)$ for the smallest absolute value of an eigenvalue of M, and $\lambda^*(M)$ for the largest one. For $x \in \mathbb{R}^2$, we denote with $A(x)$ the 2×2 matrix $(\sigma(x), [\sigma, b](x))$. For any $R > 0$, we denote with $A_R(x)$ the matrix $\big(R^{1/2}\sigma(x), R^{3/2}[\sigma, b](x)\big)$. For fixed R and x, since we suppose $A_R(x)$ invertible, we associate to $A_R(x)$ the norm on $\mathbb{R}^2$

$$|\xi|_{A_R(x)} = \sqrt{\langle (A_R A_R(x)^T)^{-1}\xi, \xi\rangle} = |A_R^{-1}(x)\xi|.$$

We suppose σ, b differentiable three times and define

$$n(x) = \sum_{k=0}^{3} \sum_{|\alpha|=k} (|\partial_x^\alpha b(x)| + |\partial_x^\alpha \sigma(x)|)$$

and $\lambda(x) = \lambda_*(A(x))$. We denote with $L(\mu, h)$ the class of non-negative functions which have the property

$$f(t) \le \mu f(s) \quad \text{for } |t - s| \le h. \tag{3.2.1}$$

3.2.2 Results

We assume that:

H1 Locally uniform weak Hörmander condition: there exists a function $\lambda_\cdot : [0, T] \to (0, 1]$ such that

$$\lambda(y) \ge \lambda_t, \quad \forall |y - x_t(\phi)| < 1, \forall t \in [0, T].$$

H2 Locally uniform bounds for derivatives: there exists a function $n_\cdot : [0, T] \to [1, \infty)$ such that

$$n(y) \le n_t, \quad \forall |y - x_t(\phi)| < 1, \forall t \in [0, T].$$

H3 Geometric condition on volatility: $\exists \kappa_\sigma : \mathbb{R}^2 \to \mathbb{R}$ s. t.

$$\partial_\sigma \sigma(x) = \kappa_\sigma(x)\sigma(x). \tag{3.2.2}$$

We suppose w.l.o.g. that $|\kappa_\sigma(x)| \le n(x)$, $|\kappa'_\sigma(x)| \le n(x)$ (this is a consequence of **H2**). If $\sigma(x) = (\sigma_1(x), 0)$, i.e. the Asian option stochastic differential equation, this property holds true with $\kappa_\sigma = \sigma'_1/\sigma_1$.

H4 Control on the growth of bounds: we suppose $|\phi_\cdot|^2$, $\lambda_\cdot$, $n_\cdot \in L(\mu, h)$, for some $h \in \mathbb{R}_{>0}$, $\mu \ge 1$.

Notice that the above hypothesis do not involve global controls of our bounds on $\mathbb{R}^2$: they concern the behavior of the coefficients only along the tube, and may vary with $t \in [0, T]$.

Under assumptions **H1**, **H2**, **H3** we prove the following Gaussian bounds for the density in short time. Define, for fixed δ, $\hat{x} = x + \delta b(x)$.

Theorem 3.3 . *There exist constants $L, L_1, L_2, K_1, K_2, \delta^*$ such that: for any $r_* > 0$, for $\delta \leq \delta^* \exp(-Lr_*^2)$, for $|y - \hat{x}|_{A_\delta(x)} \leq r_*$,*

$$\frac{K_1}{\delta^2} \exp\left(-L_1 |y - \hat{x}|^2_{A_\delta(x)}\right) \leq p_{X_\delta}(y) \leq \frac{K_2}{\delta^2} \exp\left(-L_2 |y - \hat{x}|^2_{A_\delta(x)}\right).$$

This estimate is local around the point $\hat{x}$. In this general framework it is not possible to obtain global lower bounds, since the weak Hörmander condition does not ensure the positivity of the density. See Remark 3.5 for details.

Using this estimate we prove the following result for the tube in the A_R-matrix norm:

Theorem 3.10 . *We assume that* **H1**, **H2**, **H3**, **H4** *holds, with $x_t(\phi)$ given by* (3.1.2)*. There exist K, q universal constants such that for $H_t = K\left(\frac{\mu n_t}{\lambda_t}\right)^q$, for $R \leq R_*(\phi)$ (defined in* (3.3.8)*), holds*

$$\exp\left(-\int_0^T H_t \left(\frac{1}{R} + |\phi_t|^2\right) dt\right) \leq \mathbb{P}\left(\sup_{t \leq T} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} \leq 1\right) \leq \exp\left(-\int_0^T \frac{1}{H_t}\left(\frac{1}{R} + |\phi_t|^2\right) dt\right) \quad (3.2.3)$$

Remark 3.1. Notice that μ is involved in the definition of H_t, so estimate (3.2.3) holds for the controls ϕ which belong to the class $L(\mu, h)$. In this sense, H_t depends on the "growth property" (3.2.1) of ϕ.

Both of these theorems can be stated in a control metric as well, a variant of the Caratheodory distance which looks appropriate to our framework. Here we just briefly give the definition, for more details see section 3.4.2. For $\phi \in L^2((0,1), \mathbb{R}^2)$, we define the norm

$$\|\phi\|_{(1,3)} = \left\|(\phi^1, \phi^2)\right\|_{(1,3)} = \left\|(|\phi^1|, |\phi^2|^{1/3})\right\|_{L^2(0,1)}.$$

and, given $A(x) = (\sigma(x), [\sigma, b](x))$, the set

$$C_A(x, y) = \{\phi \in L^2((0,1), \mathbb{R}^2) : dv_s = A(v_s)\phi_s ds,\ x = v_0,\ y = v_1\}.$$

We define the control norm as

$$d_c(x, y) = \inf\left\{\|\phi\|_{(1,3)} : \phi \in C_A(x, y)\right\}.$$

Just remark that this distance accounts of the different speed in the $[\sigma, b]$ direction. We define also the following quasi-distance (which is naturally associated to the norm $|\cdot|_{A_R()}$):

$$d(x,y) \le \sqrt{R} \Leftrightarrow |x-y|_{A_R(x)} \le 1.$$

In section 3.4.2 we prove that d and d_c are locally equivalent. Now we can restate theorem 3.10 as follows:

Corollary 3.2. *For $H_t = K\left(\frac{\mu n_t}{\lambda_t}\right)^q$, with K, q universal constants, for small R it holds*

$$\exp\left(-\int_0^T H_t\left(\frac{1}{R}+|\phi_t|^2\right)dt\right) \le \mathbb{P}\left(\sup_{0\le t\le T} d_c(X_t, x_t(\phi)) \le \sqrt{R}\right) \le \exp\left(-\int_0^T \frac{1}{H_t}\left(\frac{1}{R}+|\phi_t|^2\right)dt\right)$$

Example 1: Asian option (see for instance [52]). Consider a system given Black and Scholes model for the price of an asset, and an (arithmetic average) Asian option on that asset with time horizon T. This is a problem of real interest in mathematical finance. The associated SDE is

$$dX_t^1 = X_t^1(\sigma \circ dW_t + r dt);\ X_0^1 = \xi > 0, \qquad dX_t^2 = \frac{X_t^1}{T}dt;\ X_0^2 = 0.$$

(the stochastic integral is in Stratonovich form so to recover the classical formulation $r \to r + \sigma^2/2$). In this case

$$A_R^{-1}(x) = \begin{pmatrix} \sigma x^1 R^{1/2} & 0 \\ 0 & \frac{\sigma x^1}{T} R^{3/2} \end{pmatrix}^{-1} = \frac{1}{\sigma x^1}\begin{pmatrix} \frac{1}{R^{1/2}} & 0 \\ 0 & \frac{T}{R^{3/2}} \end{pmatrix}$$

Remark that this matrix is invertible for $x^1 \neq 0$. Since we are working under local non-degeneracy assumptions, our tube estimates hold for any initial condition $\xi > 0$, since this implies the positivity of the first component of the skeleton path at any time $t > 0$. On the other hand, results requiring "global" non degeneracy, such as the density estimates in [42], do not hold for this model. We take as control $\phi_t = 0$ so $x_t(\phi) = \xi\left(e^{rt}, \frac{1}{T}\int_0^t e^{rs}ds\right)$. We have

$$\begin{aligned}
|X_t - x_t(\phi)|_{A_R(x_t(\phi))} &= \frac{1}{\sigma \xi e^{rt}}\sqrt{\frac{|X_t^1 - \xi e^{rt}|^2}{R} + \frac{T^2|X_t^2 - \frac{\xi}{T}\int_0^t e^{rs}ds|^2}{R^3}} \\
&= \frac{1}{\sigma \xi e^{rt}}\sqrt{\frac{\xi^2|e^{rt}(e^{\sigma W_t}-1)|^2}{R} + \frac{\xi^2|\int_0^t e^{rs+\sigma W_s}ds - \int_0^t e^{rs}ds|^2}{R^3}} \\
&= \frac{1}{\sigma e^{rt}}\sqrt{\frac{|e^{rt}(e^{\sigma W_t}-1)|^2}{R} + \frac{|\int_0^t e^{rs}(e^{\sigma W_s}-1)ds|^2}{R^3}}
\end{aligned}$$

and applying our tube estimate we find

$$e^{-C_1 T/R} \leq \mathbb{P}\left(\sup_{t\leq T} \left\{ \frac{|e^{\sigma W_t} - 1|^2}{R\sigma^2} + \frac{|\int_0^t e^{r(s-t)}(e^{\sigma W_s} - 1)ds|^2}{R^3\sigma^2} \right\} \leq 1 \right) \leq e^{-C_2 T/R}.$$

Example 2: Geometric average Asian option ([52]). The following SDE may represent the Black and Scholes model for the log-price of an asset, and the log-price of a geometric Asian option on that asset with time horizon T:

$$dX_t^1 = \sigma \circ dW_t + rdt = \sigma dW_t + rdt;\ X_0^1 = \xi, \qquad dX_t^2 = \frac{X_t^1}{T}dt;\ X_0^2 = 0.$$

The matrix

$$A_R^{-1}(x) = \begin{pmatrix} \sigma R^{1/2} & 0 \\ 0 & \frac{\sigma}{T} R^{3/2} \end{pmatrix}^{-1} = \frac{1}{\sigma} \begin{pmatrix} \frac{1}{R^{1/2}} & 0 \\ 0 & \frac{T}{R^{3/2}} \end{pmatrix}$$

does not depend on x. With $\phi_t = 0$ we have $x_t(\phi) = \left(\xi + rt, \frac{\xi t + rt^2/2}{T}\right)$, so

$$\begin{aligned} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} &= \frac{1}{\sigma}\sqrt{\frac{|X_t^1 - \xi + rt|^2}{R} + \frac{T^2|X_t^2 - (\xi t + rt^2/2)/T|^2}{R^3}} \\ &= \frac{1}{\sigma}\sqrt{\frac{|\sigma W_t|^2}{R} + \frac{|\sigma \int_0^t W_s ds|^2}{R^3}}, \end{aligned}$$

and we find, with our tube estimate,

$$e^{-C_1 T/R} \leq \mathbb{P}\left(\sup_{t\leq T} \left\{ \frac{|W_t|^2}{R} + \frac{|\int_0^t W_s ds|^2}{R^3} \right\} \leq 1 \right) \leq e^{-C_2 T/R}.$$

3.3 Density and tube estimate of the diffusion process

We study the behavior of the diffusion X, defined in (3.1.1), on a small time interval $[0, \delta]$. We end up finding exponential lower and upper bound for the density of X_δ, for δ small enough, in the matrix norm associated to the diffusion. Recall $\hat{x} = x + \delta b(x)$. We introduce the class of constants

$$\mathcal{C} = \{C = K\,(n(x)/\lambda(x))^q\,, \exists K, q \geq 1 \text{ universal constants}\}$$

(recall $n(x), \lambda(x)$ are defined in Subsection 3.2.1). We will also denote with $1/\mathcal{C} = \{\delta : 1/\delta \in \mathcal{C}\}$. For this result we need to suppose **H1**, **H2**, **H3** locally around x, and we do not require **H4**, which is needed only in the concatenation involved in the tube estimate.

Theorem 3.3. *There exist constants $L, L_1, K_2 \in \mathcal{C}$, $L_2, K_1, \delta^* \in 1/\mathcal{C}$ such that: for any $r_* > 0$, for*

$$\delta \leq \delta^* \exp\left(-2Lr_*^2\right), \tag{3.3.1}$$

for $|y - \hat{x}|_{A_\delta(x)} \leq r_*$,

$$\frac{K_1}{\delta^2} \exp\left(-L_1|y - \hat{x}|^2_{A_\delta(x)}\right) \leq p_{X_\delta}(y) \leq \frac{K_2}{\delta^2} \exp\left(-L_2|y - \hat{x}|^2_{A_\delta(x)}\right). \tag{3.3.2}$$

Remark 3.4. Taking $r_* = L^{-1/2}$, this implies in particular the following fact: there exist constants $K_2 \in \mathcal{C}$, $K_1, \delta^*, r_* \in 1/\mathcal{C}$ such that: for $\delta \leq \delta^*$, for y with $|y - \hat{x}|_{A_\delta(x)} \leq r_*$,

$$\frac{K_1}{\delta^2} \leq p_{X_\delta}(y) \leq \frac{K_2}{\delta^2}.$$

The last two results also hold in the same form if we replace the norm $|\cdot|_{A_\delta(x)}$ with $|\cdot|_{A_\delta(\hat{x})}$, because of (3.4.5).

Remark 3.5. The weak Hörmander condition ensures the existence of the density for X_δ, but not its positivity. The fact that we have lower bounds for the density supposing just a local weak Hörmander condition might appear contradictory. In fact, our estimates are local around $\hat{x}$ (and not around x!), and this contradiction does not subsist, as we see in the following example (see for instance [40] (3.2.6)). Take

$$X_t^1 = 1 + W_t, \qquad X_t^2 = \int_0^t (X_s^1)^2 ds.$$

Clearly $p_{X_d}(y) = 0$ for any $y \in \mathbb{R}^-$. We are taking $X_0 = x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and weak Hörmander holds locally around this point, but $p_{X_d}(x) = 0 \ \forall \delta > 0$.

$$\sigma(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad b(x) = \begin{pmatrix} 0 \\ x_1^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad [\sigma, b](x) = \begin{pmatrix} 0 \\ 2x_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

The set $\{y : |y - \hat{x}|_{A_\delta(x)} \leq r_*\}$, on which (3.3.2) holds, is included in $\mathbb{R}^+$, so this is not in contrast with the fact that the density $p_{X_d} = 0$ on $\mathbb{R}^-$. Indeed y satisfies

$$|y - \hat{x}|_{A_\delta(x)} = \sqrt{\delta^{-1}(y_1 - 1)^2 + \frac{1}{4}\delta^{-3}(y_2 - \delta)^2} \leq r_*$$

For $y_2 < 0$,

$$|y - \hat{x}|_{A_\delta} \leq r_* \Rightarrow \frac{1}{2}\delta^{-1/2} \leq r_* \Rightarrow \delta \geq \frac{1}{4r_*^2} \geq \delta^* \exp(-2Lr_*^2)$$

for any choice of $r_* > 0$, if $\delta^* \leq \frac{1}{4}$, and this is in contrast with (3.3.1).

3.3.1 Development

In this section $x \in \mathbb{R}^2$ will be fixed. It represents the initial condition of the diffusion process X_t (so $X_0 = x$). In order to lighten the notation we will not mention it (so, for example, we denote A instead of $A(x)$, and so on). We write the stochastic Taylor development of X_t with a reminder of order t^2. We need to introduce some notation. Consider a small time $\delta \in (0,1)$. We define

- $\hat{x} = x + b(x)\delta$
- The matrices $\bar{A}$ and $\bar{A}_\delta$ as

$$\bar{A} = (\bar{A}_1, A_2), \qquad \text{with } \bar{A}_1 = \sigma + \delta\partial_b\sigma, \quad A_2 = [\sigma, b]$$

  and

$$\bar{A}_\delta = \left(\delta^{1/2}\bar{A}_1, \delta^{3/2}A_2\right).$$

  Remark that from **H1** these matrices are always invertible if δ is small enough.

- The Gaussian r.v.

$$\Theta = \begin{pmatrix} \Theta_1 \\ \Theta_2 \end{pmatrix} = \begin{pmatrix} \delta^{-1/2}W_\delta \\ \delta^{-3/2}\int_0^\delta(\delta - s)dW_s \end{pmatrix}$$

- The polynomial of degree 3 and direction σ (recall κ_σ defined in (3.2.2)):

$$\eta(u) = \left(\frac{\kappa_\sigma(x)}{2}u^2 + \frac{(\partial_\sigma\kappa_\sigma + \kappa_\sigma^2)(x)}{6}u^3\right)\sigma(x).$$

- $G = \Theta + \tilde{\eta}(\Theta)$, where $\tilde{\eta}(\Theta) = \bar{A}_\delta^{-1}\eta(\delta^{1/2}\Theta_1)$
- The remainder R_δ:

$$\begin{aligned} R_\delta = & \int_0^\delta \int_0^s \partial_b\sigma(X_u) - \partial_b\sigma(x)du \circ dW_s + \\ & \int_0^\delta \int_0^s \partial_\sigma b(X_u) - \partial_\sigma b(x) \circ dW_u ds + \\ & \int_0^\delta \int_0^s \partial_b b(X_u)duds + \\ & \int_0^\delta \int_0^s \int_0^u \partial_\sigma\partial_\sigma\sigma(X_v) - \partial_\sigma\partial_\sigma\sigma(x) \circ dW_v \circ dW_u \circ dW_s + \\ & \int_0^\delta \int_0^s \int_0^u \partial_b\partial_\sigma\sigma(X_v) \circ dv \circ dW_u \circ dW_s. \end{aligned}$$

  Notice that $\|R_\delta\|_{2,p} \leq C\delta^2$. We also denote $\tilde{R}_\delta := \bar{A}_\delta^{-1}R_\delta$.

We now prove that the following decomposition holds:

$$X_\delta = \hat{x} + \bar{A}_\delta(G + \tilde{R}_\delta) \tag{3.3.3}$$

This is a main tool in our approach. Indeed we are able to find exponential bounds for the density of the variable $F := G + \tilde{R}_\delta$ in the Euclidean metric of $\mathbb{R}^2$. The fact that in Theorem 3.3 the bounds for the diffusion are in the $A_\delta(x)$-norm follows from the change of variable suggested by (3.3.3).

Let us prove (3.3.3). With a stochastic Taylor development we obtain

$$X_t = x + b(x)t + U_t + R_t$$

where

$$\begin{aligned} U_t &= \sigma(x)W_t + \partial_\sigma\sigma(x)\int_0^t W_s \circ dW_s \\ &+ \partial_\sigma\partial_\sigma\sigma(x)\int_0^t\int_0^s W_u \circ dW_u \circ dW_s \\ &+ \partial_b\sigma(x)\int_0^t sdW_s + \partial_\sigma b(x)\int_0^t W_s ds \end{aligned}$$

Now we write

$$\begin{aligned} \int_0^t W_s ds &= \int_0^t (t-s)dW_s \\ \int_0^t sdW_s &= -\int_0^t (t-s)dW_s + tW_t \end{aligned}$$

Therefore

$$\begin{aligned} U_t &= (\sigma(x) + t\partial_b\sigma(x))W_t + (\partial_\sigma b(x) - \partial_b\sigma(x))\int_0^t (t-s)dW_s \\ &+ \partial_\sigma\sigma(x)\frac{W_t^2}{2} + \partial_\sigma\partial_\sigma\sigma(x)\frac{W_t^3}{6} \end{aligned}$$

So we have the following decomposition of X_t:

$$X_t = x + b(x)t + (\sigma(x) + t\partial_b\sigma(x))W_t + [\sigma, b](x)\int_0^t (t-s)dW_s + \eta(W_t) + R_t$$

where x is the initial condition. With η we denote the polynomial of degree 3. Remark that **H3** implies that both the coefficients of this polynomial have the same direction as σ:

$$\eta(u) = \frac{\partial_\sigma\sigma(x)}{2}u^2 + \frac{\partial_\sigma\partial_\sigma\sigma(x)}{6}u^3 = \left(\frac{\kappa_\sigma(x)}{2}u^2 + \frac{(\partial_\sigma\kappa_\sigma + \kappa_\sigma^2)(x)}{6}u^3\right)\sigma(x).$$

3.3.2 Density of the rescaled diffusion

We prove in this section the following theorem for $F = G + \tilde{R}_\delta$.

Lemma 3.6. *There exist L_1, L_2, K_1, K_2 positive constants, $\delta^* \in 1/\mathcal{C}$ such that: for any $r_* > 0$, for*

$$\delta \leq \delta^* \exp\left(-2L_1(r_*)^2\right)$$

and $|z| \leq r_*$,

$$K_1 \exp\left(-L_1|z|^2\right) \leq p_F(z) \leq K_2 \exp\left(-L_2|z|^2\right).$$

Remark 3.7. With a simple change of variable we have that for $|y - \hat{x}|_{\bar{A}_\delta(x)} \leq r_*$,

$$\frac{K_1}{|\det \bar{A}_\delta(x)|} \exp\left(-L_1|y - \hat{x}|^2_{\bar{A}_\delta(x)}\right) \leq p_{X_\delta}(y)$$
$$\leq \frac{K_2}{|\det \bar{A}_\delta(x)|} \exp\left(-L_2|y - \hat{x}|^2_{\bar{A}_\delta(x)}\right).$$

These estimates and (3.4.4) imply Theorem 3.3

Proof. STEP 1: In the proof, wlog, we suppose $r_* \geq 1$. In what follows, $C \in \mathcal{C}$, and may vary from line to line. We start by computing the derivatives of η:

$$\begin{aligned}
\eta(y) &= \left(\frac{\kappa_\sigma}{2} y^2 + \frac{\partial_\sigma \kappa_\sigma + \kappa_\sigma^2}{6} y^3\right)\sigma \\
\eta'(y) &= \left(\kappa_\sigma y + \frac{\partial_\sigma \kappa_\sigma + \kappa_\sigma^2}{2} y^2\right)\sigma \\
\eta''(y) &= (\kappa_\sigma + (\partial_\sigma \kappa_\sigma + \kappa_\sigma^2)y)\sigma \\
\eta'''(y) &= (\partial_\sigma \kappa_\sigma + \kappa_\sigma^2)\sigma.
\end{aligned}$$

By the definition of $\bar{A}_\delta^{-1}$,

$$\bar{A}_\delta^{-1}\delta^{1/2}(\sigma + \delta\partial_b\sigma) = (1,\, 0)^T.$$

Therefore

$$\bar{A}_\delta^{-1}\sigma = \delta^{-1/2}(1,\, 0)^T - \bar{A}_\delta^{-1}\delta\partial_b\sigma.$$

By (3.4.2) we have $|\bar{A}_\delta^{-1}\delta\partial_b\sigma| \leq \delta^{-1/2}$, so that $|\bar{A}_\delta^{-1}\sigma| \leq C\delta^{-1/2}$. We stress that this upper bound is $\delta^{-1/2}$ in contrast with $\delta^{-3/2}$ in (3.4.2), because $\bar{A}_\delta$ works in the specific direction σ. Now we can estimate the norms of $\tilde{\eta}$ and its derivatives. Since they are collinear with σ, we have

$$\begin{aligned}
|\tilde{\eta}(u)| &= |\bar{A}_\delta^{-1}\, \eta(\delta^{1/2}u_1)| \leq C(|u_1|^2\delta^{1/2} + |u_1|^3\delta) \\
|\partial_{u_1}\tilde{\eta}(u)| &= |\bar{A}_\delta^{-1}\delta^{1/2}\eta'(\delta^{1/2}u_1)| \leq C(|u_1|\delta^{1/2} + |u_1|^2\delta) \\
|\partial_{u_1}^2\tilde{\eta}(u)| &= |\bar{A}_\delta^{-1}\delta\eta''(\delta^{1/2}u_1)| \leq C(\delta^{1/2} + |u_1|\delta) \\
|\partial_{u_1}^3\tilde{\eta}(u)| &= |\bar{A}_\delta^{-1}\delta^{3/2}\eta'''(\delta^{1/2}u_1)| \leq C\delta \\
|\partial_{u_2}\tilde{\eta}(u)| &= 0.
\end{aligned}$$

So, referring to the notation of Section 2.2.3, we have

$$\begin{aligned} c_*(\tilde\eta,h) &= \sup_{|u|\le 2h}\max_{i,j}\left|\partial_i\tilde\eta^j(u)\right| \le Ch\delta^{1/2} \\ c_2(\tilde\eta) &= \max_{i,j}\sup_{|u|\le 1}\left|\partial^2_{i,j}\tilde\eta(u)\right| \le C\delta^{1/2} \\ c_3(\tilde\eta) &= \max_{i,j,k}\sup_{|u|\le 1}\left|\partial^3_{i,j,k}\tilde\eta(u)\right| \le C\delta. \end{aligned} \tag{3.3.4}$$

We first want to apply Lemma 2.13 to $G=\Theta+\tilde\eta(\Theta)$. Here $d=2$, and the covariance matrix of Θ is

$$\gamma_\Theta = Q = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

It has 2 positive eigenvalues, $0<\lambda_1<\lambda_2$, and $\det(Q)=1/12$. We are supposing here $\delta\le\delta^*\exp\left(-2L_1r_*^2\right)\le\delta^*/r_*^2$. Since

$$h_\eta = \frac{1}{64(c_2(\eta)+\sqrt{c_3(\eta)})} \ge \frac{1}{C_3\sqrt{\delta}} \ge \frac{r_*}{C_3\sqrt{\delta^*}}$$

and

$$c_*(\eta,16r_*) \le C_4 r_*\sqrt{\delta} \le C_4\sqrt{\delta^*},$$

choosing $\delta^*\le\frac{1}{16}\frac{\lambda_1}{\lambda^2}\frac{1}{C_3^2C_4^2}$ the conditions (2.2.15) are satisfied:

$$c_*(\eta,16r_*) \le \frac14\sqrt{\frac{\lambda_1}{\lambda_2}}, \qquad r_*\le h_\eta.$$

So there exist L_1,L_2,K_1,K_2 universal constants, such that for $|z|\le r_*$,

$$K_1\exp\left(-L_1|z|^2\right) \le p_{G,U}(z) \le K_2\exp\left(-L_2|z|^2\right). \tag{3.3.5}$$

STEP 2 (lower bound for p_F). From (3.3.5), using theorem 2.4, we recover estimates for p_F. We start checking that C_1,C_2 in Remark 2.5 are in $\mathcal{C}$. $n=2$, and from(2.1.4) and $r_*\ge1$,

$$m_U(1,64) \le C\left(1+\frac{\|\Theta\|_{2,64}}{r_*}\right)^2 \le C\in\mathcal{C}.$$

Now we consider $\Gamma_{G,U}=1+(\mathbb{E}_U\lambda_*(\gamma_G)^{-p})^p$.

$$\begin{aligned} \langle\gamma_G\xi,\xi\rangle &= \int\langle D_sG,\xi\rangle^2 \\ &\ge \int\frac12\langle D_s\Theta,\xi\rangle^2-\langle D_s\eta(\Theta),\xi\rangle^2ds \\ &= S_1+S_2. \end{aligned}$$

We have

$$S_2=\int\langle\nabla\eta(\Theta)D_s\Theta,\xi\rangle^2ds=\int\langle D_s\Theta,\nabla\eta(\Theta)^T\xi\rangle^2ds\le\lambda_2|\nabla\eta(\Theta)|^2|\xi|^2$$

and $S_1 \geq \lambda_1/2$, so

$$\lambda_*(\gamma_G) \geq \lambda_1 \left(\frac{1}{2} - \frac{\lambda_2}{\lambda_1} |\nabla \eta(\Theta)|^2 \right).$$

From and $c_*(\eta, 16r_*) \leq \frac{1}{4}\sqrt{\frac{\lambda_1}{\lambda_2}}$ (proved before) and $|\nabla\eta(\Theta)| \leq 2c_*(\eta, \Theta)$ follows

$$|\nabla\eta(\Theta)| \leq \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_2}},$$

and therefore $4\lambda_*(\gamma_G) \geq \lambda_1$, which implies $\Gamma_{G,U}(32n) \leq C$. Standard computations give $n_{F,G,U}(1,p) \leq C \in \mathcal{C}$, so from corollary 2.4 we have that $\exists C \in \mathcal{C}$ such that for $|z| \leq r$

$$p_F(z) \geq p_{G,U}(z) - C\|\tilde{R}_\delta\|_{64,U} \geq K_1 \exp\left(-L_1|z|^2\right) - C\|\tilde{R}_\delta\|_{64,U}.$$

Notice

$$\|\tilde{R}_\delta\|_{64,U} = \|\bar{A}_\delta^{-1} R_\delta\|_{64,U} \leq C\delta^2/\delta^{3/2} = C\sqrt{\delta},$$

so $p_F(z) \geq K_1 \exp(-L_1|z|^2) - C_5\sqrt{\delta}$, $\exists C_5 \in \mathcal{C}$. We have that if $\delta \leq \left(\frac{K_1}{2C_5}\right)^2 \exp(-2L_1 r_*^2)$,

$$p_F(z) \geq \frac{K_1}{2} \exp\left(-L_1|z|^2\right).$$

So taking

$$\delta^* \leq \frac{1}{16}\frac{\lambda_1}{\lambda_2}\frac{1}{C_3^2 C_4^2}\left(\frac{K_1}{2C_5}\right)^2$$

the estimate holds.
STEP 3 (upper bound for p_F). The proof of the upper bound follows again from in theorem 2.4. We deal with C_2 exactly as for the lower bound, with the difference that we need $\Gamma_F(64) < \infty$, instead of $\Gamma_{G,U}(64) < \infty$. This fact is proved in Lemma 3.8.

We also need to prove that $\|1-U\|_{1,14n}$ decays as $C\exp(-L|z|^2) \leq C\exp(-Lr^2)$. This follows from (2.1.5):

$$\|1-U\|_{1,28} \leq \sum_{i=1,2} \mathbb{P}(|\Theta_i| > r)^{\frac{1}{56}} C(1+1/r) \leq Ce^{-Lr_*^2}.$$

□

The moments of $\hat{\gamma}_F$ are bounded, and these bounds do not depend on δ. This result looks interesting by itself, since it means that we are able to account precisely of the of the scaling of the diffusion in the two main directions σ and $[\sigma, b]$. In this particular case this is a refinement of the classical result on the bounds of the Malliavin covariance under the (weak) Hörmander condition (see *Norris Lemma* in [79], or [71]).

Lemma 3.8. *For any fixed $p > 1$, exists $\delta_* \in 1/\mathcal{C}$, $C \in \mathcal{C}$, such that for any $\delta \leq \delta_*$, $\Gamma_F(p) \leq C_p$.*

(In particular, this is true for $p = 32n = 64$).

Proof. Following [79] we define the tangent flow of X as the derivative with respect to the initial condition of X, $Y_t := \partial_x X_t$. We also denote its inverse $Z_t = Y_t^{-1}$. They satisfy the following stochastic differential equations

$$\begin{aligned} Y_t &= Id + \int_0^t \nabla\sigma(X_s)Y_s \circ dW_s + \int_0^t \nabla b(X_s)Y_s ds \\ Z_t &= Id - \int_0^t Z_s\nabla\sigma(X_s) \circ dW_s - \int_0^t Z_s\nabla b(X_s)ds \end{aligned}$$

Then

$$D_sX_t = Y_tZ_s\sigma(X_s),$$

and

$$D_sF = D_s\bar{A}_\delta^{-1}(X_\delta - \hat{x}) = \bar{A}_\delta^{-1}Y_\delta Z_s\sigma(X_s).$$

We define

$$\bar{\gamma}_\delta = \int_0^\delta A_\delta^{-1}Z_s\sigma(X_s)\sigma(X_s)^TZ_s^TA_\delta^{-1,T}ds.$$

and then

$$\gamma_F = \langle DF, DF\rangle = \bar{A}_\delta^{-1}Y_\delta A_\delta\bar{\gamma}_\delta A_\delta^TY_\delta^T\bar{A}_\delta^{-1,T}.$$

Remark that

$$\gamma_F^{-1} = \bar{A}_\delta^TZ_\delta^TA_\delta^{-1,T}\bar{\gamma}_\delta^{-1}A_\delta^{-1}Z_\delta\bar{A}_\delta,$$

and that in this representation we have both A_δ and its "perturbed" version $\bar{A}_\delta$. We have to check the integrability of $A_\delta^{-1}Z_\delta\bar{A}_\delta$, which we expect to be close to the identity matrix for small δ, and the integrability of $\bar{\gamma}_\delta^{-1}$. We use the following representation, holding for general ϕ, which follows applying Ito's formula (details in [79])

$$Z_t\phi(X_t) = \phi(x) + \int_0^t Z_s[\sigma,\phi](X_s)dW_s^k + \int_0^t Z_s\left\{[b,\phi] + \frac{1}{2}[\sigma,[\sigma,\phi]]\right\}(X_s)ds \quad (3.3.6)$$

In our framework $d = 1$, $\sigma_1 = \sigma$, $\sigma_0 = b$. Taking $\phi = \sigma$ the representation above reduces to

$$\begin{aligned} Z_t\sigma(X_t) &= \sigma(x) + \int_0^t Z_s[b,\sigma](X_s)ds \\ &= \sigma(x) + t[b,\sigma](x) + L_t, \end{aligned}$$

with $L_t = \int_0^t Z_s[b,\sigma](X_s) - Z_0[b,\sigma](x)ds$. We have

$$\begin{aligned} A_\delta^{-1}Z_s\sigma(X_s) &= A_\delta^{-1}(\sigma(x) + s[b,\sigma](x) + L_s) \\ &= \frac{1}{\delta^{1/2}}\begin{pmatrix} 1 \\ -s/\delta \end{pmatrix} + L_s', \end{aligned}$$

with $L'_s = A_\delta^{-1} L_s$. Standard computations show

$$\mathbb{E}\left[\left|\int_0^{\delta\varepsilon} L_s L_s^T ds\right|^q\right] \le C(\delta\varepsilon)^{4q}, \quad \forall q > 0, \quad \exists C \in \mathcal{C}.$$

For constant c, that we will chose in the sequel, and fixed ε, we introduce the stopping time

$$S_\varepsilon = \inf\left\{s \ge 0 : \left|\int_0^s L_u L_u^T du\right| \ge c(\delta\varepsilon)^3\right\} \wedge \delta,$$

Remark that for any $q > 0$

$$\begin{aligned}\mathbb{P}(S_\varepsilon < \delta\varepsilon) &\le \mathbb{P}\left(\left|\int_0^{\delta\varepsilon} L_s L_s^T ds\right|^q \ge c^q(\delta\varepsilon)^{3q}\right) \\ &\le \frac{\mathbb{E}\left[\left|\int_0^{\delta\varepsilon} L_s L_s^T ds\right|^q\right]}{c^q(\delta\varepsilon)^{3q}} \\ &\le \frac{C(\delta\varepsilon)^{4q}}{c^q(\delta\varepsilon)^{3q}} \\ &\le C/c^q\,(\delta\varepsilon)^q \le \varepsilon^q\end{aligned} \tag{3.3.7}$$

for $\delta \le \delta_q$. Now we suppose to be on $\frac{S_\varepsilon}{\delta} \ge \varepsilon$. Applying the inequality

$$\langle (v+R)(v+R)^T\xi, \xi\rangle \ge \frac{1}{2}\langle vv^T\xi,\xi\rangle - \langle RR^T\xi,\xi\rangle,$$

which holds for any vectors v, R, ξ, we obtain

$$\begin{aligned}\bar{\gamma}_\delta &= \int_0^\delta A_\delta^{-1} Z_s \sigma(X_s)\sigma(X_s)^T Z_s^T A_\delta^{-1,T} ds \\ &\ge \int_0^{S_\varepsilon} A_\delta^{-1} Z_s \sigma(X_s)\sigma(X_s)^T Z_s^T A_\delta^{-1,T} ds \\ &= \int_0^{S_\varepsilon} \frac{1}{\delta}\begin{pmatrix} 1 & -s/\delta \\ -s/\delta & (s/\delta)^2\end{pmatrix} ds - A_\delta^{-1}\int_0^{S_\varepsilon} L_s L_s^T ds A_\delta^{-1,T}.\end{aligned}$$

We have

$$\int_0^{S_\varepsilon} \frac{1}{\delta}\begin{pmatrix} 1 & -s/\delta \\ -s/\delta & (s/\delta)^2\end{pmatrix} ds \ge \int_0^{\delta\varepsilon} \frac{1}{\delta}\begin{pmatrix} 1 & -s/\delta \\ -s/\delta & (s/\delta)^2\end{pmatrix} ds \ge \begin{pmatrix} \varepsilon & -\frac{\varepsilon^2}{2} \\ -\frac{\varepsilon^2}{2} & \frac{\varepsilon^3}{3}\end{pmatrix} \ge \frac{\varepsilon^3}{16}$$

and, chosing now the constant c in the definition of S_ε small enough,

$$\left|A_\delta^{-1}\int_0^{S_\varepsilon} L_s L_s^T ds A_\delta^{-1,T}\right| \le \frac{C'}{\delta^3} c(\delta\varepsilon)^3 \le \frac{\varepsilon^3}{32}.$$

so

$$\langle \bar{\gamma}_\delta \xi, \xi\rangle \ge \frac{\varepsilon^3}{32}|\xi|^2, \quad \forall |\xi| = 1.$$

Therefore, using also (3.3.7), we have that for any q, for any $\varepsilon \le \varepsilon_0$, $\delta \le \delta_q$,

$$\mathbb{P}(\langle \bar{\gamma}_\delta \xi, \xi \rangle < \varepsilon^3/32) \le \mathbb{P}[S_\varepsilon < \delta\varepsilon] \le \varepsilon^q.$$

Now we apply Lemma 3.9. Remark that we do not need all the moments but just up to a certain q, and so the estimate we find holds uniformly in δ for $\delta \le \delta_0$. We are left with the estimate of $A_\delta^{-1} Z_\delta \bar{A}_\delta$. Applying (3.3.6) and **H3**, one can prove that

$$Z_t \sigma(x) = (1 - \kappa_\sigma(x) W_t)\sigma(x) + J_t,$$

with $\mathbb{E}|J_t|^p \le C t^p$. So

$$Z_\delta \bar{A}_\delta = \left(\sqrt{\delta}(1 - \kappa_\sigma(x) W_\delta)\sigma(x),\, 0\right) + M_\delta$$

where $M_\delta \in \mathbb{R}^2$ with $\mathbb{E}|M_\delta|^q \le C\delta^{3q/2}$, $C \in \mathcal{C}$. Since $A_\delta = (\delta^{1/2}\sigma, \delta^{3/2}[\sigma, b])$, this implies

$$A_\delta^{-1} Z_\delta \bar{A}_\delta \le C \in \mathcal{C}.$$

□

The following lemma is a slight modification of Lemma 2.3.1. in [79].

Lemma 3.9. *Let γ be a symmetric nonnegative definite $n \times n$ matrix. Denoting $|C| = \sum_{1 \le i,j \le n} |\gamma^{i,j}|^2)^{1/2}$, we assume that $\mathbb{E}|C|^{p+1} < \infty$, and that for $\varepsilon \le \varepsilon_{p+2n}$,*

$$\sup_{|\xi|=1} \mathbb{P}[\langle \gamma\xi, \xi \rangle < \varepsilon] \le \varepsilon^{p+2n}$$

Then

$$\mathbb{E}\lambda_*(\gamma)^{-p} \le C\varepsilon_{p+2n}^{-p}.$$

Proof. Denote with $\lambda := \lambda_*(\gamma) = \inf_{|\xi|=1}\langle \gamma\xi, \xi \rangle$ the smallest eigenvalue of γ. It is proved in [79] that for every $p > 2n$ under our hypothesis

$$\mathbb{P}[\lambda < \varepsilon] \le K\mathbb{E}|C|^p \varepsilon^p.$$

for any $\varepsilon \le \varepsilon_{p+2n}$. We have

$$\begin{aligned}\mathbb{E}\left[\lambda^{-p}\right] &\le p\int_0^\infty \mathbb{P}[1/\lambda > x)x^{p-1}dx \\ &= p\int_0^{1/\varepsilon_{p+2n}} \mathbb{P}[1/\lambda > x)x^{p-1}dx + p\int_{1/\varepsilon_{p+2n}}^\infty \mathbb{P}[1/\lambda > x)x^{p-1}dx\end{aligned}$$

The first term is bounded by $(1/\varepsilon_{p+2n})^p$. For the second we apply the computations above:

$$\begin{aligned}p\int_{1/\varepsilon_{p+2n}}^\infty \mathbb{P}[\lambda < 1/x)x^{p-1}dx &\le pK\mathbb{E}|C|^{p+1}\int_{1/\varepsilon_{p+2n}}^\infty (1/x)^{p+1}x^{p-1}dx \\ &= K\mathbb{E}|C|^{p+1}\varepsilon_{p+2n}.\end{aligned}$$

□

3.3.3 Tubes estimates on the diffusion

As an application of Theorem 3.3, we are now able to prove the following tube estimate. We consider the diffusion on $[0,T]$, and for $\phi \in L^2[0,T]$, let

$$x_t(\phi) = x_0 + \int_0^t \sigma(x_s(\phi))\phi_s ds + \int_0^t b(x_s(\phi))ds, \qquad \text{for } t \in [0,T].$$

Recall that we suppose $|\phi_.|^2$, $\lambda_.$, $n_. \in L(\mu,h)$, for some $h \in \mathbb{R}_{>0}$, $\mu \geq 1$, where $L(\mu,h)$ is the class of non-negative functions which have the property

$$f(t) \leq \mu f(s) \quad \text{ for } |t-s| \leq h.$$

We denote in this section, for fixed $t \in [0,T]$,

$$\mathcal{C}_t = \left\{ C_t = K\left(n_t/\lambda_t\right)^q, \exists K, q \geq 1 \text{ universal constants} \right\},$$

Denote, for K_* and q_* constants,

$$R_*(\phi) = \inf_{0\leq t\leq T} \left(\frac{1}{K_*}\frac{\lambda_t}{\mu n_t}\right)^{q_*} \left(h \wedge \inf\left\{\delta / \int_t^{t+\delta} |\phi_s|^2 ds \ : \quad t \in [0,T], \delta \in [0,h]\right\}\right) \tag{3.3.8}$$

Theorem 3.10. *There exist positive constants K,q such that for $R \in]0,1]$ holds*

$$\exp\left(-K\int_0^T \left(\frac{\mu n_t}{\lambda_t}\right)^q \left(\frac{1}{h} + \frac{1}{R} + |\phi_t|^2 dt\right)\right) \leq \mathbb{P}\left(\sup_{t\leq T} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} \leq 1\right).$$

Moreover, exist constants K, q, K_*, q_* such that for $R \leq R_*(\phi)$ holds

$$\mathbb{P}\left(\sup_{t\leq T} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} \leq 1\right) \leq \exp\left(-\int_0^T \frac{1}{K}\left(\frac{\lambda_t}{\mu n_t}\right)^q \left(\frac{1}{R} + |\phi_t|^2\right) dt\right)$$

Remark 3.11. For $R \leq R_*(\phi) \leq h$ the lower bound holds as in (3.2.3)

Proof. A main point in this proof is the choice a sequence of short time intervals in a way such that we are able to apply the short time density estimate. This issue is related to the choice of a an "elliptic evolution sequence" in [11]. We write x_t for $x_t(\phi)$ to have a more readable notation. We start proving the lower bound.
STEP 1: We set, for large q_1, K_1 to be fixed in the sequel,

$$f_R(t) = K_1 \left(\frac{\mu n_t}{\lambda_t}\right)^{q_1} \left(\frac{1}{h} + \frac{1}{R} + |\phi_t|^2\right).$$

Recall **H4**: $|\phi_.|^2, n_., \lambda_. \in L(\mu,h)$, $\exists \mu \geq 1$, $0 < h \leq 1$, where

$$L(\mu,h) = \{f : f(t) \leq \mu f(s) \quad \text{for} \quad |t-s| \leq h\}.$$

This implies $f_R \in L(\mu^{2q_1+1}, h)$. We also define

$$\delta(t) = \inf_{\delta>0} \left\{ \int_t^{t+\delta} f_R(s)ds \geq \frac{1}{\mu^{2q_1+1}} \right\}. \tag{3.3.9}$$

Clearly $\delta(t) \leq h$, so we can use on the intervals $[t, t+\delta(t)]$ the fact that our bounds are in $L(\mu, h)$. If $0 < t - t' \leq h$,

$$\mu^{2q_1+1} f_R(t)\delta(t) \geq \int_t^{t+\delta(t)} f_R(s)ds = 1 = \int_{t'}^{t'+\delta(t')} f_R(s)ds \geq \mu^{-(2q_1+1)} f_R(t)\delta(t'),$$

so $\delta(t')/\delta(t) \leq \mu^{4q_1+2}$. Also the converse holds, and $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$. We set

$$\varepsilon(t) = \left(\int_t^{t+\delta(t)} |\phi_s|^2 ds \right)^{1/2}.$$

We have

$$\frac{1}{\mu^{2q_1+1}} = \int_t^{t+\delta(t)} f_R(s)ds \geq \int_t^{t+\delta(t)} \frac{f_R(t)}{\mu^{2q_1+1}} ds \geq \delta(t) \frac{f_R(t)}{\mu^{2q_1+1}},$$

so

$$\delta(t) \leq \frac{1}{f_R(t)} \leq \frac{R}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}. \tag{3.3.10}$$

Similarly,

$$\frac{1}{\mu^{2q_1+1}} \geq \int_t^{t+\delta(t)} K_1 \left(\frac{\mu n_s}{\lambda_s} \right)^{q_1} |\phi_s|^2 ds \geq \frac{1}{\mu^{2q_1}} K_1 \left(\frac{\mu n_t}{\lambda_t} \right)^{q_1} \varepsilon(t)^2,$$

and we can write both

$$\delta(t) \leq \frac{1}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}, \quad \text{and} \quad \varepsilon(t)^2 \leq \frac{1}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}. \tag{3.3.11}$$

We set our time grid as

$$t_0 = 0; \quad t_k = t_{k-1} + \delta(t_{k-1}),$$

and introduce the following notation on the grid:

$$\delta_k = \delta(t_k); \quad \varepsilon_k = \varepsilon(t_k); \quad n_k = n(x_k); \quad \lambda_k = \lambda(x_k); \quad X_k = X_{t_k}; \quad x_k = x_{t_k}.$$

We also define

$$\hat{X}_k = X_k + b(X_k)\delta_k; \quad \hat{x}_k = x_k + b(x_k)\delta_k,$$

and for $t_k \leq t \leq t_{k+1}$,

$$\hat{X}_k(t) = X_k + b(X_k)(t - t_k); \quad \hat{x}_k(t) = x_k + b(x_k)(t - t_k).$$

Moreover we denote

$$|\xi|_k = |\xi|_{A_{\delta_k}(x_k)}; \quad \mathcal{C}_k = \mathcal{C}_{t_k},$$

and $r_*^k \in \mathcal{C}_k$ the ray r_* of remark 3.4 associated to x_k. Lemmas 3.14, 3.15, 3.16, 3.17 hold for δ_k and ε_k small enough, and in particular Lemma 3.17 says that

$$\frac{1}{C_k^1}|\xi|_{A_\delta(x_k)} \leq |\xi|_{A_\delta(x_{k+1})} \leq C_k^1|\xi|_{A_\delta(x_k)}, \tag{3.3.12}$$

for some $C_k^1 \in \mathcal{C}_k$, for any $\delta \leq \delta_k$. Moreover we have $|x_{k+1} - \hat{x}_k|_k \leq C_k(\varepsilon_k \vee \delta_k)$, and for all $t_k \leq t \leq t_{k+1}$, applying also (3.4.1), $|x_t - \hat{x}_k(t)|_{A_R(x_t)} \leq C_k(\varepsilon_k \vee \delta_k)$ for all $R \geq \delta_k$. Recall (3.3.11), and we fix q_3, K_3 such that, for $q_1 \geq q_3, K_1 \geq K_3$, the lemmas 3.14, 3.15, 3.16, 3.17 hold and

$$|x_{k+1} - \hat{x}_k|_k \leq r_*^k/8 \tag{3.3.13}$$

$$|\hat{x}_k(t) - x_t|_{A_R(x_t)} \leq \frac{1}{16} \text{ for all } t_k \leq t \leq t_{k+1}, \tag{3.3.14}$$

and moreover the theorem in short time 3.3 holds in the form of remark 3.4. Now, $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$ implies $\delta_k/\delta_{k+1} \leq \mu^{4q_1+2}$ and $\delta_{k+1}/\delta_k \leq \mu^{4q_1+2}$. This, (3.3.12) and (3.4.1) give

$$\frac{1}{C_k^1\mu^{2q_1+1}}|\xi|_k \leq |\xi|_{k+1} \leq \mu^{2q_1+1}C_k^1|\xi|_k, \tag{3.3.15}$$

where C_k^1 is in $\mathcal{C}_k$, depending on K_3, q_3. We now set, for K_2, q_2 to be fixed in the sequel,

$$r_k = \frac{1}{K_2\mu^{2q_1+2q_2+1}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}, \tag{3.3.16}$$

and define

$$\Gamma_k = \{|X_k - x_k|_k \leq r_k\}, \quad D_k = \{\sup_{t_k \leq t \leq t_{k+1}} |X_t - x_t|_{A_R(x_t)} \leq 1\},$$

and $\mathbb{P}_k$ as the conditional probability

$$\mathbb{P}_k(\cdot) = \mathbb{P}\left(\cdot|W_t, t \leq t_k; X_k \in \Gamma_k\right).$$

We denote p_k the density of X_{k+1} with respect to this probability. We prove that on $\{|\cdot - x_{k+1}|_{k+1} \leq r_{k+1}\}$ we can apply Theorem 3.3 in the form of remark 3.4 to p_k and so there exists $\underline{C}_k \in \mathcal{C}_k$ such that

$$\frac{1}{\underline{C}_k\delta_k^2} \leq p_k(y) \tag{3.3.17}$$

We estimate

$$|y - \hat{X}_k|_k \leq |y - x_{k+1}|_k + |x_{k+1} - \hat{x}_k|_k + |\hat{x}_k - \hat{X}_k|_k. \tag{3.3.18}$$

We already have (3.3.13). Since we are on $|y - x_{k+1}|_{k+1} \leq r_{k+1}$, from (3.3.15) and the fact that $r_{k+1}/r_k \leq \mu^{2q_2}$

$$|y - x_{k+1}|_k \leq C_k^1\mu^{2q_1+1}|y - x_{k+1}|_{k+1} \leq C_k^1\mu^{2q_1+1}r_{k+1} \leq C_k^1\mu^{2q_1+2q_2+1}r_k \leq \frac{C_k^1}{K_2}\left(\frac{\lambda_k}{n_k}\right)^{q_2}.$$

It also holds $|\hat{x}_k - \hat{X}_k|_k \leq C_k|x_k - X_k|_k \leq C_k r_k$, for some $C_k \in \mathcal{C}_k$. Similarly, $|\hat{x}_k(t) - \hat{X}_k(t)|_{A_R(x_t)} \leq C_k r_k$, for all $t_k \leq t \leq t_{k+1}$. Recalling (3.3.16), we can fix K_2, q_2 such that $|y - x_{k+1}|_k \leq r_*^k/16$, $|\hat{x}_k - \hat{X}_k|_k \leq r_*^k/16$, and

$$|\hat{X}_k(t) - \hat{x}_k(t)|_{A_R(x_t)} \leq 1/4, \quad \text{for all } t_k \leq t \leq t_{k+1}. \tag{3.3.19}$$

From (3.3.18), (3.3.13) this implies $|y - \hat{X}_k|_k \leq r_*^k/4$. We also have, from (3.4.2), $|x_k - X_k| \leq |x_k - X_k|_k \lambda_k \sqrt{\delta_k}$, so we can also fix K_2, q_2 such that $r_k\lambda_k \leq 1/C$ in lemma 3.15. Therefore

$$\frac{1}{4}|\xi|_k \leq |\xi|_{A_{\delta_k}(X_k)} \leq 4|\xi|_k.$$

So $|y - \hat{X}_k|_{A_{\delta_k}(X_k)} \leq r_*^k$. Now, also from lemma 3.15 and (3.3.15)

$$\begin{aligned}\{|\cdot - x_{k+1}|_{A_{\delta_k}(X_k)} \leq r_{k+1}/(4C_k^1\mu^{2q_1+1})\} &\subset \{|\cdot - x_{k+1}|_k \leq r_{k+1}/(C_k^1\mu^{2q_1+1})\} \\ &\subset \{|\cdot - x_{k+1}|_{k+1} \leq r_{k+1}\},\end{aligned}$$

and $r_{k+1}/(4C_k^1\mu^{2q_1+1}) \geq r_k/(4C_k^1\mu^{2q_1+2q_2+1}) = \frac{1}{4C_k^1K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}$. So

$$Leb(|\cdot - x_{k+1}|_{k+1} \leq r_{k+1}) \geq \delta_k^2 \det A(x_{k+1}) \left(\frac{1}{4C_k^1 K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^2.$$

Now, $\det A(x_{k+1}) \geq \lambda_{k+1} \geq \lambda_k/\mu$ so, from (3.3.17),

$$\mathbb{P}_k(\Gamma_{k+1}) \geq \underline{C}_k \left(\frac{1}{4C_k^1 K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^2 \frac{\lambda_k}{\mu}$$

where $\underline{C}_k \in 1/\mathcal{C}_k$ is the constant in remark 3.4. This implies

$$2\mu^{-4q_1}\exp(-K_4(\log\mu + \log n_k - \log\lambda_k)) \leq P_k(\Gamma_{k+1})$$

for some constant K_4 (depending on K_2, K_3, q_2, q_3; on the contrary, we keep explicit the dependence in q_1, which is not fixed yet).
STEP 2: Consider now $t_k \leq t \leq t_{k+1}$. Recall the definition

$$D_k = \left\{\sup_{t_k \leq t \leq t_{k+1}} |X_t - x_t|_{A_R(x_t)} \leq 1\right\},$$

and introduce

$$E_k = \left\{\sup_{t_k \leq t \leq t_{k+1}} |X_t - \hat{X}_k(t)|_{A_R(x_t)} \leq \frac{1}{2}\right\}.$$

We decompose

$$|X_t - x_t|_{A_R(x_t)} \leq |X_t - \hat{X}_k(t)|_{A_R(x_t)} + |\hat{X}_k(t) - \hat{x}_k(t)|_{A_R(x_t)} + |\hat{x}_k(t) - x_t|_{A_R(x_t)},$$

and, from the previous part of the proof, (3.3.14) gives $|\hat{x}_k(t) - x_t|_{A_R(x_t)} \leq 1/4$, and (3.3.19) gives $|\hat{X}_k(t) - \hat{x}_k(t)|_{A_R(x_t)} \leq 1/4$. So $|X_t - x_t|_{A_R(x_t)} \leq |X_t - \hat{X}_k(t)|_{A_R(x_t)} + 1/2$, and therefore $E_k \subset D_k$. Using (3.3.3), some standard computations and the exponential martingale inequality we find that

$$\mathbb{P}_k(E_k^c) \leq \exp\left(-\frac{1}{K_5}\left(\frac{\lambda_k}{\mu n_k}\right)^{q_5}\frac{R}{\delta_k}\right)$$

for some constants K_5, q_5. From (3.3.10), $R/\delta_k \geq K_1(\mu n_k/\lambda_k)^{q_1}$, so choosing and fixing now q_1, K_1 large enough we conclude

$$\mathbb{P}_k(E_k^c) \leq \mu^{-4q_1}\exp(-K_4(\log\mu + \log n_k - \log\lambda_k)) \leq \frac{1}{2}\mathbb{P}_k(\Gamma_{k+1}),$$

so

$$\begin{aligned}\mathbb{P}_k(\Gamma_{k+1} \cap D_k) &\geq \mathbb{P}_k(\Gamma_{k+1} \cap E_k) \geq \mathbb{P}_k(\Gamma_{k+1}) - \mathbb{P}_k(E_k^c) \geq \frac{1}{2}\mathbb{P}_k(\Gamma_{k+1}) \\ &\geq \exp\left(-K_6(\log\mu + \log n_k - \log\lambda_k)\right),\end{aligned} \tag{3.3.20}$$

for some constant K_6. Let now $N(T) = \max\{k : t_k \leq T\}$. From Definition 3.3.9

$$\int_0^T f_R(t)dt = \sum_{k=1}^{N(T)}\int_{t_{k-1}}^{t_k} f_R(t)dt \geq \frac{N(T)}{\mu^{2q_1+1}}.$$

From (3.3.20),

$$\begin{aligned}\mathbb{P}\left(\sup_{t\leq T}|X_t - x_t(\phi)|_{A_R(x_t(\phi))} \leq 1\right) &\geq \mathbb{P}\left(\bigcap_{k=1}^{N(T)}\Gamma_{k+1} \cap D_k\right) \\ &\geq \prod_{k=1}^{N(T)}\exp(-K_6(\log\mu + \log n_k - \log\lambda_k)) \\ &= \exp\left(-K_6\sum_{k=1}^{N(T)}\log\mu + \log n_k - \log\lambda_k\right).\end{aligned}$$

Since

$$\begin{aligned}\sum_{k=1}^{N(T)}(\log\mu + \log n_k - \log\lambda_k) &= \mu^{2q_1+1}\sum_{k=1}^{N(T)}\int_{t_k}^{t_k+1} f_R(s)ds(\log\mu + \log n_k - \log\lambda_k) \\ &\leq \int_0^T \mu^{2q_1+1}f_R(t)\log\left(\frac{\mu^3 n_t}{\lambda_t}\right)dt,\end{aligned}$$

the lower bound follows.

STEP 3: We now prove the upper bound. Now recall (3.3.8), and $R \leq R_*(\phi)$. We define, with the same K_1, q_1 as in STEP 1 and 2,

$$g_R(t) = \frac{1}{h} + \frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7} + K_1\mu^{2q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}|\phi_t|^2.$$

for some constants $K_7 > K_1, q_7 > q_1 + 1$ to be fixed in the sequel. We define a new $\delta(t)$

$$\delta(t) = \inf_{\delta>0}\left\{\int_t^{t+\delta} g_R(s)ds \geq 1\right\}.$$

Clearly $\delta(t) \leq h$, so we can use on the intervals $[t, t+\delta(t)]$ the property of being in $L(\mu, h)$. If $0 < t - t' \leq h$,

$$\mu^{2q_7} g_R(t)\delta(t) \geq \int_t^{t+\delta(t)} g_R(s)ds = 1 = \int_{t'}^{t'+\delta(t')} g_R(s)ds \geq \mu^{-2q_7} g_R(t)\delta(t'),$$

so $\delta(t')/\delta(t) \leq \mu^{4q_7}$. Taking q_* and K_* in (3.3.8) large enough such that $q_* > 5q_1 + 1 + q_7$, $K_* > K_1K_7$,

$$\int_t^{t+\delta(t)} \frac{1}{h}ds \leq \int_t^{\delta(t)} \frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7} ds$$

and again from (3.3.8)

$$\begin{aligned}\int_t^{t+\delta(t)} K_1\mu^{2q_1+1}\left(\frac{\mu n_s}{\lambda_s}\right)^{q_1}|\phi_s|^2 ds &\leq K_1\mu^{4q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\int_t^{t+\delta(t)}|\phi_s|^2 ds \\ &\leq K_1\mu^{4q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\frac{\delta(t)}{R}\frac{1}{K_*}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_*} \\ &\leq \int_t^{t+\delta(t)} \frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7} ds.\end{aligned}$$

Therefore, since $1 = \int_t^{t+\delta(t)} g_R(s)ds$ and

$$\int_t^{t+\delta(t)} \frac{1}{R}\frac{1}{K_7\mu^{2q_7}}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7} ds \leq \int_t^{t+\delta(t)} g_R(s)ds \leq 3\int_t^{t+\delta(t)} \frac{1}{R}\frac{1}{K_7\mu^{2q_7}}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7} ds,$$

we find that forall t

$$\frac{1}{K_7\mu^{4q_7}}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7} \leq \frac{R}{\delta(t)} \leq \frac{3}{K_7}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7}$$

For q_*, K_* large enough this also implies

$$\delta(t) \leq \frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}.$$

We set

$$\varepsilon(t) = \left(\int_t^{t+\delta(t)} |\phi_s|^2 ds \right)^{1/2}.$$

We find, with the same computations as before,

$$\varepsilon(t)^2 \leq \frac{1}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}.$$

This implies that (3.3.11) also holds with this new grid, and also the lemmas used before. Since we are taking the same K_1 and q_1 as before (3.3.15) holds. For the same reason, the theorem in short time 3.3 also holds. We define

$$\Delta_k = \{|X_k - x_k|_{A_R(x_k)} \leq 1\},$$

$\tilde{\mathbb{P}}_k$ as the conditional probability $\tilde{\mathbb{P}}_k(\cdot) = \mathbb{P}\left(\cdot | W_t, t \leq t_k; X_k \in \Delta_k\right)$. As we did in STEP 1, if q_*, K_* are large enough, R is small enough and the upper bound for the density holds on Δ_{k+1}. Because of (3.3.12),

$$Leb(|\cdot - x_k|_{A_R(x_{k+1})} \leq 1) \leq Leb(|\cdot - x_k|_{A_R(x_k)} \leq 1)(C_k^1)^2 = (C_k^1)^2 \det(A(x_k))R^2.$$

Now, using the density estimate,

$$\tilde{\mathbb{P}}_k(\Delta_{k+1}) \leq (C_k^1)^2 \det(A(x_k))R^2\, \overline{C}_k\, \delta_k^{-2} \leq (C_k^1)^2 \det(A(x_k)) \overline{C}_k \left(\frac{R}{\delta}\right)^2.$$

where $\overline{C}_k$ is the constant of remark 3.4. Recall

$$\frac{R}{\delta(t)} \leq \frac{3}{K_7} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_7},$$

so we fix now K_7, q_7 large enough to have

$$\tilde{\mathbb{P}}_k(\Delta_{k+1}) \leq \exp(-K_{10})$$

for a $K_{10} > 0$. (We also fix now q_*, K_*, whose size depend on q_7, K_7). From the definition of $N(T)$

$$\int_0^T g_R(t)dt = \sum_{k=1}^{N(T)} \int_{t_{k-1}}^{t_k} g_R(t)dt = N(T).$$

As before

$$\mathbb{P}\left(\sup_{t \leq T} |X_t - x_t(\phi)|_{A_R(x_t(\phi))} \leq 1 \right) \leq \prod_{k=1}^{N(T)} \tilde{\mathbb{P}}_k(\Delta_{k+1})$$
$$\leq \prod_{k=1}^{N(T)} \exp(-K_{10}) = \exp(-K_{10}\, N(T)) \leq \exp\left(-K_{10} \int_0^T g_R(t) \right),$$

and we have the upper bound. Remark that because of the choice of $R \leq R_*(\phi)$, we can drop the dependence on h in the upper bound of Theorem 3.10. □

3.4 Matrix norm and control metric

3.4.1 Matrix norms

In this chapter we use a number of properties of norms associated to the matrix A and A_R. Recall that in general we can associate a norm to a matrix M via

$$|y|_M = \sqrt{\langle (MM^T)^{-1}y, y\rangle}.$$

In this case we take $A_R = \big(R^{1/2}\sigma, R^{3/2}[\sigma, b]\big)$, for $R > 0$. Since these are square matrices, the associated norm can be defined as well as

$$|y|_{A_R} = |A_R^{-1}y|.$$

Lemma 3.12. *For every $y \in \mathbb{R}^2$ and $0 < R \leq R' \leq 1$,*

$$(R/R')^{1/2}|y|_{A_R} \geq |y|_{A_{R'}} \geq (R/R')^{3/2}|y|_{A_R} \tag{3.4.1}$$
$$\tfrac{1}{R^{1/2}\lambda^*(A)}|y| \leq |y|_{A_R} \leq \tfrac{1}{R^{3/2}\lambda_*(A)}|y| \tag{3.4.2}$$

Proof. Writing explicitly the inequalities (3.4.1), we easily see that they are verified if $0 < R \leq R' < 1$. Taking $R' = 1$, we have

$$R^{1/2}|y|_{A_R} \geq |y|_A \geq R^{3/2}|y|_{A_R}$$

and so

$$\frac{1}{R^{1/2}\lambda^*(A)}|y| \leq |y|_{A_R} \leq \frac{1}{R^{3/2}\lambda_*(A)}|y|$$

□

Remark 3.13. Recall the following properties of matrices:

$$\forall\xi, \quad C\,|\xi|_B^2 \geq |\xi|_A^2 \Leftrightarrow C\,\left(BB^T\right)^{-1} \geq \left(AA^T\right)^{-1} \Leftrightarrow BB^T \leq C\,AA^T$$

and

$$\langle MM^T\xi, \xi\rangle = \sum_i \langle M_i, \xi\rangle^2,$$

so that for $\lambda_\#(M) := \lambda_*(MM^T)$, $\lambda^\#(M) := \lambda^*(MM^T)$

$$\lambda_\#(M) = \inf_{|\xi|=1}\sum_i \langle M_i, \xi\rangle^2 \quad \text{and} \quad \lambda^\#(M) = \sup_{|\xi|=1}\sum_i \langle M_i, \xi\rangle^2$$

where M_i are the columns of M. Taking $M = A(x) = (\sigma(x), [\sigma, b](x))$ we have in particular that

$$\lambda_*(A(x))^2|\xi|^2 \leq \langle \sigma(x), \xi\rangle^2 + \langle [\sigma, b](x), \xi\rangle^2 \leq \lambda^*(A(x))^2|\xi|^2 \quad \forall \xi \in \mathbb{R}^2 \tag{3.4.3}$$

Lemma 3.14. *There exists $C \in \mathcal{C}, \delta^* \in 1/\mathcal{C}$ such that for $\delta \leq \delta^*$, for any $\xi \in \mathbb{R}^2$,*

$$\tfrac{1}{C}|\xi|_{A_\delta(x)} \leq |\xi|_{\bar{A}_\delta(x)} \leq C|\xi|_{A_\delta(x)} \tag{3.4.4}$$
$$\tfrac{1}{C}|\xi|_{A_\delta(x)} \leq |\xi|_{A_\delta(\hat{x})} \leq C|\xi|_{A_\delta(x)} \tag{3.4.5}$$

Proof. We take $M = A_\delta(x)$ and $M = \bar{A}_\delta(x)$ in remark 3.13. Notice

$$\delta^3\langle\partial_b\sigma(x),\xi\rangle^2 \leq \delta^3 C\lambda_*(A(x))|\xi|^2 \leq C(\delta\langle\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2)$$

We have

$$\begin{aligned}\delta\langle\sigma(x)+\delta\partial_b\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2 &\leq 2\delta\langle\sigma(x),\xi\rangle^2 + 2\delta^3\langle\partial_b\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2\\ &\leq C(\delta\langle\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2),\end{aligned}$$

and so $|\xi|^2_{A_\delta} \leq C|\xi|^2_{\bar{A}_\delta(x)}$. Analogously, since

$$\delta\langle\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2 \leq C(\langle\delta\sigma(x)+\delta\partial_b\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2),$$

we have $|\xi|^2_{\bar{A}_\delta(x)} \leq C|\xi|^2_{A_\delta(x)}$. From

$$|\sigma(\hat{x})-\sigma(x)| = |\sigma(x+b(x)\delta)-\sigma(x)| \leq \int_0^\delta |\sigma'(x+b(x)t)b(x)|dt \leq C\delta,$$

applying again Remark 3.13 as in the previous point, also (3.4.5) follows. □

We prove now some estimates that will be needed in the concatenation along the tube. The following lemma establish the equivalence of matrix norms of this kind when the matrix is taken in two points that are close in such matrix norms, uniformly in δ.

Lemma 3.15. *Consider two points $x, y \in \mathbb{R}^2$. It exist $\rho, \delta^* \in 1/\mathcal{C}$ such that: If $|x-y|_{A_\delta(x)} \leq \rho$, for any $\delta \leq \delta^*$, for any $\xi \in \mathbb{R}^2$,*

$$\frac{1}{4}|\xi|_{A_\delta(x)} \leq |\xi|_{A_\delta(y)} \leq 4|\xi|_{A_\delta(x)}.$$

Proof. We write C for a constant in $\mathcal{C}$ that may vary from line to line. Remark that (3.4.2) implies

$$|x-y| \leq \delta^{1/2}C|x-y|_{A_\delta(x)} \leq \rho C\delta^{1/2} \leq 1.$$

From

$$\sigma(x) = \sigma(y) + \int_0^1 \sigma'(x)(x-y)ds + \int_0^1 (\sigma'(y+(x-y)s)-\sigma'(x))(x-y)ds$$

follows

$$\begin{aligned}\langle\sigma(x),\xi\rangle^2 \leq\ & 4\langle\sigma(y),\xi\rangle^2 + 4\langle\sigma'(x)(x-y),\xi\rangle^2\\ &+ 4\langle\int_0^1 (\sigma'(y+(x-y)s)-\sigma'(x))(x-y)ds,\xi\rangle^2\end{aligned}$$

Since $A_\delta(x)$ is invertible,

$$\sigma'(x)(x-y) = \sigma'(x)A_\delta(x)A_\delta^{-1}(x)(x-y).$$

From Cauchy-Schwartz inequality and $|A_\delta^{-1}(x)(x-y)| \le \rho$

$$\begin{aligned}\langle \sigma'(x)(x-y), \xi\rangle &= \langle A_\delta^{-1}(x)(x-y), (\sigma'(x)A_\delta(x))^T\xi\rangle \\ &\le \rho|(\sigma'(x)A_\delta(x))^T\xi|.\end{aligned}$$

Recalling **H3**

$$\begin{aligned}\sigma'(x)A_\delta(x) &= \sigma'(x)(\delta^{1/2}\sigma(x), \delta^{3/2}[\sigma,b](x)) \\ &= (\delta^{1/2}\kappa_\sigma(x)\sigma(x), \delta^{3/2}\partial_{[\sigma,b]}\sigma(x))\end{aligned}$$

so

$$|(\sigma'(x)A_\delta(x))^T\xi|^2 = \delta\kappa_\sigma^2(x)\langle\sigma(x),\xi\rangle^2 + \delta^3\langle\partial_{[\sigma,b]}\sigma(x),\xi\rangle^2$$

and therefore

$$\begin{aligned}\langle \sigma'(x)(x-y), \xi\rangle &\le \rho(\delta\kappa_\sigma^2(x)\langle\sigma(x),\xi\rangle^2 + \delta^3\langle\partial_{[\sigma,b]}\sigma(x),\xi\rangle^2) \\ &\le C\rho\delta\langle\sigma(x),\xi\rangle^2 + C\rho\delta^3|\xi|^2\end{aligned}$$

Now

$$\begin{aligned}\langle\int_0^1 (\sigma'(y+(x-y)s) - \sigma'(x))(x-y)ds, \xi\rangle^2 &\le \int_0^1 |n(x)||y-x|^2(1-s)ds|^2|\xi|^2 \\ &\le C\rho^4\delta^2|\xi|^2.\end{aligned}$$

So

$$\langle\sigma(x),\xi\rangle^2 \le 4\langle\sigma(y),\xi\rangle^2 + C\rho\delta\langle\sigma(x),\xi\rangle^2 + C\rho\delta^3|\xi|^2 + C\rho^4\delta^2|\xi|^2.$$

Taking δ, ρ small enough in $1/\mathcal{C}$, this implies

$$\langle\sigma(x),\xi\rangle^2 \le 4\langle\sigma(y),\xi\rangle^2 + \rho\delta^2|\xi|^2.$$

In the direction $[\sigma, b]$:

$$\begin{aligned}\langle[\sigma,b](x),\xi\rangle^2 &= \langle[\sigma,b](y) + \int_0^1 [\sigma,b]'(y+(x-y)s)(x-y)ds, \xi\rangle^2 \\ &\le 2\langle[\sigma,b](y),\xi\rangle^2 + C|x-y|^2|\xi|^2 \\ &\le 2\langle[\sigma,b](y),\xi\rangle^2 + C\delta\rho^2|\xi|^2.\end{aligned}$$

We can conclude that

$$\delta\langle\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2 \le 4\delta\langle\sigma(y),\xi\rangle^2 + 2\delta^3\langle[\sigma,b](y),\xi\rangle^2 + C\rho\delta^3|\xi|^2.$$

So taking ρ small enough in $\mathcal{C}$, we have

$$\delta\langle\sigma(x),\xi\rangle^2 + \delta^3\langle[\sigma,b](x),\xi\rangle^2 \le 16\delta\langle\sigma(y),\xi\rangle^2 + 16\delta^3\langle[\sigma,b](y),\xi\rangle^2.$$

We have shown $|\xi|_{A_\delta(x)} \le 4|\xi|_{A_\delta(y)}$. The converse inequality follows from an analogous reasoning.

□

We prove now that moving along a control $\phi \in L^2[0,T]$ for a small time, the trajectory remains close to the initial point in the A_δ-norm. Define, for fixed δ,

$$\varepsilon = \left(\int_0^\delta |\phi_s|^2 ds \right)^{1/2}.$$

For

$$x_t(\phi) = x_0 + \int_0^t \sigma(x_s(\phi))\phi_s ds + \int_0^t b(x_s(\phi)) ds,$$

we have:

Lemma 3.16. *There exist $\delta_*, \varepsilon_* \in 1/\mathcal{C}$ such that for $\delta \le \delta_*$, $\varepsilon \le \varepsilon_*$*

$$|x_\delta(\phi) - (x_0 + b(x_0)\delta)|_{\bar{A}_\delta(x_0)} \le C(\varepsilon \vee \delta^{1/2}).$$

Proof. Via computations analogous to Decomposition 3.3.3 it is possible to write

$$x_\delta(\phi) - (x_0 + b(x_0)\delta) = \bar{A}_\delta(G_\phi + \tilde{R}_{\phi,\delta}) \tag{3.4.6}$$

where

$$G_\phi = \Theta_\phi + \tilde{\eta}(\Theta_\phi), \quad \Theta_\phi = \begin{pmatrix} \delta^{-1/2} \int_0^\delta \phi_s ds \\ \delta^{-3/2} \int_0^\delta (\delta - s)\phi_s ds \end{pmatrix}$$

and

$$|\tilde{R}_{\phi,\delta}| \le C\delta^{1/2}.$$

Remark that, by Hölder inequality,

$$|\delta^{-1/2} \int_0^\delta \phi_s ds| \le \varepsilon, \quad |\delta^{-3/2} \int_0^\delta (\delta - s)\phi_s ds| \le \varepsilon$$

so $|\Theta_\phi| \le 2\varepsilon$ and by (3.3.4) $|\tilde{\eta}(\Theta_\phi)| \le 4\varepsilon^2$. Therefore $|G_\phi| \le 4\varepsilon$ and

$$|\bar{A}_\delta^{-1}(x_\delta(\phi) - (x_0 + b(x_0)\delta))| = |G_\phi + \tilde{R}_{\phi,\delta}| \le C(\varepsilon \vee \delta^{1/2}).$$

□

Lemma 3.17. *There exist $\delta_*, \varepsilon_* \in 1/\mathcal{C}$, $C \in \mathcal{C}$ such that for $\delta \le \delta_*$, $\varepsilon \le \varepsilon_*$*

$$\frac{1}{C}|\xi|_{A_\delta(x_0)} \le |\xi|_{A_\delta(x_\delta)} \le C|\xi|_{A_\delta(x_0)}$$

Proof. Applying in this order (3.4.5), Lemma (3.4.4), Lemma 3.16 we obtain

$$|x_\delta - \hat{x}|_{A_\delta(\hat{x})} \le C|x_\delta - (x_0 + b(x_0)\delta)|_{A_\delta(x_0)} \le C|x_\delta - (x_0 + b(x_0)\delta)|_{\bar{A}_\delta(x_0)} \le C(\varepsilon \vee \delta^{1/2}).$$

Now, choosing δ_*, ε_* small enough, we can apply Lemma 3.15 to the points $x_\delta, \hat{x}$, and

$$\frac{1}{4}|\xi|_{A_\delta(\hat{x})} \le |\xi|_{A_\delta(x_\delta)} \le 4|\xi|_{A_\delta(\hat{x})}.$$

Now again (3.4.5) concludes the proof. □

3.4.2 The control metric

Here we write $A_R(\cdot)$ instead of A_R because we need to consider this matrix on different points. A natural way to associate a quasi-distance to the matrix norm $|\cdot|_{A_R(\cdot)}$ used in this chapter is to define

$$d(x,y) < \sqrt{R} \Leftrightarrow |x-y|_{A_R(x)} < 1.$$

d is a quasi-distance on $\Omega = \{x \in \mathbb{R}^2 : \det A(x) \neq 0\}$, verifying the following three properties (see [78]):

i) for every $r > 0$, the set $\{y \in \Omega \,:\, d(x,y) < r\}$ is open;

ii) $d(x,y) = 0$ if and only if $x = y$;

iii) for every compact set $K \Subset \Omega$ there exists $C > 0$ such that $d(x,y) \leq C\big(d(x,z) + d(z,y)\big)$ holds for every $x, y, z \in K$.

We say that two quasi-distances $d_1 : \Omega \times \Omega \to \mathbb{R}^+$ and $d_2 : \Omega \times \Omega \to \mathbb{R}^+$ are equivalent if for every compact set $K \Subset \Omega$ there exists a constant C such that for every $x, y \in K$

$$\frac{1}{C} d_1(x,y) \leq d_2(x,y) \leq C d_1(x,y). \tag{3.4.7}$$

In particular if d_1 is a distance and d_2 is equivalent with d_1 then d_2 is a quasi-distance. d_1 and d_2 are locally equivalent if for every $x_0 \in \Omega$ there exist a neighborhood V of x_0 and a constant C such that (3.4.7) holds for every $x, y \in V$.

On the other hand, the distance usually considered in the framework of stochastic differential equations is the *control distance* defined as follows: denote

$$C(x,y) = \{\phi \in L^2(0,1) : dv_s = \sigma(v_s)\phi_s ds,\ x = v_0,\ y = v_1\}. \tag{3.4.8}$$

The control distance d_c between x and y is

$$d_c(x,y) = \inf \left\{ \left(\int_0^1 |\phi_s|^2 ds \right)^{1/2} : \phi \in C(x,y) \right\}.$$

Geometrically speaking, this corresponds to take the geodesic (i.e. the shortest distance curve) joining x and y on the sub-Riemannian manifold associated with the diffusion coefficient σ. In our case this notion looks inadequate: we are supposing just a weak Hörmander condition, and this means that we have to use the drift coefficient b to generate the whole space $\mathbb{R}^2$. Therefore any reasonable associated norm should incorporate b as well. Moreover it should account of the different speed associated to the vector field given by $[\sigma, b]$. This is the reason for the following

Definition 3.18. We first introduce a norm for the control which accounts of the scales associated to the different directions. For $\phi = (\phi_s^1, \phi_s^2) \in L^2((0,1), \mathbb{R}^2)$, we define the norm

$$\|\phi\|_{(1,3)} = \left\| (|\phi_s^1|, |\phi_s^2|^{1/3}) \right\|_{L^2(0,1)}.$$

Now we generalize (3.4.8) to

$$C_A(x,y) = \{\phi \in L^2((0,1),\mathbb{R}^2) : dv_s = A(v_s)\phi_s ds,\ x = v_0,\ y = v_1\}. \tag{3.4.9}$$

This is non-empty if $A = (\sigma, [\sigma, b])$ is invertible. We introduce

$$d_c(x,y) = \inf \left\{ \|\phi\|_{(1,3)} : \phi \in C_A(x,y) \right\}.$$

We are interested in establishing an equivalence between d, the semi-distance coming from the matrix-norm, and d_c, the distance in term of the control.

Lemma 3.19. *d and d_c are locally equivalent.*

Proof. We use in this proof some notions on similar metrics and pseudo-metrics for which we refer to [78]. Define

$$\rho(x,y) = \inf\{\delta > 0 | \exists \phi \in C_A(x,y),\ |\phi_s^1| < \delta,\ |\phi_s^2| < \delta^3\}.$$

It is also possible to allow only constant linear combinations of the vector fields:

$$\bar{C}_A(x,y) = \{\theta \in \mathbb{R}^2 : dv_s = A(v_s)\theta ds,\ x = v_0,\ y = v_1\}, \tag{3.4.10}$$

and define

$$\rho_2(x,y) = \inf\{\delta > 0 | \exists \theta \in \bar{C}_A(x,y),\ |\theta^1| < \delta,\ |\theta^2| < \delta^3\}.$$

In [78] the pseudo-distances ρ and ρ_2 are proved equivalent. We use here only the trivial inequality $\rho \leq \rho_2$. Remark that the difference between ρ and d_c is that we take $\|\phi\|_\infty = \sup_{0\leq s\leq 1} |\phi_s|$ instead of $\|\phi\|_2 = \left(\int_0^1 |\phi_s|^2 ds\right)^{1/2}$. So $d_c \leq \rho$ follows easily from the fact that the $L^2(0,1)$ norm is dominated by the $L_\infty(0,1)$ norm.

For fixed x, we consider a compact K containing x, and define

$$\mathcal{C}_K = \left\{ C = \sup_{x\in K} L\left(n(x)/\lambda(x)\right)^q, \exists L, q \geq 1 \text{ universal constants} \right\},$$

$1/\mathcal{C}_K = \{\delta : 1/\delta \in \mathcal{C}_K\}$. We prove that

$$d(x,y) < \sqrt{R} \Rightarrow \rho_2(x,y) < C\sqrt{R},$$

for $R \leq R_* \in 1/\mathcal{C}_K$, $C \in \mathcal{C}_K$. By definition, $d(x,y) < \sqrt{R}$ means $|x-y|_{A_R(x)} < 1$. We prove that this implies the existence of $\theta \in \bar{C}_A(x,y)$ with $|\theta^1| < CR^{1/2}$, $|\theta^2| < CR^{3/2}$. Indeed, consider the function

$$\Phi(\theta) = \int_0^1 A(v_s(\theta))\theta ds,$$

with v satisfying $dv_s = A(v_s)\theta ds$, $v_0(\theta) = x$. Remark that $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$, $\Phi(0) = 0$ and $\Phi'(0) = A(x)$, which is non-degenerate because of **H1**. Therefore it is locally invertible. Recall (from (3.4.2))

$$|x-y|_{A_R(x)} < 1 \Rightarrow |x-y| < C\sqrt{R} \leq C\sqrt{R_*},$$

$C \in \mathcal{C}_A$. For $R_* \in 1/\mathcal{C}$ small enough, we have that it exists $\theta \in \bar{C}_A(x, y)$. We now show

$$|\theta^1| < CR^{1/2}, \ |\theta^2| < CR^{3/2}.$$

It is clear that $|\theta| < CR^{\frac{1}{2}}$. Now,

$$\begin{aligned} A_R(x)^{-1}(x-y) &= A_R(x)^{-1} \int_0^1 A(v_s(\theta))\theta ds \\ &= \left(\frac{\theta^1}{R^{1/2}}, \frac{\theta^2}{R^{3/2}} \right) + J(\theta, R) \end{aligned}$$

with

$$J(\theta, R) = A_R(x)^{-1} \int_0^1 (A(v_s(\theta)) - A(x))\theta ds.$$

Using as usual **H3** and development (3.4.6), it is possible to prove $|(J(\theta, R))_2| \leq C|\theta||\theta^2|R^{-3/2}$. So, supposing $R_* \in 1/\mathcal{C}_K$, we have that $|\theta^2| < CR^{-3/2}$.

It also holds

$$d_c(x, y) < \frac{\sqrt{R}}{C^*} \Rightarrow d(x, y) < \sqrt{R},$$

if C^* is large enough constant. With the same notation as above, in particular supposing $\phi \in C_A(x, y)$ with $\|\phi\|_{1,3} \leq \frac{\sqrt{R}}{K}$, and applying **H3**,

$$\begin{aligned} |x-y|_{A_R(x)} &= \left| A_R(x)^{-1} \int_0^1 A(v_s)\phi_s ds \right| \\ &\leq C \sqrt{\frac{\left(\int_0^1 \phi_s^1 ds\right)^2}{R} + \frac{\left(\int_0^1 \phi_s^2 ds\right)^2}{R^3}} \\ &\leq C/C^* < 1 \end{aligned}$$

This concludes the local equivalence of d, d_c, ρ, ρ_2. □

3.5 Density estimates for a chain of stochastic differential equations

3.5.1 Setting, notations and results

In this last section we consider a different model. We work with the system of stochastic differential equations considered in [42] (related models are studied in [77], [59]). We apply the techniques we introduced in chapter 2, and find a local Gaussian density estimate in short time coherent with the result of [42]. Differently from [42], here coefficients do not depend on time. We work under local non degeneracy, finding local estimates, whereas in the original work hypothesis and results are global. We take a Brownian Motion in $W \in \mathbb{R}^d$, and a chain of n differential equations in dimension d:

$$\begin{aligned} dX^1_t &= B_1(X^1_t, \ldots, X^n_t)dt + \sigma(X^1_t, \ldots, X^n_t) \circ dW_t \\ dX^2_t &= B_2(X^1_t, \ldots, X^n_t)dt \\ dX^3_t &= B_3(X^2_t, \ldots, X^n_t)dt \\ &\cdots \\ dX^n_t &= B_n(X^{n-1}_t, X^n_t)dt \end{aligned} \tag{3.5.1}$$

each X^i_t being $\mathbb{R}^d$ valued as well. We write the equation in Stratonovic form, whereas in [42] the stochastic integral is written in Ito's form. The equivalence between the two is clear, since the correction term we need to add when converting the integrals is non-zero only in the first d components. Therefore, the structure of the differential equation does not change. (3.5.1) corresponds to a diffusion $X \in \mathbb{R}^{nd}$,

$$X_0 = x, \quad dX_t = \bar{\sigma}(X_t) \circ dW_t + B(X_t)dt$$

where the coefficients have the specific form

$$B = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} \quad \text{and} \quad \bar{\sigma} = \begin{pmatrix} \sigma \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

and for $i > 1$ $B_i(X)$ depends just on $X_{i-1}, \ldots, X_n$. We denote with σ^i and $\bar{\sigma}^i$ the $i-th$ columns of σ and $\bar{\sigma}$. We take $B, \sigma \in \mathcal{C}^\infty$ and suppose

(H_1) $\lambda_* \big(\sigma(x)\, D_{x_1}B_2(x) \ldots D_{x_{n-1}}B_n(x)\big) \geq \lambda(x) > 0$,

(H_2) $\forall k \in \mathbb{N}$, $x \in \mathbb{R}^{nd}$, $\quad |\nabla^k B(x)| + |\nabla^k \sigma(x)| \leq C < \infty$.

(H_1) implies that the diffusion satisfies the weak Hörmander condition at x, so X_t admits a density. This is explained in detail in [42].

We introduce some notations. For $m \in \mathbb{N}$, let $\mathcal{M}(m)$ be the set of all $m \times m$ matrices on $\mathbb{R}$.

- For fixed $\delta > 0$, $\mathbb{T}_\delta \in \mathcal{M}(nd)$ is a diagonal matrix given by n diagonal blocks in $\mathcal{M}(d)$, with $\delta^i Id_d$ as i^{th} diagonal block, where $Id_d \in \mathcal{M}(d)$ is the identity matrix.

- $A \in \mathcal{M}(nd)$ is a block-diagonal matrix given by n blocks in $\mathcal{M}(d)$, with the product $D_{x_{i-1}}B_i \ldots D_{x_1}B_2\sigma(x)$ as i^{th} diagonal block:

$$A = \begin{pmatrix} \sigma(x) & 0 & \ddots & 0 \\ 0 & D_{x_1}B_2\sigma(x) & 0 & \ddots \\ \ddots & 0 & \ddots & 0 \\ 0 & \ddots & 0 & D_{x_{n-1}}B_n \ldots D_{x_1}B_2\sigma(x) \end{pmatrix}. \tag{3.5.2}$$

This matrix is invertible because of (H_1). For fixed δ we also define

$$A_\delta = \frac{A\mathbb{T}_\delta}{\sqrt{\delta}} = \frac{\mathbb{T}_\delta A}{\sqrt{\delta}}. \tag{3.5.3}$$

We do not write the dependence on x, since we are working with the diffusion in short time and the initial condition is fixed.

- Q is a symmetric positive definite block matrix in $\mathcal{M}(nd)$, given by n^2 diagonal blocks: for $1 \le i, j \le n$, the block in position (i, j) is

$$\frac{Id_d}{(i+j-1)(i-1)!(j-1)!} \tag{3.5.4}$$

(see remark 3.24). We also define $M_\delta = \mathbb{T}_\delta A\sqrt{Q/\delta} = A_\delta\sqrt{Q}$, and recall that for any M square, invertible matrix $|\xi|_M = |M^{-1}\xi|$.

- We also introduce the following ODE

$$\bar{X}_0 = x, \quad d\bar{X}_t = B(\bar{X}_t)dt,$$

and its solution at time δ, $\bar{X}_\delta$.

In [42] the following Gaussian two sided bound is proved, using the *parametrix method*, under some regularity of the coefficients and a version of (H_1) uniform in space.

Theorem 3.20. *Let X_t be the solution of* (3.5.1) *with initial condition x, with final time horizon T. X_t admits a density p_{X_δ} in y, and there exists a constant C_T such that, for any $\delta \in (0, T]$,*

$$\begin{aligned} &C_T^{-1}\delta^{-n^2d/2}\exp\left(-C_T\delta|y-\bar{X}_\delta|^2_{\mathbb{T}_\delta}\right) \\ &\quad\le p_{X_\delta}(y) \le C_T\delta^{-n^2d/2}\exp\left(-C_T^{-1}\delta|y-\bar{X}_\delta|^2_{\mathbb{T}_\delta}\right). \end{aligned} \tag{3.5.5}$$

We prove here, with the *Malliavin calculus* techniques of chapter 2, the following result:

Theorem 3.21. *There exists $C > 0$ such that for any $\varepsilon > 0$, $r > 0$, for $\delta \le \frac{\varepsilon^2\exp(-r^2)}{C}$, for y such that $|y-\bar{X}_\delta|_{M_\delta} \le r$*

$$\begin{aligned} &\frac{1-\varepsilon}{c\delta^{n^2d/2}}\exp\left(-\frac{|y-\bar{X}_\delta|^2_{M_\delta}}{2}\right) \\ &\quad\le p_{X_\delta}(y) \le \frac{1+\varepsilon}{c\delta^{n^2d/2}}\exp\left(-\frac{|y-\bar{X}_\delta|^2_{M_\delta}}{2}\right) \end{aligned} \tag{3.5.6}$$

where $c = \left(\prod_{i=1}^{n-1} i! \Big/ \sqrt{\prod_{i=1}^{2n-1} i!}\right)^d (2\pi)^{nd/2}\det A.$

Remark 3.22. We first notice that $\exists C > 0$ such that $C^{-1}\sqrt{\delta}|\xi|_{\mathbb{T}_\delta} \leq |\xi|_{M_\delta} \leq C\sqrt{\delta}|\xi|_{\mathbb{T}_\delta}$. Therefore, our result implies

$$\begin{aligned} &C^{-1}\delta^{-n^2d/2}\exp\left(-C\delta|y-\bar{X}_\delta|^2_{\mathbb{T}_\delta}\right) \\ &\quad\leq p_{X_\delta}(y) \leq C\delta^{-n^2d/2}\exp\left(-C^{-1}\delta|y-\bar{X}_\delta|^2_{\mathbb{T}_\delta}\right), \end{aligned} \tag{3.5.7}$$

which is analogous to (3.5.5), but weaker in the sense that it holds just locally around $\bar{X}_\delta$ and for small δ. On the other hand, our result holds under local hypoellipticity, whereas in [42] global hypoellipticity is required. Moreover, in (3.5.6) there is no constant in the exponential, and ε can be taken abitrarily small (but also δ must be taken depending on ε), so our estimate is more significant in small time.

We present in the following sections the proof of this result, which is based on the Malliavin calculus techniques presented in chapter 2. In section 3.5.3 we consider a related control distance.

3.5.2 Development

Remark 3.23. [Lie Brackets] Take $F = (F_1, \dots F_n)$, $G = (G_1, \dots G_n)$ functions in $\mathbb{R}^{nd}$, each F_i, G_i being in $\mathbb{R}^d$. We consider the Lie Brackets $[G,F] = (\nabla F)G - (\nabla G)F$. Supposing $F_{k+1} = F_{k+2} = \dots = F_n = 0$ and $G_{l+1} = G_{l+2} = \dots = G_l = 0$, we have $[G,F]_i = 0$ for $i > \max(k,l)$. We have

$$\nabla B = \begin{pmatrix} D_{x_1}B_1 & \times & \dots & & & \\ D_{x_1}B_2 & D_{x_2}B_2 & \times & \dots & & \\ 0 & D_{x_2}B_3 & \ddots & & & \\ \vdots & 0 & \ddots & \times & \vdots & \vdots \\ & & \ddots & D_{x_{n-2}}B_{n-1} & D_{x_{n-1}}B_{n-1} & \times \\ 0 & & \dots & 0 & D_{x_{n-1}}B_n & D_{x_n}B_n \end{pmatrix}$$

In particular for $k < n$, $[F,B]_{k+1} = (D_{x_k}B_{k+1})F_k$, and $[F,B]_i = 0$ for $i > k+1$.

Define the following r.v. in $\mathbb{R}^d$, for $k = 1, \dots, n$:

$$J_t^k = \int_0^t \dots \int_0^{s_{k-2}} W_{s_{k-1}} ds_{k-1} \dots ds_1 = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW_s.$$

With a first order stochastic Taylor development we obtain:

$$X_t - \bar{X}_t = \bar{\sigma}(x)J_t^1 + \int_0^t \bar{\sigma}(X_s) - \bar{\sigma}(x) \circ dW_s + \int_0^t B(X_s) - B(\bar{X}_s)ds.$$

So in the first d components (i.e., in the space where the first stochastic differential equation lives) we have

$$X_t - \bar{X}_t = \bar{\sigma}(x)J_t^1 + \begin{pmatrix} L_t^1 \\ \times \\ \vdots \\ \times \end{pmatrix} = \begin{pmatrix} \sigma(x)J_t^1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} L_t^1 \\ \times \\ \vdots \\ \times \end{pmatrix}.$$

where $L_t^1 \in \mathbb{R}^d$ is of order t, meaning $\mathbb{E}|L_t^1|^q \le C_q t^q$ for any q, $t \le t_0$. When we push further the development find

$$
\begin{aligned}
X_t - \bar{X}_t &= \bar{\sigma}(x) J_t^1 + (\nabla B)\bar{\sigma}(x) J_t^2 \\
&\quad + \int_0^t \int_0^s (\nabla \bar{\sigma})\bar{\sigma}(X_u) \circ dW_u \circ dW_s + \int_0^t \int_0^s (\nabla \bar{\sigma}) B(X_u) du \circ dW_s \\
&\quad + \int_0^t \int_0^s (\nabla B)\bar{\sigma}(X_u) - (\nabla B)\bar{\sigma}(x) \circ dW_u ds \\
&\quad + \int_0^t \int_0^s (\nabla B) B(X_u) - (\nabla B) B(\bar{X}_u) du ds.
\end{aligned}
$$

From Remark 3.23

$$
(\nabla B)\bar{\sigma}(x) = \begin{pmatrix} \times \\ D_{x_1} B_2 \sigma(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

and, denoting with $L_t^2 \in \mathbb{R}^d$ the vector of the components from $d+1$ to $2d$ of the integrals above, for any $q > 0$ it holds $\mathbb{E}|L_t^2|^q \le C_q t^{2q}$. Therefore

$$
X_t - \bar{X}_t = \begin{pmatrix} \sigma(x) J_1 \\ D_{x_1} B_2 \sigma(x) J_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} L_t^1 \\ L_t^2 \\ \times \\ \vdots \\ \times \end{pmatrix}
$$

where for any $q > 0$ $\mathbb{E}|L_t^1|^q \le C_q t^q$ and $\mathbb{E}|L_t^2|^q \le C_q t^{2q}$. Iterating this procedure n times we find

$$
X_t - \bar{X}_t = \begin{pmatrix} \sigma(x) J_t^1 & + & L_t^1 \\ D_{x_1} B_2 \sigma(x) J_t^2 & + & L_t^2 \\ \vdots & + & \ddots \\ D_{x_{n-1}} B_n \dots D_{x_1} B_2 \sigma(x) J_t^n & + & L_t^n \end{pmatrix}
$$

where $E|L_t^i|^q \le C_q t^{iq}$. Clearly

$$
F := A_\delta^{-1}(X_\delta - \bar{X}_\delta) = \Theta + A_\delta^{-1} L_\delta, \tag{3.5.8}
$$

where

$$
\Theta = \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_n \end{pmatrix}, \qquad \Theta_k = \frac{J_\delta^k}{\delta^{k-1/2}} \in \mathbb{R}^d.
$$

Θ is a non-degenerate Gaussian since its covariance can be expressed as a block matrix as

$$Q = Cov(\Theta) = \left(\frac{Id_d}{(i+j-1)(i-1)!(j-1)!} \right)_{1 \le i,j \le n},$$

and $\det(Q) > 0$ (see (3.5.11)).

Remark 3.24. In linear algebra a *Hilbert matrix* is a square matrix with entries given by $H_{i,j} = \frac{1}{i+j-1}$, $i, j = 1, \dots n$. For this matrix an explicit expression for the determinant is known:

$$\det \left(\frac{1}{i+j-1} \right)_{1 \le i,j \le n} = \frac{(\prod_{i=1}^{n-1} i!)^4}{\prod_{i=1}^{2n-1} i!}. \tag{3.5.9}$$

In our setting it is more convenient to define H as a block matrix in $\mathcal{M}(nd)$, given by n^2 diagonal $d \times d$ blocks. For $1 \le i, j \le n$, the block in position i, j is $\frac{Id_d}{(i+j-1)}$. Using (3.5.9), we can see that

$$\det H = \left(\frac{(\prod_{i=1}^{n-1} i!)^4}{\prod_{i=1}^{2n-1} i!} \right)^d. \tag{3.5.10}$$

We also set U as a diagonal matrix in $\mathcal{M}(dn)$ where on the diagonal we have n blocks given by $\frac{Id_d}{(i-1)!}$, $i = 1, \dots n$. Clearly $\det U = (\prod_{i=1}^{n-1} i!)^{-d}$. For Q defined in (3.5.4) holds $Q = UHU$, so from (3.5.10)

$$\det(Q) = \det(H)\det(Q)^2 = \left(\frac{(\prod_{i=1}^{n-1} i!)^2}{\prod_{i=1}^{2n-1} i!} \right)^d \neq 0 \tag{3.5.11}$$

3.5.3 Density estimate

Lemma 3.25. *For $y \in \mathbb{R}^{nd}$, for $\delta \le 1$,*

$$\begin{aligned} &\frac{1}{(2\pi)^{nd/2} \det M_\delta} \left(\exp\left(-\frac{|y - \bar{X}_\delta|^2_{M_\delta}}{2} \right) - C\sqrt{\delta} \right) \\ &\le p_{X_\delta}(y) \\ &\le \frac{1}{(2\pi)^{nd/2} \det M_\delta} \left(\exp\left(-\frac{|y - \bar{X}_\delta|^2_{M_\delta}}{2} \right) + C\sqrt{\delta} \right) \end{aligned} \tag{3.5.12}$$

Remark 3.26. Theorem 3.21 follows directly from this lemma. Remark, from 3.5.11,

$$(2\pi)^{nd/2} \det M_\delta = \det A \left(\prod_{i=1}^{n-1} i! \Big/ \sqrt{\prod_{i=1}^{2n-1} i!} \right)^d (2\pi)^{nd/2} \, \delta^{n^2 d/2}.$$

Proof. We start using (3.5.8) to give an estimate for the density of F. We apply theorem 2.4 with $U = 1$, $F = F$ and $G = \Theta$ and $R = A_\delta^{-1} L_\delta$. We have

$$p_\Theta(y) - C\Gamma_\Theta(32nd)\|A_\delta^{-1}L_\delta\|_{2,32nd} \le p_F(y) \le p_\Theta(y) + C\Gamma_F(32nd)\|A_\delta^{-1}L_\delta\|_{2,32nd}.$$

Standard computations give $\|A_\delta^{-1}L_\delta\|_{2,32nd} \le C_q\delta^{1/2}$. Θ is a non degenerate Gaussian, so $\Gamma_\Theta(32nd) < \infty$. We prove non-degeneracy of the Malliavin covariance of F in lemma 3.27. More precisely, we prove that there exists $\delta_* > 0$ such that, for any $\delta \le \delta_*$, $\Gamma_F(32nd) < C < \infty$. We obtain that for any $\delta \le \delta_*$.

$$p_\Theta(y) - C\sqrt{\delta} \le p_F(y) \le p_\Theta(y) + C\sqrt{\delta}.$$

Since $\Theta \sim N(0, Q)$, multiplying by $Q^{-1/2}$ we have $Q^{-1/2}F = M_\delta^{-1}(X_\delta - \bar{X}_\delta)$ and

$$\begin{aligned} &\frac{1}{(2\pi)^{dn/2}} \exp\left(-\frac{|y|^2}{2}\right) - C\sqrt{\delta} \\ &\le p_{M_\delta^{-1}(X_\delta - \bar{X}_\delta)}(y) \le \frac{1}{(2\pi)^{dn/2}} \exp\left(-\frac{|y|^2}{2}\right) + C\sqrt{\delta}. \end{aligned} \tag{3.5.13}$$

(3.5.12) follows from the change of variable $y \to M_\delta y$. □

Lemma 3.27. *There exists $C > 0$ such that, for any $\delta \le 1$, $\Gamma_F(32nd) < C < \infty$.*

Proof. We actually prove $\Gamma_F(p) < C < \infty$ for any $p > 1$. The proof is analogous to the proof of lemma 3.8. Recall that $\Gamma_F(p) = 1 + (\mathbb{E}|\lambda_*(\gamma_F(\gamma))|^{-p})^{1/p}$.

Following [79] we define the tangent flow of X as the derivative with respect to the initial condition of X, $Y_t := \partial_x X_t$. We also denote its inverse $Z_t = Y_t^{-1}$. They satisfy the following stochastic differential equations in Stratonovic form:

$$\begin{aligned} Y_t &= Id + \sum_k \int_0^t \nabla\bar{\sigma}^k(X_s)Y_s \circ dW_s^k + \int_0^t \nabla B(X_s)Y_s ds \\ Z_t &= Id - \sum_k \int_0^t Z_s\nabla\bar{\sigma}^k(X_s) \circ dW_s^k - \int_0^t Z_s\nabla B(X_s)ds \end{aligned} \tag{3.5.14}$$

Then

$$D_s X_t = Y_t Z_s \bar{\sigma}(X_s),$$

and, for $\phi \in C^2$, applying Ito's formula (see([79])),

$$Z_t\phi(X_t) = \phi(x) + \int_0^t Z_s \sum_{k=1}^d [\bar{\sigma}^k, \phi](X_s)dW_s^k + \int_0^t Z_s \left\{[B, \phi] + \frac{1}{2}\sum_{k=1}^d [\bar{\sigma}^k, [\bar{\sigma}^k, \phi]]\right\}(X_s)ds. \tag{3.5.15}$$

We compute the derivative of F:

$$D_s F = D_s A_\delta^{-1}(X_\delta - \bar{X}_\delta) = A_\delta^{-1} Y_\delta Z_s \bar{\sigma}(X_s). \tag{3.5.16}$$

Now, multipling by $Id_{nd} = A_\delta A_\delta^{-1}$, we write

$$\gamma_F = \langle DF, DF\rangle = A_\delta^{-1} Y_\delta A_\delta \bar{\gamma}_\delta A_\delta^T Y_\delta^T A_\delta^{-1,T}, \tag{3.5.17}$$

with

$$\bar{\gamma}_\delta = \int_0^\delta A_\delta^{-1} Z_s \bar{\sigma}(X_s)\bar{\sigma}(X_s)^T Z_s^T A_\delta^{-1,T} ds. \tag{3.5.18}$$

Remark that

$$\gamma_F^{-1} = A_\delta^T Z_\delta^T A_\delta^{-1,T} \bar{\gamma}_\delta^{-1} A_\delta^{-1} Z_\delta A_\delta. \tag{3.5.19}$$

We now deal with $\bar{\gamma}_\delta^{-1}$. We apply formula (3.5.15) to σ^i, $i = 1 \dots d$, and then to the Lie Brackets appearing in the development, iterating until the order of the remainder is small enough to find

$$Z_t \sigma^i(X_t) = \begin{pmatrix} \sigma^i(x) & + & O(t^{1/2}) \\ D_{x_1} B_2 \sigma^i(x) t & + & O(t^{3/2}) \\ D_{x_2} B_3 D_{x_1} B_2 \sigma^i(x) \frac{t^2}{2!} & + & O(t^{5/2}) \\ \vdots & + & \ddots \\ D_{x_{n-1}} B_n \dots D_{x_1} B_2 \sigma^i(x) \frac{t^{n-1}}{(n-1)!} & + & O(t^{n-1/2}) \end{pmatrix}.$$

We have denoted as $O(t^\alpha)$ the integrals we find iterating (3.5.15). They involve Z_s and σ^i, B and their Lie brackets up to a certain finite order. Writing $O(t^\alpha)$ we mean that for any $q > 0$ the q-moment of this quantity is bounded by $C_q t^\alpha$, and this follows from standard computations.

We can write the result above as a matrix product:

$$Z_t \sigma(X_t) = \begin{pmatrix} \sigma(x) & + & R_t^1 \\ D_{x_1} B_2 \sigma(x) t & + & R_t^2 \\ D_{x_2} B_3 D_{x_1} B_2 \sigma(x) \frac{t^2}{2!} & + & R_t^3 \\ \vdots & + & \ddots \\ D_{x_{n-1}} B_n \dots D_{x_1} B_2 \sigma(x) \frac{t^{n-1}}{(n-1)!} & + & R_t^n. \end{pmatrix} \tag{3.5.20}$$

The remainder R_t^i, for $i = 1, \dots, n$, is a square $d \times d$ matrix, and it is of order $t^{i-1/2}$, meaning $\mathbb{E}|R_t^i|^q \le C_q t^{q(i-1/2)}$, for any $q > 0$. From (3.5.20), using the block-diagonal structure of A_δ we have

$$A_\delta^{-1} Z_s \sigma(X_s) = \delta^{-1/2} \begin{pmatrix} Id_d \\ Id_d \frac{s}{\delta} \\ \vdots \\ \frac{Id_d}{(n-1)!} \left(\frac{s}{\delta}\right)^{n-1} \end{pmatrix} + \begin{pmatrix} \tilde{R}_t^1/(\delta^{1/2}) \\ \tilde{R}_t^2/\delta \\ \vdots \\ \tilde{R}_t^n/(\delta^{n/2}) \end{pmatrix}$$

where $\mathbb{E}|\tilde{R}_t^i|^q \le C_q t^{q(i-1/2)}$ still holds. For fixed ε we introduce the stopping time

$$S_\varepsilon = \inf\left\{ s \ge 0 : \exists i, j = 1, \dots, n \quad \int_0^s |\tilde{R}_u^i \tilde{R}_u^l| du \ge \frac{c_{i,j}(\delta\varepsilon)^{i+j-1}}{(i+j-1)(i-1)!(j-1)!} \right\} \wedge \delta, \tag{3.5.21}$$

with $c_{i,j}$ universal constants small enough to have (3.5.23). Remark

$$\begin{aligned}
\mathbb{P}[S_\varepsilon < \delta\varepsilon] &\leq \sum_{i,j} \mathbb{P}\left[\left(\int_0^{\delta\varepsilon} |R_u^i R_u^l| du\right)^q \geq c(\delta\varepsilon)^{q(i+j-1)}\right] \\
&\leq \sum_{i,j} \frac{\mathbb{E}\left(\int_0^{\delta\varepsilon} |R_u^i R_u^l| du\right)^q}{c_q(\delta\varepsilon)^{q(i+j-1)}} \\
&\leq \sum_{i,j} \frac{C_q(\delta\varepsilon)^{q(i+j)}}{c_q(\delta\varepsilon)^{q(i+j-1)}} \leq \varepsilon^q
\end{aligned} \tag{3.5.22}$$

for $\delta \leq \delta_q$, $\varepsilon \leq 1$, for any $q > 0$.

Now we suppose to be on $\frac{S_\varepsilon}{\delta} \geq \varepsilon$. Applying the inequality

$$\langle (M+R)(M+R)^T \xi, \xi \rangle \geq \frac{1}{2}\langle MM^T\xi, \xi\rangle - \langle RR^T\xi, \xi\rangle,$$

which holds for any matrix M, R, vector ξ, we obtain

$$\begin{aligned}
\bar{\gamma}_\delta &= \int_0^\delta A_\delta^{-1} Z_s \sigma(X_s)\sigma(X_s)^T Z_s^T A_\delta^{-1,T} ds \\
&\geq \int_0^{S_\varepsilon} A_\delta^{-1} Z_s \sigma(X_s)\sigma(X_s)^T Z_s^T A_\delta^{-1,T} ds \\
&\geq \frac{1}{2}\int_0^{S_\varepsilon} \delta^{-1} \begin{pmatrix} Id_d \\ Id_d \frac{s}{\delta} \\ \vdots \\ \frac{Id_d}{(n-1)!}\left(\frac{s}{\delta}\right)^{n-1} \end{pmatrix} \begin{pmatrix} Id_d \\ Id_d \frac{s}{\delta} \\ \vdots \\ \frac{Id_d}{(n-1)!}\left(\frac{s}{\delta}\right)^{n-1} \end{pmatrix}^T ds \\
&\quad - \int_0^{S_\varepsilon} \begin{pmatrix} \tilde{R}_s^1/(\delta^{1/2}) \\ \tilde{R}_s^2/(\delta^{3/2}) \\ \vdots \\ \tilde{R}_s^n/(\delta^{n-1/2}) \end{pmatrix} \begin{pmatrix} \tilde{R}_s^1/(\delta^{1/2}) \\ \tilde{R}_s^2/(\delta^{3/2}) \\ \vdots \\ \tilde{R}_s^n/(\delta^{n-1/2}) \end{pmatrix}^T ds \\
&= \left(\frac{(S_\varepsilon/\delta)^{(i+j-1)}}{(i+j-1)(i-1)!(j-1)!} Id_d - \int_0^{S_\varepsilon} \tilde{R}_s^i \tilde{R}_s^j / \delta^{(i+j-1)} ds\right)_{1\leq i,j\leq n} \\
&= \left(\frac{Id_d - \frac{(i+j-1)(i-1)!(j-1)!}{S_\varepsilon^{(i+j-1)}} \int_0^{S_\varepsilon} \tilde{R}_s^i \tilde{R}_s^j ds}{(i+j-1)(i-1)!(j-1)!} \left(\frac{S_\varepsilon}{\delta}\right)^{(i+j-1)}\right)_{1\leq i,j\leq n}
\end{aligned}$$

From (3.5.21) we have

$$\begin{aligned}
&\left|\frac{(i+j-1)(i-1)!(j-1)!}{S_\varepsilon^{(i+j-1)}}\int_0^{S_\varepsilon}\tilde{R}_s^i\tilde{R}_s^j ds\right| \\
&\quad\le \frac{(i+j-1)(i-1)!(j-1)!}{S_\varepsilon^{(i+j-1)}}\int_0^{S_\varepsilon}|\tilde{R}_s^i\tilde{R}_s^j|ds \\
&\quad\le c_{i,j}\left(\frac{\delta\varepsilon}{S_\varepsilon}\right)^{i+j-1} \\
&\quad\le c_{i,j}
\end{aligned}$$

Choosing $c_{i,j}$ small enough, independently of δ, we have:

$$\begin{aligned}
\langle\bar{\gamma}_\delta\xi,\xi\rangle &\ge \frac{1}{2}\left\langle\left(\frac{Id_d}{(i+j-1)(i-1)!(j-1)!}\left(\frac{S_\varepsilon}{\delta}\right)^{i+j-1}\right)_{1\le i,j\le n}\xi,\xi\right\rangle \\
&\ge c\left(\frac{S_\varepsilon}{\delta}\right)^{2n-1}\ge c\varepsilon^{2n-1},
\end{aligned} \tag{3.5.23}$$

since $\frac{S_\varepsilon}{\delta}\ge\varepsilon$. So $\exists\varepsilon_*$ such that taking $\varepsilon\le\varepsilon_*$, $\forall\delta\le 1$, $\forall|\xi|=1$,

$$\langle\bar{\gamma}_\delta\xi,\xi\rangle\ge\varepsilon^{2n}.$$

Using also (3.5.22) we have that for any q, for any $\varepsilon\le\varepsilon_*$, $\delta\le 1$,

$$\mathbb{P}(\langle\bar{\gamma}_\delta\xi,\xi\rangle<\varepsilon^{2n})\le\mathbb{P}[S_\varepsilon<\delta\varepsilon]\le C_q\varepsilon^q.$$

Now we apply Lemma 3.9, as in lemma 3.8 and we have $(\mathbb{E}|\lambda_*(\bar{\gamma}_\delta)|^{-p})^{1/p}\le C$.

The estimate of $A_\delta^{-1}Z_\delta A_\delta$ is standard, and from (3.5.19) we have the estimate for $(\mathbb{E}|\lambda_*(\gamma_F)|^{-p})^{1/p}\le C$. □

3.5.4 On the control distance

As in previous sections, we can define the norm $|\xi|_{A_R}=|A_R^{-1}\xi|$. It is straightforward to see that $\exists C>0$ such that $C^{-1}\sqrt{\delta}|\xi|_{\mathbb{T}_\delta}\le|\xi|_{A_\delta}\le C\sqrt{\delta}|\xi|_{\mathbb{T}_\delta}$. In what follows we establish a local equivalence between $|\cdot|_{A_R}$ and an appropriate control distance as in section 3.4.2, and so (3.5.7) could be stated in this control norm as well. We write $A_R(\cdot)$ instead of A_R because here we need to consider this matrix on different points. As before, we associate a semi-distance to the matrix norm $|\cdot|_{A_R(\cdot)}$:

$$d(x,y)<\sqrt{R}\Leftrightarrow|x-y|_{A_R(x)}<1.$$

We give the following definition in analogy to definition 3.18:

Definition 3.28. For $\phi \in L^2((0,1), \mathbb{R}^{nd})$, we introduce the formal degrees $d_j = 2i-1$, for $j = (i-1)d+1, \dots id$, for $i = 1 \dots, n$. We define the norm

$$\|\phi\|_{we} = \left\|(|\phi_s^j|^{1/d_j})_{j=1,\dots,nd}\right\|_{L^2(0,1)}.$$

Now, as for (3.4.9), we generalize 3.4.8 to

$$C_A(x,y) = \{\phi \in L^2((0,1), \mathbb{R}^{nd}) : dv_s = A(v_s)\phi_s ds,\ x = v_0,\ y = v_1$$

. and we introduce

$$d_c(x,y) = \inf\left\{\|\phi\|_{we} : \phi \in C_A(x,y)\right\}.$$

Remark that when $n = 1$, i.e. we have just one elliptic d-dim differential equation, this corresponds to the usual definition of the Caratheodory distance.

We are interested in establishing a local equivalence between d, the semi-distance coming from the matrix-norm, and d_c, the distance in term of the control.

Proposition 3.29. *d, d_c are locally equivalent.*

Proof. Recall that in our notation d is the dimension of σ, d_j are the degrees associated to the directions j. We use in this proof some notions on similar metrics and pseudo-metrics for which we refer to [78]. Define

$$\rho(x,y) = \inf\{\delta > 0 | \exists \phi \in C_A(x,y),\ |\phi_s^j| < \delta^{d_j}, j = 1, \dots, nd\}.$$

It is also possible to allow only constant linear combinations of the vector fields:

$$\bar{C}_A(x,y) = \{\theta \in \mathbb{R}^{nd} : dv_s = A(v_s)\theta ds,\ x = v_0,\ y = v_1\}, \tag{3.5.24}$$

and define

$$\rho_2(x,y) = \inf\{\delta > 0 | \exists \theta \in \bar{C}_A(x,y),\ |\theta^j| < \delta^{d_j}, j = 1, \dots, nd\}.$$

In [78] the pseudo-distances ρ and ρ_2 are proved equivalent. We use here only the trivial inequality $\rho \le \rho_2$. Remark that an equivalent definition of ρ would be to define it as d_c taking the norm $\|\phi\|_\infty = \sup_{0 \le s \le 1} |\phi_s|$ instead of $\|\phi\|_2 = \left(\int_0^1 |\phi_s|^2 ds\right)^{1/2}$. So $d_c \le \rho$ follows easily from the fact that the $L^2(0,1)$ norm is dominated by the $L^\infty(0,1)$ norm.

The proof of

$$d(x,y) < \sqrt{R} \Rightarrow \rho_2(x,y) < C\sqrt{R},$$

for $R \le R_x$ is analogous to what we have done in Lemma 3.19. By definition, $d(x,y) < \sqrt{R}$ means $|x-y|_{A_R(x)} < 1$. This implies that it exists $\theta \in \bar{C}_A(x,y)$ such that $|\theta^{i(d-1)+1,\dots,id}| < (2\sqrt{R})^{2i-1}$, $v(0) = x$, $v(1) = y$. Indeed, consider the function

$$\Phi(\theta) = \int_0^1 A(v_s(\theta))\theta ds, \tag{3.5.25}$$

with $v(\theta)$ satisfying $dv_s(\theta) = A(v_s(\theta))\theta ds$, $v_0(\theta) = x$. Remark that $\Phi : \mathbb{R}^{nd} \to \mathbb{R}^{nd}$, $\Phi(0) = 0$ and $\Phi'(0) = A(x)$, which is invertible because of (H_1). Recall

$$|x - y|_{A_R(x)} < 1 \Rightarrow |x - y| < C_x\sqrt{R} \leq C_x\sqrt{R_x}.$$

For R_x small enough, we have that there exist θ such that $\Phi(\theta) = x - y$ (local inversion theorem), so $\theta \in \bar{C}_A$. From (3.5.25) it is also clear that $|\theta| < CR^{\frac{1}{2}}$. We now show

$$|\theta^{i(d-1)+1,\dots,id}| < (2\sqrt{R})^{2i-1}, \quad j = 1\dots, nd.$$

We have

$$\begin{aligned}
A_R(x)^{-1}(x-y) &= A_R(x)^{-1}\int_0^1 A(v_s(\theta))\theta ds \\
&= A_R(x)^{-1}A(x)\theta + A_R(x)^{-1}\int_0^1 (A(v_s(\theta)) - A(x))\theta ds \\
&= \left(\frac{\theta_j}{R^{d_j/2}}\right)_{j=1,\dots,nd} + J(\theta, R).
\end{aligned}$$

Consider

$$J(\theta, R) = A_R(x)^{-1}\int_0^1 (A(v_s(\theta)) - A(x))\theta ds.$$

Remark $|A(v_s(\theta)) - A(x)| \leq C|\theta|$. Because of the block-triangular structure of A,

$$(J(\theta, R)_k)_{k=n(d-1)+1,\dots,nd} = \frac{(D_{x_{n-1}}B_n \dots D_{x_1}B_2\sigma(x))^{-1}}{R^{n-1/2}} \times$$
$$\left(\int_0^1 [D_{x_{n-1}}B_n \dots D_{x_1}B_2\sigma(v_s(\theta)) - D_{x_{n-1}}B_n \dots D_{x_1}B_2\sigma(x)]\theta^{n(d-1)+1,\dots,nd}ds\right)$$

so, since $|\theta| < CR^{\frac{1}{2}}$,

$$|(J(\theta, R)_k)_{k=n(d-1)+1,\dots,nd}| \leq C|\theta|\frac{|\theta^{n(d-1)+1,\dots,nd}|}{R^{n-1/2}}.$$

For $R \leq R_x$ small enough, $C|\theta| \leq CR_x^{1/2} < 1/2$. So

$$|(J(\theta, R)_k)_{k=n(d-1)+1,\dots,nd}| < \frac{1}{2}\frac{|\theta^{n(d-1)+1,\dots,nd}|}{R^{n-1/2}}.$$

Recall that

$$\begin{aligned}
1 \geq |A_R(x)^{-1}(x-y)| &= \left|\left(\frac{\theta_j}{R^{d_j/2}}\right)_{j=1\dots nd} + J(\theta, R)\right| \\
&\geq \left|\left(\frac{\theta_j}{R^{d_j/2}}\right)_{j=n(d-1)+1,\dots,nd} + (J(\theta, R)_j)_{j=n(d-1)+1,\dots,nd}\right| \\
&\geq \left|\left(\frac{\theta^{n(d-1)+1,\dots,nd}}{R^{n-1/2}}\right)_{j=n(d-1)+1,\dots,nd}\right| - |(J(\theta, R)_j)_{j=n(d-1)+1,\dots,nd}| \\
&> \frac{1}{2}\frac{|\theta^{n(d-1)+1,\dots,nd}|}{R^{n-1/2}}.
\end{aligned}$$

Then we have

$$|\theta^{n(d-1)+1,\dots,nd}| < 2R^{n-1/2} < (2\sqrt{R})^{2n-1}.$$

With the same method, considering $n-j$ with an induction j, we can prove that

$$|\theta^{i(d-1)+1,\dots,id}| < (2\sqrt{R})^{2i-1}, \quad 1 \le i \le n.$$

It also holds

$$d_c(x,y) < \frac{\sqrt{R}}{K} \Rightarrow d(x,y) < \sqrt{R},$$

if K is large enough. With the same notation as above, in particular supposing $\exists \phi \in C_A(x,y)$ with $\|\phi\|_{we} \le \frac{\sqrt{R}}{K}$,

$$\begin{aligned}
|x-y|_{A_R(x)} &= \left| A_R(x)^{-1} \int_0^1 A(v_s)\phi_s ds \right| \\
&\le C \left| \left(\frac{\int_0^1 \phi_s^j ds}{R^{d_j/2}} \right)_{j=1\dots,nd} \right| \\
&\le C/K < 1
\end{aligned}$$

So $d(x,y) \le \sqrt{R}$, and this concludes the local equivalence of d, d_c, ρ, ρ_2. □

Chapter 4

Tubes estimates for diffusion processes under a local Hörmander condition of order one

4.1 Introduction

In this chapter we consider a diffusion process in $\mathbb{R}^n$ solution of

$$dX_t = \sum_{j=1}^{d} \sigma_j(t, X_t) \circ dW_t^j + b(t, X_t)dt, \qquad X_0 = x. \tag{4.1.1}$$

where $W = (W^1, ..., W^d)$ is a standard Brownian motion and $\circ dW_t^j$ denotes the Stratonovich integral. We assume $\sigma_j, b : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ three time differentiable in $x \in \mathbb{R}^n$ and one time differentiable in time $t \in \mathbb{R}^+$, and that the derivatives with respect to the space $x \in \mathbb{R}^n$ are one time differentiable with respect to t. We assume that the coefficients σ_j, b verify the strong Hörmander condition of order one (involving σ_j and the first order Lie brackets $[\sigma_i, \sigma_j]$) locally around a skeleton path

$$dx_t(\phi) = \sum_{j=1}^{d} \sigma_j(t, x_t(\phi))\phi_t^j dt + b(t, x_t(\phi))dt. \tag{4.1.2}$$

As in chapter 3, we use a norm which reflects the non isotropic structure of the problem, i.e. the fact that the diffusion process X_t moves with speed $\sqrt{t}$ in the direction of the diffusion vector fields σ_j and with speed $t = \sqrt{t} \times \sqrt{t}$ in the direction of $[\sigma_i, \sigma_j]$. We prove that this norm, that we denote $|\xi|_{A_R}$ (see (4.2.3)) is equivalent with the standard control metric d_c. We find exponential lower and upper bounds for the probability that the diffusion remains in a small tube around the skeleton path, i.e. $\mathbb{P}\left(\sup_{t\leq T} |X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} \leq 1\right)$. The proof of this result is based on a diagonal two-sided bound for the density in short time, and a concatenation procedure. Our density estimate is interesting in comparison with the classical result (1.1.6) (see [72]), since also here we work in a strong Hörmander framework, but we allow for a general drift, and moreover our coefficients may depend on t (see [34]). This work is the continuation of [6], where the lower bound was proved.

4.2 Notations and main results

For $(t,x) \in \mathbb{R}^+ \times \mathbb{R}^n$ we denote by $n(t,x)$ a constant such that for every $s \in [(t-1)\vee 0, t+1], y \in B(x,1)$ and for every multi index α of length less than or equal to three

$$|\partial_x^\alpha b(s,y)| + |\partial_t \partial_x^\alpha b(s,y)| + \sum_{j=1}^d |\partial_x^\alpha \sigma_j(s,y)| + |\partial_t \partial_x^\alpha \sigma_j(s,y)|) \leq n(t,x). \qquad (4.2.1)$$

Here, $\alpha = (\alpha_1, ..., \alpha_k) \in \{1, ..., n\}^k$ represents a multi-index, $|\alpha| = k$ the length of α and $\partial_x^\alpha = \partial_{x_{\alpha_1}} ... \partial_{x_{\alpha_k}}$. We do not assume global Lipschitz continuity or sublinear growth properties for the coefficients, so the above SDE might not have a unique solution. We only assume to work with a continuous adapted process X solving (4.1.1) on the time interval $[0,T]$.

We need to recall some notations. For $f, g : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ we define the directional derivative $\partial_g f(t,x) = \sum_{i=1}^n g^i(t,x) \partial_{x_i} f(t,x)$, and we recall that the Lie bracket (with respect to the space variable x) is defined as $[g,f](t,x) = \partial_g f(t,x) - \partial_f g(t,x)$. Let $M \in \mathcal{M}_{n\times m}$ be a matrix with full row rank. We write M^T for the transposed matrix, and MM^T is invertible. We denote by $\lambda_\#(M)$ (respectively $\lambda^\#(M)$) the smallest (respectively the largest) eigenvalue of MM^T and we consider the following norm on $\mathbb{R}^n$:

$$|y|_M = \sqrt{\langle (MM^T)^{-1} y, y \rangle}. \qquad (4.2.2)$$

Here and all along this chapter $d^2 = m$. We are concerned with the matrix $A(t,x) = (A_l(t,x))_{l=1,...,m}$, defined as follows. Let $l = (p-1)d + i \in \{1, \dots, m\}$, with $p, i \in \{1, \dots, d\}$.

$$\begin{aligned} A_l(t,x) &= [\sigma_i, \sigma_p](t,x) \quad if \quad i \neq p, \\ &= \sigma_i(t,x) \quad if \quad i = p. \end{aligned}$$

For $R > 0$, we define $\overline{R}$ the diagonal $m \times m$ matrix with $\overline{R}_{l,l} = R$ for $i \neq p$ and $\overline{R}_{l,l} = \sqrt{R}$ for $i = p$. Moreover,

$$A_R(t,x) = A(t,x)\overline{R} = (\sqrt{R}\sigma_i(t,x), [\sqrt{R}\sigma_j, \sqrt{R}\sigma_p](t,x))_{i,j,p=1,\dots,d, j\neq p}. \qquad (4.2.3)$$

We denote by $\lambda(t,x)$ the smallest eigenvalue of $A(t,x)A(t,x)^T$, i.e.

$$\lambda(t,x) = \inf_{|\xi|=1} \sum_{i=1}^m \langle A_i(t,x), \xi \rangle^2, \qquad (4.2.4)$$

Consider now some $x \in \mathbb{R}^n, t \geq 0$ such that $(\sigma_i(t,x), [\sigma_j, \sigma_p](t,x))_{i,j,p=1,\dots,d,j\neq p}$ span $\mathbb{R}^n$. Then $A_R A_R^T(t,x)$ is invertible and we may define $|y|_{A_R(t,x)}$. We also denote $\langle \sigma(t,x) \rangle$ the subspace of $\mathbb{R}^n$ spanned by $\sigma_i(t,x), i = 1, ..., d$.

For $\mu \geq 1$ and $0 < h \leq 1$ we denote by $L(\mu, h)$ the class of non negative functions $f : \mathbb{R}^+ \to \mathbb{R}^+$ which have the property

$$f(t) \leq \mu f(s) \quad \text{for} \quad |t-s| \leq h.$$

We will make the following hypothesis: there exists some functions $n : [0,T] \to [1,\infty)$ and $\lambda : [0,T] \to (0,1]$ such that for some $\mu \geq 1$ and $0 < h \leq 1$ we have

$$\begin{array}{lll} (H_1) & n(t, x_t(\phi)) & \leq n_t, \forall t \in [0,T], \\ (H_2) & \lambda(t, x_t(\phi)) & \geq \lambda_t > 0, \forall t \in [0,T], \\ (H_3) & |\phi_\cdot|^2, n_\cdot, \lambda_\cdot & \in L(\mu, h). \end{array} \tag{4.2.5}$$

Remark 4.1. Hypothesis (H_2) implies that for each $t \in (0,T)$, the space $\mathbb{R}^n$ is spanned by the vectors $(\sigma_i(t,x_t), [\sigma_j, \sigma_p](t,x_t))_{i,j,p=1,\dots,d,j<p}$, so the Hörmander condition holds along the curve $x_t(\phi)$.

The main result in this chapter is the following:

Theorem 4.14 . *Suppose that* (H_1), (H_2) *and* (H_3) *hold and that* $X_0 = x_0(\phi)$. *There exist* K, q *universal constants such that for* $H_t = K\left(\frac{\mu n_t}{\lambda_t}\right)^q$, *for* $R \leq R_*(\phi)$ *(defined in* (4.6.1)*),*

$$\exp\left(-\int_0^T H_t\left(\frac{1}{R} + |\phi_t|^2\right)dt\right) \\ \leq \mathbb{P}\left(\sup_{t\leq T} |X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} \leq 1\right) \leq \exp\left(-\int_0^T \frac{1}{H_t}\left(\frac{1}{R} + |\phi_t|^2\right)dt\right) \tag{4.2.6}$$

Remark 4.2. Suppose $X_t = W_t$ and $x_t(\phi) = 0$, so that $n_t = 1$, $\lambda_t = 1$, $\mu = 1$ and $\phi_t = 0$. Then $|X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} = R^{-1/2}W_t$ and we obtain $\exp(-C_1T/R) \leq \mathbb{P}(\sup_{t\leq T}|W_t| \leq \sqrt{R}) \leq \exp(-C_2T/R)$ which is coherent with the standard estimate (see [64]).

The proof of Theorem 4.14 relies on the following two-sided bound for the density of equation (4.1.1) in short time. The estimate is *diagonal*, meaning that it is local around the drifted initial condition $X_0 + b(0,X_0)\delta$.

Theorem 4.5 . *Suppose that* (H_1) *and* (H_2) *hold locally around* X_0. *Then there exist constants* r, δ_*, C *such that for* $\delta \leq \delta_*$, $|y - X_0 - b(0,X_0)\delta|_{A_\delta(0,X_0)} \leq r$,

$$\frac{1}{C\delta^{n-\frac{dim\langle\sigma(0,X_0)\rangle}{2}}} \leq p_{X_\delta}(y) \leq \frac{C}{\delta^{n-\frac{dim\langle\sigma(0,X_0)\rangle}{2}}}.$$

A global two-sided bound for the density of X_t is proved in [72], under the *strong* Hörmander non-degeneracy condition. It is also assumed that the coefficients do not depend time, i.e. $b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$, and that $b(x) = \sum_{j=1}^d \alpha_i\sigma_i(x)$, with

$\alpha_i \in C_b^\infty(\mathbb{R}^n)$ (i.e. the drift is generated by the vector fields of the diffusive part, which is a quite restrictive hypothesis). This bound is Gaussian in the control metric that we now define. For $x, y \in \mathbb{R}^n$ we denote by $C(x,y)$ the set of controls $\psi \in L^2([0,1];\mathbb{R}^d)$ such that the corresponding skeleton solution of

$$du_t(\psi) = \sum_{j=1}^d \sigma_j(u_t(\psi))\psi_t^j dt, \qquad u_0(\psi) = x$$

satisfies $u_1(\psi) = y$. The control (Caratheodory) distance is defined as

$$d_c(x,y) = \inf\Big\{\Big(\int_0^1 |\psi_s|^2\, ds\Big)^{1/2} : \psi \in C(x,y)\Big\}.$$

Their result is the following: there exist a constant $M \geq 1$ such that

$$\frac{1}{M|B_{d_c}(x,t^{1/2})|}\exp\left(-\frac{Md_c(x,y)^2}{t}\right) \\ \leq p_t(x,y) \leq \frac{M}{|B_{d_c}(x,t^{1/2})|}\exp\left(-\frac{d_c(x,y)^2}{Mt}\right)$$

for $(t,x,y) \in (0,1] \times \mathbb{R}^n \times \mathbb{R}^n$, where $B_d(x,r) = \{y \in \mathbb{R}^n : d(x,y) < r\}$. It is natural at this point to wonder if d_c and $|\cdot|_{A_R}$ are somehow related. Recall that now, as in [72], $b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$. We define the semi distance d via: $d(x,y) < \sqrt{R}$ if $|x-y|_{A_R(x)} < 1$, and prove in section 4.7.2 the local equivalence of d and d_c. This allows us to state theorem 4.14 in the control metric:

Corollary 4.3. *There exist K, q constants such that for $H_t = K\left(\frac{\mu n_t}{\lambda_t}\right)^q$, for $R \leq R_*(\phi)$,*

$$\exp\left(-\int_0^T H_t\left(\frac{1}{R} + |\phi_t|^2\right)dt\right) \\ \leq \mathbb{P}\left(\sup_{0\leq t\leq T} d_c(X_t, x_t(\phi)) \leq \sqrt{R}\right) \leq \exp\left(-\int_0^T \frac{1}{H_t}\left(\frac{1}{R} + |\phi_t|^2\right)dt\right). \quad (4.2.7)$$

We present now two examples of application:
Example 1. Consider the two dimensional diffusion process

$$X_t^1 = x_1 + W_t^1, \quad X_t^2 = x_2 + \int_0^t X_s^1 dW_s^2.$$

Since $A_R A_R^T(x) = \begin{bmatrix} R & 0 \\ 0 & R(x_1 + 2R) \end{bmatrix}$, the associated norm is $|\xi|^2_{A_R(x)} = \frac{\xi_1^2}{R} + \frac{\xi_2^2}{R(x_1+2R)}$. On $\{x_1 = 0\}$, $|\xi|^2_{A_R(x)} = \frac{\xi_1^2}{R} + \frac{\xi_2^2}{2R^2}$ and consequently for R small $\{\xi : |\xi|_{A_R(x)} \leq 1\}$ is an ellipsoid.

If we take a path $x(t)$ with $x_1(t)$ which keeps far from zero then we have ellipticity along the path and we may use estimates for elliptic SDEs (see [10]). If $x_1(t) = 0$ for some $t \in [0, T]$ we need our estimate. Let us compare the norm in the two cases: if $x_1 > 0$ the diffusion matrix is non-degenerate and we can consider the norm $|\xi|_{B_R(x)}$ with $B_R(x) = R\sqrt{\sigma\sigma^T(x)}$. We have

$$|\xi|^2_{B_R(x)} = \frac{1}{R}\xi_1^2 + \frac{1}{Rx_1^2}\xi_2^2 \geq \frac{1}{R}\xi_1^2 + \frac{1}{R(2R + x_1)}\xi_2^2 = |\xi|^2_{A_R(x)},$$

and the two norms are equivalent for R small. Let us now take $x_t(\phi) = (0, 0)$. We have $n_s = 1$ and $\lambda_s = 1$ and $X_t - x_t = (W_t^1, \int_0^t W_s^1 dW_s^2)$, so we obtain

$$\begin{aligned} e^{-C_1T/R} &\leq \mathbb{P}\left(\sup_{t\leq T}\left\{\frac{1}{R}\left|W_t^1\right|^2 + \frac{1}{2R^2}\left|\int_0^t W_s^1 dW_s^2\right|^2\right\} \leq 1\right) \\ &= \mathbb{P}\left(\sup_{t\leq T}(|X_t - x_t|^2_{A_R(x_t)} \leq 1\right) \leq e^{-C_2T/R}. \end{aligned}$$

Example 2: The principal invariant diffusion on the Heisenberg group. Consider on $\mathbb{R}^3$ the vector fields $\partial_x - \frac{y}{2}\partial_z$ and $\partial_y - \frac{x}{2}\partial_z$. The associated Markov process is a Brownian motion on $\mathbb{R}^2$ and its Lï¿$\frac{1}{2}$vy area.

$$X_t^1 = x_1 + W_t^1, \quad X_t^2 = x_2 + W_t^2, \quad X_t^3 = x_3 + \frac{1}{2}\int_0^t X_s^1 dW_s^2 - \frac{1}{2}\int_0^t X_s^2 dW_s^1.$$

(cf. [47], [3], [73], where gradient bounds for the heat kernel are obtained, and [14]). Since the diffusion is in dimension $n = 3$ and the driving Brownian in dimension $d = 2$, ellipticity cannot hold. Direct computations give

$$\sigma_1(x) = \begin{pmatrix} 1 \\ 0 \\ -x_2/2 \end{pmatrix}, \quad \sigma_2(x) = \begin{pmatrix} 0 \\ 1 \\ x_1/2. \end{pmatrix}, \quad [\sigma_1, \sigma_2](x) = \partial_{\sigma_1}\sigma_2 - \partial_{\sigma_2}\sigma_1 = \begin{pmatrix} 0 \\ 0 \\ 1. \end{pmatrix}$$

Therefore $\sigma_1(x), \sigma_2(x), [\sigma_1, \sigma_2](x)$ span $\mathbb{R}^3$ and hypoellipticity holds. In $x = 0$ we have $|\xi|^2_{A_R(0)} = \frac{\xi_1^2+\xi_2^2}{R} + \frac{\xi_3^2}{2R^2}$, so taking as control $\phi_t = 0$, denoting $A_t(W) = \frac{1}{2}\int_0^t X_s^1 dW_s^2 - \frac{1}{2}\int_0^t X_s^2 dW_s^1$ (the Lï¿$\frac{1}{2}$vy area) we obtain

$$\begin{aligned} \mathbb{P}\left(\sup_{t\leq T/R}|W_t^1|^2 + |W_t^2|^2 + \frac{|A_t(W)|^2}{2} \leq 1\right) &= \mathbb{P}\left(\sup_{t\leq T}\frac{|W_t^1|^2 + |W_t^2|^2}{R} + \frac{|A_t(W)|^2}{2R^2} \leq 1\right) \\ &= \mathbb{P}\left(\sup_{t\leq T}|X_t|^2_{A_R(x_t(\phi))} \leq 1\right). \end{aligned}$$

Appling our tube estimate we have

$$e^{-C_1T/R} \leq \mathbb{P}\left(\sup_{t\leq T/R}|W_t^1|^2 + |W_t^2|^2 + \frac{|A_t(W)|^2}{2} \leq 1\right) \leq e^{-C_2T/R}.$$

4.3 Decomposition

We start with the decomposition of the process that will allow us to produce the lower bound in short time. Writing this decomposition we also introduce some notations that we will employ in this chapter, mainly in section 4.4.

4.3.1 Development

Using a development in stochastic Taylor series of order two we write

$$X_t = X_0 + Z_t + b(0, X_0)t + R_t \tag{4.3.1}$$

where

$$Z_t = \sum_{i=1}^{d} a_i W_t^i + \sum_{i,j=1}^{d} a_{i,j} \int_0^t W_s^i \circ dW_s^j \tag{4.3.2}$$

with $a_i = \sigma_i(0, X_0)$, $a_{i,j} = \partial_{\sigma_i}\sigma_j(0, X_0)$, and

$$\begin{aligned} R_t = & \sum_{j,i=1}^{d} \int_0^t \int_0^s (\partial_{\sigma_i}\sigma_j(u, X_u) - \partial_{\sigma_i}\sigma_j(0, X_0)) \circ dW_u^i \circ dW_s^j \\ & + \sum_{i=1}^{d} \int_0^t \int_0^s \partial_b \sigma_i(u, X_u) du \circ dW_s^i + \sum_{i=1}^{d} \int_0^t \int_0^s \partial_u \sigma_j(u, X_u) du \circ dW_s^i \\ & + \sum_{i=1}^{d} \int_0^t \int_0^s \partial_{\sigma_i} b(u, X_u) \circ dW_u^i ds + \int_0^t \int_0^s \partial_b b(u, X_u) du ds. \end{aligned} \tag{4.3.3}$$

Since $\mathcal{O}(R_t) = t^{3/2}$, we expect the behavior of X_t and Z_t to be somehow close. Our first goal is to give a decomposition for Z_t in (4.3.2). We start introducing some notation: we will write $[a]_{i,j} = a_{i,j} - a_{j,i} = [\sigma_i, \sigma_j](0, X_0)$. We fix $\delta > 0$ and denote $s_k(\delta) = \frac{k}{d}\delta$ and

$$\Delta_k^i(\delta, W) = W_{s_k(\delta)}^i - W_{s_{k-1}(\delta)}^i, \quad \Delta_k^{i,j}(\delta, W) = \int_{s_{k-1}(\delta)}^{s_k(\delta)} (W_s^i - W_{s_{k-1}}^i) \circ dW_s^j.$$

Notice that $\Delta_k^{i,j}(\delta, W)$ is the Stratonovich integral, but for $i \neq j$ it coincides with the Ito integral. When no confusion is possible we use the short notation $s_k = s_k(\delta), \Delta_k^i =$

$\Delta_k^i(\delta,W), \Delta_k^{i,j} = \Delta_k^{i,j}(\delta,W)$. Moreover for $p = 1,...,d$ we define

$$\begin{aligned}
\mu_p(\delta,W) &= \sum_{i\neq p}\Delta_i^p \\
\psi_p(\delta,W) &= \sum_{i\neq j, i\neq p, j\neq p} a_{i,j}\Delta_p^{i,j} + \sum_{l=p+1}^d \sum_{i\neq p}\sum_{j\neq l}^d a_{i,j}\Delta_l^j\Delta_p^i + \frac{1}{2}\sum_{i\neq p}^d a_{i,i}\left|\Delta_p^i\right|^2 \\
\varepsilon_p(\delta,W) &= \sum_{l>p}^d\sum_{j\neq l}^d a_{p,j}\Delta_l^j + \sum_{p>l}^d\sum_{j\neq l}^d a_{j,p}\Delta_l^j + \sum_{j\neq p} a_{p,j}\Delta_p^j \\
\eta_p(\delta,W) &= \frac{1}{2}a_{p,p}\left|\Delta_p^p\right|^2 + \sum_{l>p}^d a_{p,l}\Delta_l^l\Delta_p^p + \Delta_p^p\varepsilon_p.
\end{aligned} \tag{4.3.4}$$

We write $\eta(\delta,W) = \sum_{p=1}^d \eta_p(\delta,W)$ and $\psi(\delta,W) = \sum_{p=1}^d \psi_p(\delta,W)$. Our aim is to prove the following decomposition.

$$Z_\delta = \sum_{p=1}^d a_p(\Delta_p^p(\delta,W)+\mu_p(\delta,W)) + \sum_{p=1}^d\sum_{i\neq p}[a]_{i,p}\Delta_p^{i,p}(\delta,W)+\eta(\delta,W)+\psi(\delta,W) \tag{4.3.5}$$

Remark 4.4. The reason of this decomposition is the following. We split the time interval $(0,\delta)$ in d sub intervals of length δ/d. We also split the Brownian motion in corresponding increments: $(W_s^p - W_{s_{k-1}}^p)_{s_{k-1}\leq s\leq s_k}, p = 1,...,d$. Let us fix p. For $s\in(s_{p-1},s_p)$ we have the processes $(W_s^i - W_{s_{p-1}}^i)_{s_{p-1}\leq s\leq s_p}, i = 1,...,d$. Our idea is to settle a calculus which is based on W^p and to take conditional expectation with respect to $W^i, i\neq p$. So $(W_s^i - W_{s_{p-1}}^i)_{s_{p-1}\leq s\leq s_p}, i\neq p$ will appear as parameters (or controls) which we may choose in an appropriate way. The random variables on which the calculus is based are $\Delta_p^p = W_{s_p}^p - W_{s_{p-1}}^p$ and $\Delta_p^{i,p} = \int_{s_{p-1}}^{s_p}(W_s^i - W_{s_{p-1}}^i)dW_s^p, j\neq p$. These are the r.v. that we have emphasized in the decomposition of Z_δ. Notice that, conditionally to the controls $(W_s^i - W_{s_{p-1}}^i)_{s_{p-1}\leq s\leq s_p}, i\neq p$, this is a centered Gaussian vector and, under appropriate hypothesis on the controls this Gaussian vector is non degenerate (we treat in section 4.4.3 the problem of the choice of the controls). In order to handle the term $\Delta_p^{p,i} = \int_{s_{p-1}}^{s_p}(W_s^p - W_{s_{p-1}}^p)dW_s^i$. we use the identity $\Delta_p^{p,i} = \Delta_p^i\Delta_p^p - \Delta_p^{i,p}$. This is the reason for which $(a_{p,i} - a_{i,p}) = [a]_{p,i}$ appears.

We now prove (4.3.5). We decompose

$$Z_\delta = \sum_{l=1}^d Z(s_l) - Z(s_{l-1}) = \sum_{l=1}^d\left(\sum_{i=1}^d a_i\Delta_l^i + \sum_{i,j=1}^d a_{i,j}\int_{s_{l-1}}^{s_l} W_s^i\circ dW_s^j\right)$$

and write

$$\int_{s_{l-1}}^{s_l} W_s^i\circ dW_s^j = W_{s_{l-1}}^i\Delta_l^j + \Delta_l^{i,j} = (\sum_{p=1}^{l-1}\Delta_p^i)\Delta_l^j + \Delta_l^{i,j}.$$

Then

$$Z_\delta = \sum_{l=1}^{d}\sum_{i=1}^{d} a_i\Delta_l^i + \sum_{l=1}^{d}\sum_{i,j=1}^{d} a_{i,j}(\sum_{p=1}^{l-1}\Delta_p^i)\Delta_l^j + \sum_{l=1}^{d}\sum_{i,j=1}^{d} a_{i,j}\Delta_l^{i,j} =: S_1 + S_2 + S_3.$$

Notice first that

$$S_1 = \sum_{l=1}^{d} a_l\Delta_l^l + \sum_{l=1}^{d}\sum_{i\neq l} a_i\Delta_l^i.$$

We treat now S_3. We will use the identities

$$\left|\Delta_l^i\right|^2 = 2\Delta_l^{i,i} \quad \text{and} \quad \Delta_l^i\Delta_l^j = \Delta_l^{i,j} + \Delta_l^{j,i}.$$

Then

$$\begin{aligned}
S_3 &= \sum_{l=1}^{d}\sum_{i=1}^{d} a_{i,i}\Delta_l^{i,i} + \sum_{l=1}^{d}\sum_{i\neq j} a_{i,j}\Delta_l^{i,j} \\
&= \sum_{l=1}^{d}\sum_{i=1}^{d} a_{i,i}\Delta_l^{i,i} + \sum_{l=1}^{d}\sum_{i\neq l} a_{i,l}\Delta_l^{i,l} + \sum_{l=1}^{d}\sum_{j\neq l} a_{l,j}\Delta_l^{l,j} + \sum_{l=1}^{d}\sum_{i\neq j, i\neq l j\neq l} a_{i,j}\Delta_l^{i,j} \\
&= \frac{1}{2}\sum_{l=1}^{d}\sum_{i=1}^{d} a_{i,i}\left|\Delta_l^i\right|^2 + \sum_{l=1}^{d}\sum_{i\neq l} a_{i,l}\Delta_l^{i,l} \\
&\quad + \sum_{l=1}^{d}\sum_{j\neq l} a_{l,j}\left(\Delta_l^j\Delta_l^l - \Delta_l^{j,l}\right) + \sum_{l=1}^{d}\sum_{i\neq j, i\neq l, j\neq l} a_{i,j}\Delta_l^{i,j} \\
&= \frac{1}{2}\sum_{i=1}^{d} a_{i,i}\left|\Delta_i^i\right|^2 + \frac{1}{2}\sum_{l=1}^{d}\sum_{i\neq l}^{d} a_{i,i}\left|\Delta_l^i\right|^2 + \sum_{l=1}^{d}\sum_{i\neq l}(a_{i,l} - a_{l,i})\Delta_l^{i,l} \\
&\quad + \sum_{l=1}^{d}\left(\sum_{j\neq l} a_{l,j}\Delta_l^j\right)\Delta_l^l + \sum_{l=1}^{d}\sum_{i\neq j, i\neq l, \neq j\neq} a_{i,j}\Delta_l^{i,j}.
\end{aligned}$$

We treat now S_2. We want to emphasize the terms containing Δ_i^i. We have

$$S_2 = \sum_{l>p}^{d}\sum_{i,j=1}^{d} a_{i,j}\Delta_p^i\Delta_l^j = S_2' + S_2'' + S_2''' + S_2^{iv}$$

with $\sum_{l>p}^d = \sum_{p=1}^d\sum_{l=p+1}^d$ and

$$\begin{aligned}
S_2' &= \sum_{l>p}^{d} a_{p,l}\Delta_p^p\Delta_l^l, \quad S_2'' = \sum_{l>p}^{d}\sum_{j\neq l}^{d} a_{p,j}\Delta_p^p\Delta_l^j \\
S_2''' &= \sum_{l>p}^{d}\sum_{i\neq p}^{d} a_{i,l}\Delta_p^i\Delta_l^l, \quad S_2^{iv} = \sum_{l>p}^{d}\sum_{i\neq p, j\neq l}^{d} a_{i,j}\Delta_p^i\Delta_l^j.
\end{aligned}$$

We have

$$S_2'' = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=p+1}^{d} \sum_{j\neq l}^{d} a_{p,j} \Delta_l^j \right)$$

and

$$S_2''' = \sum_{l=1}^{d} \Delta_l^l \left(\sum_{p=1}^{l-1} \sum_{i\neq p}^{d} a_{i,l} \Delta_p^i \right) = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=1}^{p-1} \sum_{j\neq l}^{d} a_{j,p} \Delta_l^j \right)$$

so that

$$S_2'' + S_2''' = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=p+1}^{d} \sum_{j\neq l}^{d} a_{p,j} \Delta_l^j + \sum_{l=1}^{p-1} \sum_{j\neq l}^{d} a_{j,p} \Delta_l^j \right).$$

Finally

$$\begin{aligned} Z_\delta \;=\; & \sum_{l=1}^{d} a_l \Delta_l^l + \sum_{l=1}^{d} \sum_{i\neq l} a_i \Delta_l^i \\ & + \sum_{l>p}^{d} a_{p,l} \Delta_p^p \Delta_l^l + \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l>p}^{d} \sum_{j\neq l}^{d} a_{p,j} \Delta_l^j + \sum_{p>l}^{d} \sum_{j\neq l}^{d} a_{j,p} \Delta_l^j \right) \\ & + \sum_{l>p}^{d} \sum_{i\neq p, j\neq l}^{d} a_{i,j} \Delta_p^i \Delta_l^j + \frac{1}{2} \sum_{i=1}^{d} a_{i,i} \left|\Delta_i^i\right|^2 + \frac{1}{2} \sum_{l=1}^{d} \sum_{i\neq l}^{d} a_{i,i} \left|\Delta_l^i\right|^2 \\ & + \sum_{l=1}^{d} \sum_{i\neq l} (a_{i,l} - a_{l,i}) \Delta_l^{i,l} + \sum_{l=1}^{d} \left(\sum_{j\neq l} a_{l,j} \Delta_l^j \right) \Delta_l^l + \sum_{l=1}^{d} \sum_{i\neq j, i\neq l, j\neq l} a_{i,j} \Delta_l^{i,j}. \end{aligned}$$

We want to compute the coefficient of Δ_p^p : this term appears in $\sum_{p=1}^{d} \Delta_p^p (a_p + \varepsilon_p)$, with

$$\varepsilon_p = \sum_{l>p}^{d} \sum_{j\neq l}^{d} a_{p,j} \Delta_l^j + \sum_{p>l}^{d} \sum_{j\neq l}^{d} a_{j,p} \Delta_l^j + \sum_{j\neq p} a_{p,j} \Delta_p^j.$$

We consider now $\Delta_p^{i,p}$. It appears in

$$\sum_{p=1}^{d} \sum_{i\neq p} (a_{i,p} - a_{p,i}) \Delta_p^{i,p}$$

The other terms are

$$\begin{aligned} & \sum_{l=1}^{d} \sum_{i\neq l} a_i \Delta_l^i + \sum_{l>p}^{d} \sum_{i\neq p, j\neq l}^{d} a_{i,j} \Delta_p^i \Delta_l^j + \frac{1}{2} \sum_{i=1}^{d} a_{i,i} \left|\Delta_i^i\right|^2 + \frac{1}{2} \sum_{l=1}^{d} \sum_{i\neq l}^{d} a_{i,i} \left|\Delta_l^i\right|^2 \\ & + \sum_{l=1}^{d} \sum_{i\neq j, i\neq l, j\neq l} a_{i,j} \Delta_l^{i,j} + \sum_{l=p+1}^{d} a_{p,l} \Delta_p^p \Delta_l^l. \end{aligned}$$

We put everything together and (4.3.5) is proved.

4.3.2 Main Gaussian component

Let $l=(p-1)d+i\in\{1,...,m\}$ with $p,i\in\{1,...,d\}$. We define $B_t=\delta^{-1/2}W_{t\delta}$ and denote

$$\begin{aligned}\Theta_l &= \frac{1}{\delta}\Delta_p^{i,p}=\int_{\frac{p-1}{d}}^{\frac{p}{d}}(B_s^i-B_{\frac{p-1}{d}}^i)dB_s^p \quad if \quad i\neq p \\ &= \frac{1}{\sqrt{\delta}}\Delta_p^p=B_{\frac{p}{d}}^p-B_{\frac{p-1}{d}}^p \quad if \quad i=p.\end{aligned} \tag{4.3.6}$$

We will also denote $l(p)=(p-1)d+p$ so that $\Theta_{l(p)}=\frac{1}{\sqrt{\delta}}\Delta_p^p$. We consider the σ field

$$\mathcal{G}:=\sigma(W_s^j-W_{s_{p-1}(\delta)}^j,s_{p-1}(\delta)\leq s\leq s_p(\delta),p=1,...d,j\neq p). \tag{4.3.7}$$

For $p=1,...,d$ we denote $\Theta_{(p)}=(\Theta_{(p-1)d+1},...,\Theta_{pd})$. Notice that conditionally to $\mathcal{G}$ the random variables $\Theta_{(p)},p=1,...,d$ are independent centered Gaussian d dimensional vectors and the covariance matrix Q_p of $\Theta_{(p)}$ is given by

$$\begin{aligned}Q_p^{p,j} &= Q_p^{j,p}=\int_{\frac{p-1}{d}}^{\frac{p}{d}}\left(B_s^j-B_{\frac{p-1}{d}}^j\right)ds, \quad j\neq p, \\ Q_p^{i,j} &= \int_{\frac{p-1}{d}}^{\frac{p}{d}}\left(B_s^j-B_{\frac{p-1}{d}}^j\right)\left(B_s^i-B_{\frac{p-1}{d}}^i\right)ds, \quad j\neq p,i\neq p, \\ Q_p^{p,p} &= \frac{1}{d}.\end{aligned}$$

It is easy to see that $\det Q_p\neq 0$ almost surely. It follows that conditionally to $\mathcal{G}$ the random variable $\Theta=(\Theta_{(1)},...,\Theta_{(d)})$ is a centered $m=d^2$ dimensional Gaussian vector. Its covariance matrix Q is a block-diagonal matrix built with Q_p, $p=1,\ldots,d$. In particular $\det Q=\prod_{p=1}^d\det Q_p\neq 0$ almost surely, and $\lambda_*(Q)=\min_{p=1,\ldots,d}\lambda_{*,p}(Q)$. We also have $\lambda^*(Q)=\max_{p=1,\ldots,d}\lambda_p^*(Q)$. We will need to work on subsets where we have a quantitative control of this quantities.

4.3.3 Decomposition.

Recall (4.7.10). We have

$$\begin{aligned}A_l(0,X_0) &= [a]_{i,p} \quad if \quad i\neq p, \\ &= a_p \quad if \quad i=p,\end{aligned} \tag{4.3.8}$$

We denote by $A_\delta^i\in\mathbb{R}^m,i=1,...,n$ the rows of the matrix A_δ. We also denote $S=\langle A_\delta^1,...,A_\delta^n\rangle\subset\mathbb{R}^m$ and $S^\perp$ its orthogonal. If hypothesis (H_2) holds the columns of A_δ span $\mathbb{R}^n$ so the rows are linearly independent. It follows that $S^\perp$ has dimension $m-n$. We take $\Gamma_\delta^i,i=n+1,...,m$ to be an orthonormal basis in $S^\perp$ and we denote $\Gamma_\delta^i=A_\delta^i(0,X_0)$ for $i=1,...,n$. We also denote $\underline{\Gamma}_\delta$ the $(m-n)\times m$ matrix with rows

$\Gamma_\delta^i, i = n+1, \ldots, m$. Finally we denote by Γ_δ the $m \times m$ dimensional matrix with rows $\Gamma_\delta^i, i = 1, ..., m$. Notice that

$$\Gamma_\delta \Gamma_\delta^T = \begin{pmatrix} A_\delta A_\delta^T(0, X_0) & 0 \\ 0 & Id_{m-n} \end{pmatrix} \tag{4.3.9}$$

where Id_{m-n} is the identity matrix in $\mathbb{R}^{m-n}$. It follows that for a point $y = (y_{(1)}, y_{(2)}) \in \mathbb{R}^m$ with $y_{(1)} \in \mathbb{R}^n, y_{(2)} \in \mathbb{R}^{m-n}$ we have

$$|y|_{\Gamma_\delta}^2 = \left|y_{(1)}\right|_{A_\delta(0,X_0)}^2 + \left|y_{(2)}\right|^2 \tag{4.3.10}$$

where we recall that $|y|_{\Gamma_\delta}^2 = \left\langle (\Gamma_\delta \Gamma_\delta^T)^{-1} y, y \right\rangle$. For $a \in \mathbb{R}^m$ we define the immersion $J_a : \mathbb{R}^n \to \mathbb{R}^m$ by

$$z \in \mathbb{R}^n \to J_a(z) = y \in \mathbb{R}^m \text{ with } y_i = z_i, i = 1, ..., n \quad \text{and} \quad y_i = \left\langle \Gamma_\delta^i, a \right\rangle, i = n+1, ..., m. \tag{4.3.11}$$

In particular $J_0(z) = (z, 0, ..., 0)$ and

$$|J_0 z|_{\Gamma_\delta} = |z|_{A_\delta(0,X_0)}. \tag{4.3.12}$$

Finally we denote

$$V_\omega = \sum_{p=1}^d a_p \mu_p(\delta, W) + \psi(\delta, W), \tag{4.3.13}$$

$$\eta_\omega(\Theta) = \sum_{p=1}^d \left(\frac{a_{p,p}}{2} \delta \Theta_{l(p)}^2 + \delta^{1/2} \Theta_{l(p)} \varepsilon_p + \sum_{q>p}^d a_{p,q} \delta \Theta_{l(q)} \Theta_{l(p)} \right)$$

where $\mu_p(\delta, W), \psi(\delta, W)$ and ε_p are defined in (4.3.4). Recall that $\Theta_{l(p)} = \delta^{-1/2} \Delta_p^p$ so that $\eta_\omega(\Theta) = \sum_{p=1}^d \eta_p(\delta, W)$. Now the decomposition (4.3.5) may be written as

$$Z_\delta = V_\omega + A_\delta(0, X_0)\Theta + \eta_\omega(\Theta).$$

We embed this relation in $\mathbb{R}^m$ and obtain

$$J_\Theta(Z_\delta) = J_0(V_\omega) + \Gamma_\delta \Theta + J_0(\eta_\omega(\Theta)).$$

Multiplying with Γ_δ^{-1} we set $\widetilde{V}_\omega = \Gamma_\delta^{-1} J_0(V_\omega)$, $\widetilde{\eta}_\omega(\Theta) = \Gamma_\delta^{-1} J_0(\eta_\omega(\Theta))$, $\widetilde{Z} = \Gamma_\delta^{-1} J_\Theta(Z_\delta)$. Now we define

$$G = \Theta + \widetilde{\eta}_\omega(\Theta), \tag{4.3.14}$$

and we have

$$\widetilde{Z} = \widetilde{V}_\omega + \Theta + \widetilde{\eta}_\omega(\Theta) = \widetilde{V}_\omega + G. \tag{4.3.15}$$

4.4 Lower bound for the density in short time

In this section we prove the lower bound of Theorem 4.5. First, using the local inversion theorem, we deal with the "Gaussian" part of the diffusion. We find a lower bound for the density conditioning in an appropriate way on the Brownian paths. Regarding this, we refer to section 4.4.3 for some technical results. The second step requires Malliavin calculus techniques, presented in Chapter 2. Those techniques allow us to handle the "non-Gaussian" remainder, and find lower bounds for the density of the rescaled diffusion F (cf. (4.4.14)). We recover the result for p_{X_δ} with a suitable change of variable.

Let X_t be solution of (4.1.1), and X_0 its initial condition. We introduce the class of constants

$$\mathcal{C} = \{C = K\left(n(0,X_0)/\lambda(0,X_0)\right)^q, \exists K, q \geq 1 \text{constants}\}, \qquad 1/\mathcal{C} = \{1/C : C \in \mathcal{C}\}.$$

In what follows when we write C we mean constants in this class, which may vary from line to line. In this section, since the initial condition is fixed and the notation heavy, we write $\sigma = \sigma(0,X_0), b = b(0,X_0), A_\delta = A_\delta(0,X_0)$. We recall that $\langle\sigma\rangle$ is the subspace of $\mathbb{R}^n$ spanned by $\sigma_i, i = 1,\ldots,d$. In what follows we prove the lower bound of the following theorem:

Theorem 4.5. *There exist $r_*, \delta_* \in 1/\mathcal{C}$, $C \in \mathcal{C}$ such that for $\delta \leq \delta_*$, $|y - X_0 - b\delta|_{A_\delta} \leq r$*

$$\frac{1}{C\delta^{n-\frac{dim\langle\sigma\rangle}{2}}} \leq p_{X_\delta}(y)\frac{C}{\delta^{n-\frac{dim\langle\sigma\rangle}{2}}}. \tag{4.4.1}$$

4.4.1 Localized density

We need to determine a set of Brownian trajectories where we have a quantitative control on the "non-degeneracy" of the main Gaussian Θ, and for this purpose we use the notion of *localization* introduced in [7]. Suppose to have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $U \in [0,1]$. We denote $d\mathbb{P}_U = Ud\mathbb{P}$, and $p_{F,U}$ the density of a r.v. F when we endow Ω with the measure $Ud\mathbb{P}$. For details on this aspect (in particular in relation with Malliavin calculus) we refer to section 2.1.4. We denote

$$q_p(B) = \sum_{j\neq p}\left|B^j_{\frac{p}{d}} - B^j_{\frac{p-1}{d}}\right| + \sum_{j\neq p, i\neq p}\left|\int_{\frac{p-1}{d}}^{\frac{p}{d}}(B^j_s - B^j_{\frac{i-1}{d}})dB^i_s\right| \quad \text{and} \quad q(B) = \sum_{p=1}^d q_p(B).$$

For fixed $\varepsilon, \rho > 0$, for $p = 1,...,d$ we define the sets

$$\Lambda_{\rho,\varepsilon,p} = \Big\{\det Q_p \geq \varepsilon^\rho, \sup_{\frac{p-1}{d}\leq t\leq\frac{p}{d}}\sum_{j\neq p}\left|B^j_t - B^j_{\frac{p-1}{d}}\right| \leq \varepsilon^{-\rho}, q_p(B) \leq \varepsilon\Big\}.$$

By (4.4.25), using the scale invariance of Brownian Motion, we may find some constants c and ε_* such that

$$\mathbb{P}(\Lambda_{\rho,\varepsilon,p}) \geq c\varepsilon^{\frac{1}{2}d(d+1)} \quad \text{for} \quad \varepsilon \leq \varepsilon_*.$$

Let $\Lambda_{\rho,\varepsilon} = \cap_{p=1}^{d} \Lambda_{\rho,\varepsilon,p}$. Using the independence property we obtain

$$\mathbb{P}\big(\Lambda_{\rho,\varepsilon}\big) \geq c \times \varepsilon^{\frac{1}{2}d^2(d+1)}. \tag{4.4.2}$$

On the set $\Lambda_{\rho,\varepsilon} \in \mathcal{G}$ we have $\det Q = \prod_{p=1}^{d} \det Q_p \geq \varepsilon^{d\rho}$ and $\lambda^*(Q) \leq \varepsilon^{-2\rho}$. Remark that

$$\frac{\lambda^*(Q)}{\sqrt{m}} \leq |Q|_l := \left(\frac{1}{m} \sum_{1 \leq i,j \leq m} Q_{i,j}^2 \right)^{1/2} \leq \lambda^*(Q). \tag{4.4.3}$$

For $a > 0$ we introduce the following function,

$$\psi_a(x) = 1_{|x| \leq a} + \exp\left(1 - \frac{a^2}{a^2 - (x-a)^2}\right) 1_{a < |x| < 2a},$$

which is a mollified version of $1_{[0,a]}$. We can now define our *localization variables.* Taking $a_1 = \varepsilon^{-d\rho}, a_2 = \varepsilon^{-2\rho}, a_3 = d\varepsilon$, we set

$$\tilde{U} = (\psi_{a_1}(1/\det Q))\psi_{a_2}(|Q|_l)\psi_{a_3}(q(B)).$$

The following inclusions hold:

$$\Lambda_{\rho,\varepsilon} \subset \Big\{ \det Q \geq \varepsilon^{d\rho}, |Q|_l \leq \varepsilon^{-2\rho}, q(B) \leq d\varepsilon \Big\} = \{\tilde{U} = 1\} \subset \{\tilde{U} \neq 0\}. \tag{4.4.4}$$

We can consider $\tilde{U}$ as a smooth version of the indicator function of $\Lambda_{\rho,\varepsilon}$. We also define, for fixed r, $\bar{U} = \prod_{i=1^n} \psi_r(\Theta_i)$. Remark that $\tilde{U}$ depends on ε and $\bar{U}$ depends on r, and that $\tilde{U}$ is $\mathcal{G}$ measurable. We set $U = \bar{U}\tilde{U}$.

Lemma 4.6. *There exist $C \in \mathcal{C}$, $\varepsilon, r \in 1/\mathcal{C}$ such that for $|z| \leq r/2$,*

$$\frac{1}{C} \leq p_{\tilde{Z},U}(z). \tag{4.4.5}$$

Proof. STEP 1: We start proving that there exist $C \in \mathcal{C}$, $\varepsilon, r \in 1/\mathcal{C}$ such that, on $\tilde{U} \neq 0$, for $|z| \leq r/2$

$$\frac{1}{C} \leq p_{\tilde{Z},\bar{U},|\mathcal{G}}(z).$$

Here $p_{\tilde{Z},\bar{U},|\mathcal{G}}$ represents the localized density of $\tilde{Z}$ conditioned to $\mathcal{G}$, i.e.

$$\mathbb{E}[f(\tilde{Z})\bar{U}|\mathcal{G}] = \int f(z) p_{\tilde{Z},\bar{U},|\mathcal{G}}(z) dz. \tag{4.4.6}$$

for f positive, measurable, with support included in $B(0, r/2)$. On $\tilde{U} \neq 0$, $\lambda^*(Q) \leq 2\sqrt{m}\varepsilon^{-2\rho}$, and

$$\frac{\varepsilon^{d\rho}}{2} \leq \det Q \leq \lambda_*(Q)\lambda^*(Q)^{m-1} \leq \lambda_*(Q)(2\sqrt{m})^{m-1}\varepsilon^{-2\rho(m-1)},$$

and this gives $\lambda_*(Q) \geq \frac{\varepsilon^{3m\rho}}{(2\sqrt{m})^m}$. So, fixing $\rho = 1/(8m)$, for $\varepsilon \leq \varepsilon^*$,

$$\frac{1}{16m^2} \frac{\lambda_*(Q)}{\lambda^*(Q)} \geq C_m \varepsilon^{3m\rho+2\rho} \geq \varepsilon. \tag{4.4.7}$$

To apply (2.2.16) to $G = \Theta + \widetilde{\eta}_\omega(\Theta)$ (with $m = d$), we need to check the hypothesis of Lemma 2.13. We use in the following the notation of section 2.2.3, in particular for $c_*(\tilde{\eta}_\omega, r)$ and $c_i(\tilde{\eta}_\omega)$. The third order derivatives of η_ω are null so we have $c_3(\tilde{\eta}_\omega) = 0$. For $i = l(p)$ and $j = l(q)$ we have $\partial_{i,j}\eta_\omega(\Theta) = \delta a_{ij}$, otherwise we get $\partial_{i,j}\eta_\omega(\Theta) = 0$. So $|\partial_{i,j}\eta_\omega(\Theta)| \leq \delta \sum_{i,j} a_{i,j}$. Using (4.7.2) we obtain

$$|\partial_{i,j}\tilde{\eta}_\omega(\Theta)| = |J_0(\partial_{i,j}\eta_\omega(\Theta))|_{\Gamma_\delta} = |\partial_{i,j}\eta_\omega(\Theta)|_{A_\delta} \leq C \in \mathcal{C}.$$

So

$$\frac{1}{C_1} \leq \frac{1}{16m^2(c_2(\tilde{\eta}_\omega) + \sqrt{c_3(\tilde{\eta}_\omega)})}, \quad \exists C_1 \in \mathcal{C} \tag{4.4.8}$$

We compute now the first order derivatives. For $j \notin \{l(p) : p = 1, ..., d\}$ we have $\partial_j \eta_\omega = 0$ and for $j = l(p)$ we have

$$\partial_p \eta_\omega(\Theta) = \delta \sum_{q=p}^{d} a_{p\wedge q, p\vee q} \Theta_{l(q)} + \sqrt{\delta}\varepsilon_p.$$

So, as above, we obtain $|\partial_j \tilde{\eta}_\omega(\Theta)| \leq C(|\Theta| + |\varepsilon_j|/\sqrt{\delta})$. Remark now that on $\{\bar{U} \neq 0\}$ we have $|\Theta| \leq Cr$, and on $\{\tilde{U} \neq 0\}$ we have $q(B) \leq 2d\varepsilon$, so

$$\sum_{j=1}^{d} |\varepsilon_j| \leq C\sqrt{\delta} q(B) \leq C\sqrt{\delta}\varepsilon. \tag{4.4.9}$$

Therefore

$$c_*(\tilde{\eta}_\omega, 16r) \leq C_2(r + \varepsilon), \quad \exists C_2 \in \mathcal{C}. \tag{4.4.10}$$

We also consider the following estimate of $\left|\widetilde{V}_\omega\right| = |V_\omega|_{A_\delta}$. We have

$$\left|\sum_{p=1}^{d} a_p \mu_p(\delta, W)\right|_{A_\delta} = \frac{1}{\sqrt{\delta}} |A_\delta J_0(\mu_p(\delta, W))|_{A_\delta} \leq \frac{1}{\sqrt{\delta}} |\mu_p(\delta, W))| \leq Cq(B)$$

and

$$|\psi(\delta, W)|_{A_\delta} \leq \frac{|\psi(\delta, W)|}{\delta\sqrt{\lambda_\#(A)}} \leq Cq(B)$$

so

$$\left|\widetilde{V}_\omega\right| \leq Cq(B) \leq C_3\varepsilon, \quad \exists C_3 \in \mathcal{C}. \tag{4.4.11}$$

We consider (4.4.11), and fix $\frac{r}{\varepsilon} = 2C_3 \in \mathcal{C}$, so $|\widetilde{V}_\omega| \leq r/2$. Then we consider (4.4.10), and

$$c_*(\tilde{\eta}_\omega, 4r) \leq C_2(2C_3+1)\varepsilon \leq \varepsilon^{1/2}, \quad for\, \varepsilon \leq \frac{1}{(4C_2C_3)^2} \in \frac{1}{\mathcal{C}}.$$

Moreover, looking at (4.4.8)

$$r = 2C_3\varepsilon \leq \frac{1}{C_1} \quad for\, \varepsilon \leq \frac{1}{2C_1C_3} \in \frac{1}{\mathcal{C}}.$$

So, with

$$\varepsilon = \varepsilon^* \wedge \frac{1}{(4C_2C_3)^2} \wedge \frac{1}{2C_1C_3} \in \frac{1}{\mathcal{C}},$$

and $r = 2C_3\varepsilon$ we have

$$|\widetilde{V}_\omega| \leq \frac{r}{2}; \qquad c_*(\tilde{\eta}_\omega, 4r) \leq \varepsilon^{1/2}; \qquad r \leq \frac{1}{C_1}.$$

Now, using also (4.4.7) and (4.4.8), we prove that (2.2.15) holds, and we can apply Lemma 2.13. We obtain

$$\frac{1}{K \det Q^{1/2}} \exp\left(-\frac{K}{\lambda_*(Q)}|z|^2\right) \leq p_{G,\bar{U},|\mathcal{G}}(z)$$

for $|z| \leq r$, where K does not depend on σ, b. Remark that, using $\lambda_*(Q) \geq \frac{\varepsilon^{3m\rho}}{(2\sqrt{m})^m}$, $\rho = \frac{1}{8m}$, $r/\varepsilon = 2C_1$ and $\varepsilon \leq 1/(4C_2C_1)^2$,

$$\frac{|z|^2}{\lambda_*(Q)} \leq \frac{(2\sqrt{m})^m r^2}{\varepsilon^{3m\rho}} \leq (2\sqrt{m})^m \frac{r^2}{\varepsilon} \leq (2\sqrt{m})^m \frac{r^2}{\varepsilon^2}\varepsilon \leq (2\sqrt{m})^m (2C_1)^2 \varepsilon \leq \bar{K} \quad (4.4.12)$$

where $\bar{K}$ does not depend on σ, b. Therefore $\frac{1}{C} \leq p_{G,\bar{U},|\mathcal{G}}(z)$, for $|z| \leq r$, for some $C \in \mathcal{C}$, on $\tilde{U} \neq 0$. Now recall $|\widetilde{V}_\omega| \leq r/2$ and (4.3.15). Hence for $|z| \leq r/2$

$$\frac{1}{C} \leq p_{\tilde{Z},\bar{U},|\mathcal{G}}(z). \quad (4.4.13)$$

STEP 2: We want to get rid of the conditioning on $\mathcal{G}$ to have non-conditional bound for $p_{\tilde{Z},U}$. Since $\tilde{U}$ is $\mathcal{G}$ measurable, for f measurable, non-negative, with support included in $B(0, r/2)$,

$$\mathbb{E}[f(\tilde{Z})U] = \mathbb{E}[\mathbb{E}[f(\tilde{Z})\bar{U}|\mathcal{G}]\tilde{U}]$$

and, by (4.4.6) and (4.4.13), we obtain

$$\frac{1}{C}\mathbb{E}[\tilde{U}] \int f(z)dz \leq \mathbb{E}[f(\tilde{Z})U]$$

$\Lambda_{\rho,\varepsilon} \subset \{\tilde{U} = 1\}$, (4.4.2) and $\varepsilon \in 1/\mathcal{C}$ imply $\mathbb{E}[\tilde{U}] \geq \frac{1}{C}$, so (4.4.5) is proved. □

4.4.2 Lower bound for the density

We use results and notations of chapter 2. In particular, we denote with D the Malliavin derivative with respect to W, the Brownian Motion driving equation (4.1.1). We recall (4.3.3) and $\Gamma_\delta = \begin{pmatrix} A_\delta \\ \underline{\Gamma}_\delta \end{pmatrix}$ and define

$$F = \tilde{Z} + \tilde{R}, \qquad \text{with } \tilde{R} = \Gamma_\delta^{-1}(R_\delta, 0_{m-n}). \tag{4.4.14}$$

Lemma 4.7. *There exist $C \in \mathcal{C}$, $\delta_*, r \in 1/\mathcal{C}$ such that for $\delta \le \delta_*$, $|z| \le r/2$,*

$$\frac{1}{C} \le p_{F,U}(z). \tag{4.4.15}$$

Proof. We want to apply Theorem 2.4 with $F = F$ and $G = \tilde{Z}$ defined in (4.3.15). Fix $p = 32n$. We are going to check that $C_1 \in \mathcal{C}$. Moreover, from (4.4.14),

$$\|\Delta_2(F - \tilde{Z})\|_p = \|\tilde{R}\|_{2,p} \le |\Gamma_\delta^{-1}| \|R\|_{2,p} \le C\delta^{-1}\delta^{3/2} = C'\sqrt{\delta}, \quad C' \in \mathcal{C}.$$

Choosing an appropriate $\delta \le \delta_* \in 1/\mathcal{C}$, since (4.4.5),

$$\frac{1}{C} \le p_{F,U}(z), \quad \text{for} \quad |z| \le r/2.$$

Let us check that the quantities involved in the definition of C_1 in Theorem 2.4 are in $\mathcal{C}$. For $n_{F,G,U}(1,p)$ this is elementary. We start with $m_U(k,p)$. Standard computations and (4.4.22) give $\forall p$

$$\|1/\det Q\|_{2,p} + \| \, |Q|_l \|_{2,p} + \|q(B)\|_{2,p} + \|\Theta\|_{2,p} \le C,$$

so we can apply (2.1.4) and conclude

$$m_U(1,p) \le C, \quad \|1 - U\|_{1,p} \le C, \quad \exists C \in \mathcal{C}. \tag{4.4.16}$$

Now (see (4.3.15)),

$$\begin{aligned} \langle \gamma_{\tilde{Z}}\xi, \xi\rangle &= \sum_{i=1}^d \int_0^\delta \langle D_s^i \tilde{Z}, \xi\rangle^2 = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i \tilde{Z}, \xi\rangle^2 = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i G, \xi\rangle^2 \\ &\ge \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \frac{1}{2}\langle D_s^i \Theta, \xi\rangle^2 - \langle D_s^i \eta(\Theta), \xi\rangle^2 ds \\ &= S_1 + S_2. \end{aligned}$$

We have

$$S_2 = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle \nabla\eta(\Theta) D_s^i \Theta, \xi\rangle^2 ds = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i \Theta, \nabla\eta(\Theta)^T \xi\rangle^2 ds \le \lambda^*(Q)|\nabla\eta(\Theta)|^2|\xi|^2$$

and $S_1 = \sum_{i=1}^{d} \int_{s_{i-1}(\delta)}^{s_i(\delta)} \frac{1}{2} \langle D_s^i \Theta, \xi \rangle^2 \geq \frac{\lambda_*(Q)}{2}$, so

$$\lambda_*(\gamma_{\tilde{Z}}) \geq \lambda_*(Q) \left(\frac{1}{2} - \frac{\lambda^*(Q)}{\lambda_*(Q)} |\nabla \eta(\Theta)|^2 \right).$$

On $\{\tilde{U} \neq 0\}$, $c_*(\eta, \Theta) \leq \frac{\sqrt{\lambda_*(Q)/\lambda^*(Q)}}{2m}$ (proved in lemma 4.6). Since $|\nabla \eta(\Theta)| \leq mc_*(\eta, \Theta)$,

$$|\nabla \eta(\Theta)| \leq \frac{1}{2} \sqrt{\frac{\lambda_*(Q)}{\lambda^*(Q)}},$$

and therefore $4\lambda_*(\gamma_{\tilde{Z}}) \geq \lambda_*(Q) \geq \varepsilon^{3m\rho}$, which implies $\Gamma_{\tilde{Z},U}(p) \leq 4\Gamma_{\Theta,U}(p) \leq C \in \mathcal{C}$. □

We now make the change of variable that gives us (4.4.1), proving that there exist $C \in \mathcal{C}$, $\delta_*, r_* \in 1/\mathcal{C}$ such that for $\delta \leq \delta_*$, $|y - X_0 - b\delta|_{A_\delta} \leq r_*$,

$$\frac{1}{C\delta^{n - \frac{dim\langle\sigma\rangle}{2}}} \leq p_{X_\delta}(y).$$

Proof. We take the same δ_*, r as in Lemma 4.7. Writing

$$\tilde{X} = \underline{\Gamma}_\delta \Theta, \quad \bar{X}_\delta = (X_\delta, \tilde{X}), \quad \bar{X}_0 = (X_0 + b\delta, 0_{m-n})$$

and recalling (4.3.1), (4.3.15), (4.4.14), we have

$$\bar{X}_\delta = \bar{X}_0 + \Gamma_\delta F. \tag{4.4.17}$$

We denote with $Pr : \mathbb{R}^m \to \mathbb{R}^n$ the projection of the first n components, and with $\underline{Pr} : \mathbb{R}^m \to \mathbb{R}^{m-n}$ the projection of the last $m-n$ components. From (4.4.17) we have

$$X_\delta = X_0 + b\delta + Pr\Gamma_\delta F, \qquad \tilde{X} = \underline{Pr}\Gamma_\delta F. \tag{4.4.18}$$

Writing (4.4.17) in coordinates as

$$y = (y_{(1)}, y_{(2)}) = (X_0 + b\delta + Pr\Gamma_\delta z, \underline{Pr}\Gamma_\delta z) = \bar{X}_0 + \Gamma_\delta z,$$

we want to recover a lower bound for $p_{X_\delta}(y_{(1)})$ from (4.4.15). From (4.3.10) follows

$$\{|X_\delta - X_0 - b\delta|_{A_\delta} \leq r/4\} \cap \{|\tilde{X}| \leq r/4\} \subset \{|\bar{X}_\delta - \bar{X}_0|_{\Gamma_\delta} \leq r/2\} = \{|F| \leq r/2\} \tag{4.4.19}$$

We consider a positive, measurable function $f : \mathbb{R}^n \to \mathbb{R}$ with support included in

$$\{y_{(1)} : \ |y_{(1)} - X_0 + b\delta|_{A_\delta} \leq r/4\} \tag{4.4.20}$$

Applying (4.4.18), (4.4.19) and (4.4.15) we obtain

$$\begin{aligned}\mathbb{E}[f(X_\delta)] &\geq \mathbb{E}_U[f(X_\delta)] = \mathbb{E}_U[f(X_0 + b\delta + Pr\Gamma_\delta F)] \\ &= \int f(X_0 + b\delta + Pr\Gamma_\delta z) p_{F,U}(z) dz \\ &\geq \frac{1}{C} \int_{\{|z| \leq r/2\}} f(X_0 + b\delta + Pr\Gamma_\delta z) dz \\ &\geq \frac{1}{C|\det \Gamma_\delta|} \int_{\{|y_{(1)} - X_0 - b\delta|_{A_\delta} \leq r/4\} \cap \{|y_{(2)}| \leq r/4\}} f(y_{(1)}) dy_{(1)} dy_{(2)} \\ &= \frac{1}{C\sqrt{|\det A_\delta A_\delta^T|}} \left(\frac{r}{4}\right)^{m-n} \int_{\{|y_{(1)} - X_0 - b(0,X_0)\delta|_{A_\delta} \leq r/4\}} f(y_1) dy_1.\end{aligned}$$

From (4.2.3) and Cauchy-Binet formula we obtain

$$\frac{1}{C} \delta^{n - \frac{\dim\langle\sigma\rangle}{2}} \leq \sqrt{|\det A_\delta A_\delta^T|} \leq C\delta^{n - \frac{\dim\langle\sigma\rangle}{2}} \tag{4.4.21}$$

Since $r \in 1/\mathcal{C}$, this gives that for some $C \in \mathcal{C}$, $r_* \in 1/\mathcal{C}$, for $|y - X_0 + b(0, X_0)\delta|_{A_\delta} \leq r_*$

$$\frac{1}{C\delta^{n - \frac{\dim\langle\sigma\rangle}{2}}} \leq p_{X_\delta}(y).$$

□

4.4.3 Support Property

In this section we prove ((4.4.25)). Let $B = (B^1, ..., B^{d-1})$ be a standard Brownian motion. We consider the analogues of the covariance matrix $Q_i(B)$ considered in the previous sections: we define a symmetric square matrix of dimension $d \times d$ by

$$\begin{aligned} Q^{d,d} &= 1, \quad Q^{d,j} = Q^{j,d} = \int_0^1 B_s^j ds, \quad j = 1, ..., d-1, \\ Q^{j,p} &= Q^{p,j} = \int_0^1 B_s^j B_s^p ds, \quad j, p = 1, ..., d-1 \end{aligned}$$

and we denote by $\lambda_*(Q)$ (respectively by $\lambda^*(Q)$) the lower (respectively larger) eigenvalue of Q.

For a measurable function $g : [0,1] \to R^{d-1}$ we denote

$$\begin{aligned} \alpha_g(\xi) &= \xi_d + \int_0^1 \langle g_s, \xi_* \rangle ds, \quad \beta_g(\xi) = \int_0^1 \langle g_s, \xi_* \rangle^2 ds - \left(\int_0^1 \langle g_s, \xi_* \rangle ds \right)^2 \quad \textit{with} \\ \xi &= (\xi_1, ..., \xi_d) \in R^d \quad \textit{and} \quad \xi_* = (\xi_1, ..., \xi_{d-1}). \end{aligned}$$

We need the following two preliminary lemmas.

Lemma 4.8. *With $g(s) = B_s, s \in [0,1]$ we have*

$$\langle Q\xi, \xi\rangle = \alpha_B^2(\xi) + \beta_B(\xi).$$

As a consequence, one has

$$\lambda_*(Q) = \inf_{|\xi|=1}(\alpha_B^2(\xi) + \beta_B(\xi)) \quad \text{and} \quad \lambda^*(Q) \le \sup_{|\xi|=1}(\alpha_B^2(\xi) + \beta_B(\xi)) \le \big(1 + \sup_{t\le 1}|B_t|\big)^2.$$

Taking $\xi_* = 0$ and $\xi_d = 1$ we obtain $\langle Q\xi, \xi\rangle = 1$ so that $\lambda_*(Q) \le 1 \le \lambda^*(Q)$.

Proof. By direct computation

$$\begin{aligned}\langle Q\xi, \xi\rangle &= \xi_d^2 + 2\xi_d \int_0^1 \langle B_s, \xi_*\rangle\, ds + \Big(\int_0^1 \langle B_s, \xi_*\rangle\, ds\Big)^2 \\ &\quad + \int_0^1 \langle B_s, \xi_*\rangle^2\, ds - \Big(\int_0^1 \langle B_s, \xi_*\rangle\, ds\Big)^2 \\ &= \Big(\xi_d + \int_0^1 \langle B_s, \xi_*\rangle\, ds\Big)^2 + \int_0^1 \langle B_s, \xi_*\rangle^2\, ds - \Big(\int_0^1 \langle B_s, \xi_*\rangle\, ds\Big)^2.\end{aligned}$$

The remaining statements follow straightforwardly. □

Proposition 4.9. *For each $p \ge 1$ one has*

$$E(|\det Q|^{-p}) \le C_{p,d} < \infty \tag{4.4.22}$$

where $C_{p,d}$ is a constant depending on p, d only.

Proof. By Lemma 7-29, pg 92 in [20], for every $p \in (0,\infty)$ one has

$$\frac{1}{|\det Q|^p} \le \frac{1}{\Gamma(p)} \int_{R^d} |\xi|^{d(2p-1)}\, e^{-\langle Q\xi,\xi\rangle} d\xi.$$

Let $\theta(\xi_*) := \int_0^1 \langle B_s, \xi_*\rangle\, ds$. Using the previous lemma

$$\begin{aligned}\int_{R^d} |\xi|^{d(2p-1)}\, e^{-\langle Q\xi,\xi\rangle} d\xi &= \int_{R^d} (\xi_d^2 + |\xi_*|^2)^{d(2p-1)/2} e^{-(\xi_d+\theta(\xi_*))^2 - \beta_B(\xi_*)} d\xi \\ &\le C \int_{R^{d-1}} ((1+\theta^2(\xi_*))^{d(2p-1)/2} + |\xi_*|^{d(2p-1)}) e^{-\beta_B(\xi_*)} d\xi_* \\ &\le C \int_{R^{d-1}} \sup_{t\le 1} 1 \vee |B_t|^{d(2p-1)}\, (1 + |\xi_*|^{d(2p-1)+1}) e^{-\beta_B(\xi_*)} d\xi_*.\end{aligned}$$

We integrate and we use Schwartz inequality in order to obtain

$$E\Big(\frac{1}{|\det Q|^p}\Big) \le C + C \int_{\{|\xi_*|\ge 1\}} (E((1 + |\xi_*|^{d(2p-1)+1})^2 e^{-2\beta_B(\xi_*)}))^{1/2} d\xi_*.$$

For each fixed ξ_* the process $b_{\xi_*}(t) := |\xi_*|^{-1} \langle B_t, \xi_* \rangle$ is a standard Brownian motion and $\beta_B(\xi_*) = |\xi_*|^2 \int_0^1 (b_{\xi_*}(t) - \int_0^1 b_{\xi_*}(s) ds)^2 dt =: |\xi_*|^2 V_{\xi_*}$ where V_{ξ_*} is the variance of b_{ξ_*} with respect to the time. Then it is proved in [46] (see (1.f), p. 183) that

$$E(e^{-2\beta_B(\xi_*)}) = E(e^{-2|\xi_*|^2 V_{\xi_*}}) = \frac{2|\xi_*|^2}{\sinh 2|\xi_*|^2}.$$

We insert this in the previous inequality and we obtain $E(|\det Q|^{-p}) < \infty$. □

We are now able to give the main result in this section. We define

$$q(B) = \sum_{i=1}^{d-1} |B_1^i| + \sum_{j \neq p} \left| \int_0^1 B_s^j dB_s^p \right| \tag{4.4.23}$$

and for $\varepsilon, \rho > 0$ we denote

$$\Lambda_{\rho,\varepsilon}(B) = \{\det Q \geq \varepsilon^\rho, \sup_{t \leq 1} |B_t| \leq \varepsilon^{-\rho}, q(B) \leq \varepsilon\}. \tag{4.4.24}$$

Proposition 4.10. *There exist some universal constants $c_{\rho,d}, \varepsilon_{\rho,d} \in (0,1)$ (depending on ρ and d only) such that for every $\varepsilon \in (0, \varepsilon_{\rho,d})$ one has*

$$P(\Lambda_{\rho,\varepsilon}(B)) \geq c_{\rho,d} \times \varepsilon^{\frac{1}{2}d(d+1)}. \tag{4.4.25}$$

Proof. Using the previous proposition and Chebyshev's inequality we get

$$P(\det Q < \varepsilon^\rho) \leq \varepsilon^{p\rho} E |\det Q|^{-p} \leq C_{p,d} \varepsilon^{p\rho} \quad \text{and} \quad P(\sup_{t \leq 1} |B_t| > \varepsilon^{-\rho}) \leq \exp(-\frac{1}{C\varepsilon^{2\rho}}).$$

Let $q'(B) = \sum_{i=1}^{d-1} |B_1^i| + \sum_{j<p} \left| \int_0^1 B_s^j dB_s^p \right|$. Since $\left| \int_0^1 B_s^j dB_s^p \right| \leq |B_1^j| \, |B_1^p| + \left| \int_0^1 B_s^p dB_s^j \right|$ we have $q(B) \leq 2q'(B) + q'(B)^2$ so that $\{q'(B) \leq \frac{1}{3}\varepsilon\} \subset \{q(B) \leq \varepsilon\}$. We will now use the following fact: consider the diffusion process $X = (X^i, X^{j,p}, \ i = 1, ..., d, 1 \leq j < p \leq d)$ solution of the equation $dX_t^i = dB_t^i, dX_t^{j,p} = X_t^j dB_t^p$. The strong Hörmander condition holds for this process and the support of the law of X_1 is the whole space. So the law of X_1 is absolutely continuous with respect to the Lebesgue measure and has a continuous and strictly positive density p. This result is well known (see for example [72] or [7]). We denote $c_d := \inf_{|x| \leq 1} p(x) > 0$ and this is a constant which depends on d only. Then, by observing that $q'(B) \leq \sqrt{m} |X_1|$, where $m = \frac{1}{2}d(d+1)$ is the dimension of the diffusion X, we get

$$P(q(B) \leq \varepsilon) \geq P\Big(q'(B) \leq \frac{\varepsilon}{3}\Big) \geq P\Big(|X_1| \leq \frac{\varepsilon}{3\sqrt{m}}\Big) \geq \frac{\varepsilon^m}{(3\sqrt{m})^m} \times \bar{c}_d,$$

with $\bar{c}_d > 0$. So finally we obtain

$$P(\Lambda_{\rho,\varepsilon}(B)) \geq \bar{c}_d \varepsilon^{\frac{1}{2}d(d+1)} - C_{p,d} \varepsilon^{p\rho} - \exp(-\frac{1}{C\varepsilon^{2\rho}}).$$

Choosing $p > \frac{1}{2\rho} d(d+1)$ and ε small we obtain our inequality. □

4.5 Upper bound for the density in short time

4.5.1 Upper bound for the density

We prove here the upper bound of theorem (4.5).

Theorem 4.5 . *(Upper bound). There exist $C \in \mathcal{C}$, $\delta_* \in 1/\mathcal{C}$ such that for $\delta \le \delta_*$,*

$$p_{X_\delta}(y) \le \frac{C}{\delta^{n-\frac{dim\langle\sigma\rangle}{2}}}.$$

Proof. As for the lower bound, we rescale X_δ, but using a different matrix α that we are now going to define.

We consider the matrix $A_\delta(t,x) = A_\delta$, for fixed $\delta > 0$, and take a singular value decomposition (SVD): $A_\delta = U\Sigma V^T$, where U is a $n \times n$ orthogonal matrix and V^T denotes the transpose of the $m \times m$ orthogonal matrix V. Σ is a $n \times m$ matrix, $\Sigma = [\bar{\Sigma}\ 0]$ where $\bar{\Sigma}$ is a $n \times n$ diagonal matrix with positive real numbers on the diagonal (since A_δ has full row rank). We define $\alpha = U\bar{\Sigma}$. Note that α is a $n \times n$ matrix. We also define the change of variable

$$T_\alpha : \mathbb{R}^n \to \mathbb{R}^n, \qquad T_\alpha(y) = \alpha^{-1}y,$$

and its adjoint $T^*_\alpha(y) = \alpha^{-1,T}y$.
Properties:

$$|y|_{A_\delta} = |T_\alpha y| = |y|_\alpha, \quad \forall y \in \mathbb{R}^m, \quad \text{and } \det\alpha = \sqrt{A_\delta A_\delta^T} \tag{4.5.1}$$

$$\forall v \in \mathbb{R}^n \text{ with } |v| = 1, \quad \exists j = 1,\dots,m : \quad |T^*_\alpha v \cdot A^j_\delta| \ge \frac{1}{m} \tag{4.5.2}$$

$$\forall j = 1,\dots,d, \quad \sqrt{\delta}|T_\alpha \sigma_j| \le C \tag{4.5.3}$$

Proof. 113 (4.5.1) follows easily from $\alpha = U\bar{\Sigma}$ and the definition of $|\cdot|_M$. Now, $(T^*_\alpha v)^T A_\delta = v^T\alpha^{-1}A_\delta = [v^T 0]V^T$. So $(T^*_\alpha v)^T A_\delta = 1$, and therefore $\exists j = 1,\dots,m$: $|T^*_\alpha v \cdot A^j_\delta| \ge \frac{1}{m}$, so (4.5.2) is proved. Moreover, $T_\alpha A_\delta = [Id_n\, 0]V^T$ and so $\forall i = 1,\dots,m$, $|T_\alpha A^i_\delta| \le 1$. For $A^i_\delta = \sigma_j\sqrt{\delta}$ we have (4.5.3). □

We define now

$$F = \alpha^{-1}(X_\delta - X_0) = T_\alpha(X_\delta - X_0)$$

We use now some estimates from chapter 2. From Hölder inequality applied to representation 2.1.2 with $p = 2n$

$$p_F(z) \le \mathbb{E}[\nabla Q_n(F-z)H(F,1)] \le \|H(F,1)\|_{2n}\|Q_n(F-z)\|_{2n/(2n-1)},$$

and since $\|Q_n(F-z)\|_{2n/(2n-1)} \le \|H(F,1)\|_{2n}^{4n(n-1)}$ (proved in [5], see equation (2.2.6) in this work), we have

$$p_F(z) \le C\|H(F,1)\|_{2n}^{4n^2}.$$

Equation (2.2.2) with $V = U = 1$, $k = 0$, $p = 2n$, $p_3 = p_4 = 10n$ gives $\|H(F,1)\|_{2n} \leq C\Gamma_F(10n)^2 n_F(0,10n)^3$. In next section, lemma 4.12 we prove $\Gamma_F(10n) \leq C \in \mathcal{C}$. $n_F(0,10n) \leq C \in \mathcal{C}$ from standard computation. We prove just $\|F\|_p \leq C$, $\forall p$, for the Malliavin derivatives the proof is heavier but analogous.

$$\begin{aligned} F &= T_\alpha \left(\sum_{j=1}^{d} \int_0^\delta \sigma_j(t, X_t) \circ dW_t^j + \int_0^\delta b(t, X_t)dt \right) \\ &= T_\alpha \left(\sum_{j=1}^{d} \sigma_j(0, X_0) W_\delta^j + \sum_{j=1}^{d} \int_0^\delta \sigma_j(t, X_t) - \sigma_j(0, X_0) \circ dW_t^j + \int_0^\delta b(t, X_t)dt \right) \\ &= T_\alpha \left(\sum_{j=1}^{d} \sigma_j(0, X_0) W_\delta^j + B_\delta \right). \end{aligned}$$

Therefore

$$\begin{aligned} |F| &= \left| T_\alpha \left(\sum_{j=1}^{d} \sigma_j(0, X_0) W_\delta^j + B_\delta \right) \right| \\ &\leq \sum_{j=1}^{d} |T_\alpha \sigma_j(0, X_0) W_\delta^j| + |T_\alpha B_\delta|. \end{aligned}$$

(4.5.3) implies $\mathbb{E}|T_\alpha \sigma_j(0, X_0) W_\delta^j|^p \leq C$, $j = 1, \ldots, d$. $|T_\alpha B_\delta| \leq |B_\delta|_{A_\delta} \leq |B_\delta|/\delta$, and since B_δ is of order δ we conclude that $\mathbb{E}|F|^p \leq C \in \mathcal{C}$.

$\Gamma_F(p)$, $\forall p$ is bounded uniformly in $\delta \downarrow 0$, as we prove in the following section, and we conclude that $p_F(z) \leq C \in \mathcal{C}$. The upper bound for the density of X_δ comes from a simple change of variable. For a positive, measurable $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathbb{E}f(X_\delta) = \mathbb{E}f(X_0 + \alpha F) = \int f(X_0 + \alpha z) p_F(z) dz \leq C \int f(X_0 + \alpha z) dz \leq \frac{C}{|\det(\alpha)|} \int f(y) dy.$$

so $p_{X_\delta}(y) \leq \frac{C}{|\det \alpha|} = \frac{C}{\sqrt{A_\delta A_\delta^T}} \leq \frac{C}{\delta^{n - \frac{dim\langle\sigma\rangle}{2}}}$ (from (4.4.21)). □

4.5.2 Covariance Matrix of the rescaled diffusion

Recall $\Gamma_F(p) = 1 + (\mathbb{E}|\lambda_*(\gamma_F)|^{-p})^{1/p}$. We need to prove that this quantity is bounded uniformly in $\delta \downarrow 0$, (in particular for $p = 10n$), to be able to prove the upper bound for p_F (see Lemma 4.7). We need this preliminary result.

Lemma 4.11. *Take a_t, b_t, $t \in [0,T]$ stochastic processes a.s. increasing, $b_0 = 0$, p fixed and $\mathbb{E}[b_t^p] \leq C_p t^{(k+1)p}$. Suppose*

$$a_t \geq \frac{t^k - b_t}{\delta^k}.$$

Fix $\delta \leq \frac{1}{2C_p^{1/p}}$. For $t \leq \delta$, $\varepsilon \leq 1$,

$$\mathbb{P}(a_\delta \leq \varepsilon) \leq \varepsilon^p.$$

Proof. Set

$$S_\varepsilon = \inf\left\{ s \geq 0 : b_s \geq \frac{(\delta\varepsilon)^k}{2} \right\} \wedge \delta,$$

Remark that for any $p > 0$

$$\begin{aligned}
\mathbb{P}(S_\varepsilon < \delta\varepsilon) &= \mathbb{P}\left(b^p_{\delta\varepsilon} \geq \left(\frac{(\delta\varepsilon)^k}{2}\right)^p \right) \\
&\leq 2^p \frac{\mathbb{E} b^p_{\delta\varepsilon}}{(\delta\varepsilon)^{kp}} \\
&\leq 2^p C_p (\delta\varepsilon)^p \\
&\leq \varepsilon^p.
\end{aligned}$$

On the other hand, on $S_\varepsilon \geq \delta\varepsilon$,

$$a_{S_\varepsilon} \geq a_{\delta\varepsilon} \geq \frac{(\delta\varepsilon)^k - (\delta\varepsilon)^k/2}{\delta^k} \geq \varepsilon^k/2.$$

Therefore for any p, for any $\varepsilon \leq 1$,

$$\mathbb{P}(a_\delta < \varepsilon^k/2) \leq \mathbb{P}(a_{S_\varepsilon} < \varepsilon^k/2) \leq \mathbb{P}(S_\varepsilon < \delta\varepsilon) \leq \varepsilon^p.$$

This implies that for for any $p \geq 1$, for $\delta \leq \delta_p$, $\varepsilon \leq 1$,

$$\mathbb{P}(a_\delta < \varepsilon) \leq \varepsilon^p.$$

□

Lemma 4.12. *For any $q > 0$ there exists $\delta_* \in 1/\mathcal{C}$ such that, for $\delta \leq \delta_* \in 1/\mathcal{C}$, $\Gamma_F(q) \leq C \in \mathcal{C}$.*

Proof. We need a bound for the moments of the inverse of

$$\gamma_F = \sum_{k=1}^d \int_0^\delta D^k_s F D^k_s F^T ds.$$

Following [79] we define the tangent flow of X as the derivative with respect to the initial condition of X, $Y_t := \partial_x X_t$. We also denote its inverse $Z_t = Y_t^{-1}$. They satisfy the following stochastic differential equations (remark that the equations we consider for X, Y and Z are all in Stratonovich form):

$$\begin{aligned}
Y_t &= Id + \sum_{k=1}^d \int_0^t \nabla\sigma_k(s, X_s) Y_s \circ dW^k_s + \int_0^t \nabla b(s, X_s) Y_s ds \\
Z_t &= Id - \sum_{k=1}^d \int_0^t Z_s \nabla\sigma_k(s, X_s) \circ dW^k_s - \int_0^t Z_s \nabla b(s, X_s) ds,
\end{aligned}$$

where $\nabla\sigma_k$ and ∇b are the Jacobian matrix with respect to the space variable. It holds

$$D_s X_\delta = Y_\delta Z_s \sigma(s, X_s).$$

Applying Ito's formula we have the following representation for $\phi \in C^2$:

$$\begin{aligned} Z_t\phi(t, X_t) &= \phi(0, x) + \int_0^t Z_s \sum_{k=1}^d [\sigma_k, \phi](s, X_s) dW_s^k \\ &+ \int_0^t Z_s \left\{ [b, \phi] + \frac{1}{2} \sum_{k=1}^d [\sigma_k, [\sigma_k, \phi]] + \frac{d\phi}{ds} \right\} (s, X_s)\, ds. \end{aligned} \tag{4.5.4}$$

(Details in [79] for the autonomous case. Remark that here we are taking an Ito integral). We now compute

$$D_s F = \alpha^{-1} D_s X_\delta = \alpha^{-1} Y_\delta Z_s \sigma(s, X_s) = \alpha^{-1} Y_\delta \alpha \alpha^{-1} Z_s \sigma(s, X_s)$$

so

$$\gamma_F = \alpha^{-1} Y_\delta \alpha\, \bar{\gamma}_F\, (\alpha^{-1} Y_\delta \alpha)^T \quad \text{where} \quad \bar{\gamma}_F = \alpha^{-1} \int_0^\delta Z_s \sigma(s, X_s) \sigma(s, X_s)^T Z_s^T ds\, \alpha^{-1,T},$$

and

$$\hat{\gamma}_F = (\alpha^{-1} Y_\delta \alpha)^{-1,T}\, \bar{\gamma}_F^{-1}\, (\alpha^{-1} Y_\delta \alpha)^{-1}$$

Since

$$(\alpha^{-1} Y_\delta \alpha)^{-1} = \alpha^{-1} Z_\delta \alpha = Id_d + \alpha^{-1}(Z_\delta - Id_d)\alpha$$

standard computations give

$$\mathbb{E}\lambda^* \left((\alpha^{-1} Y_\delta \alpha)^{-1} \right)^q \le C \in \mathcal{C}, \quad \forall q.$$

We now need to estimate the reduced matrix, i.e. prove

$$\mathbb{E}\lambda_*(\bar{\gamma}_F)^{-q} \le C \in \mathcal{C}, \quad \forall q. \tag{4.5.5}$$

We recall lemma 3.9, which is a slight modification of Lemma 2.3.1. in [79], and we use it again in this proof.

Lemma: *Let γ be a symmetric nonnegative definite $n \times n$ matrix. Denoting $|C| = \sum_{1 \le i,j \le n} |\gamma^{i,j}|^2)^{1/2}$, we assume that $\mathbb{E}|C|^{p+1} < \infty$, and that for $\varepsilon \le \varepsilon_{p+2n}$,*

$$\sup_{|\xi|=1} \mathbb{P}[\langle \gamma\xi, \xi \rangle < \varepsilon] \le \varepsilon^{p+2n}$$

Then

$$\mathbb{E}\lambda_*(\gamma)^{-p} \le C\varepsilon_{p+2n}^{-p}.$$

We show now that for any $p > 0$, $\sup_{|v|=1} \mathbb{P}\left(\langle \bar{\gamma}_F v, v \rangle \le \varepsilon\right) \le \varepsilon^p$, for $\delta \le \delta_p$ for $\varepsilon \le \varepsilon_p \in 1/\mathcal{C}$ not depending on δ. Together with the previous lemma this implies (4.5.5).

Denote $\xi = T_\alpha^* v$. From (4.5.2) and the definition (4.2.3) of A_δ we have two possible cases: A) $|\xi \cdot \sigma_j(t,x)| \geq \frac{1}{m\delta^{1/2}}$ for some $j = 1\dots d$, or B) $|\xi \cdot [\sigma_j, \sigma_l](t,x)| \geq \frac{1}{m\delta}$ for some $j, l = 1\dots d$, $j \neq l$. Moreover

$$\alpha \bar{\gamma}_F \alpha^T = \int_0^\delta Z_s \sigma(s, X_s)\sigma(s, X_s)^T Z_s^T ds. \tag{4.5.6}$$

Therefore, with $\xi = T_\alpha^* v$, we have for any $q > 1$

$$\mathbb{P}(\langle \bar{\gamma}_F v, v\rangle \leq \varepsilon^q) = \mathbb{P}\left(\xi^T \int_0^\delta Z_s \sigma(s, X_s)\sigma(s, X_s)^T Z_s^T ds\, \xi \leq \varepsilon^q\right)$$

We decompose this probability:

$$\begin{aligned}
\mathbb{P}(\langle \bar{\gamma}_F v, v\rangle \leq \varepsilon^q) &= \mathbb{P}\left(\sum_{i=1}^d \int_0^\delta |\xi^T Z_t \sigma_i(t, X_t)|^2 dt \leq \varepsilon^q\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^d \int_0^\delta |\xi^T Z_t \sigma_i(t, X_t)|^2 dt \leq \varepsilon^q \text{ and } \sum_{i,k=1}^d \int_0^\delta |\xi^T Z_t [\sigma_i, \sigma_k](t, X_t)|^2 dt \leq \frac{\varepsilon}{\delta}\right) \\
&+ \mathbb{P}\left(\sum_{i=1}^d \int_0^\delta |\xi^T Z_t \sigma_i(t, X_t)|^2 dt \leq \varepsilon^q \text{ and } \sum_{i,k=1}^d \int_0^\delta |\xi^T Z_t [\sigma_i, \sigma_k](t, X_t)|^2 dt > \frac{\varepsilon}{\delta}\right) \\
&=: I_1 + I_2
\end{aligned}$$

To estimate I_1 we distinguish two cases:
Case A): $|\xi \cdot \sigma_j(t,x)| \geq \frac{1}{m\delta^{1/2}}$ for some $j = 1\dots d$. We fix this j. For $\bar{t} \leq \delta$,

$$\begin{aligned}
I_1 &\leq \mathbb{P}\left(\int_0^\delta |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \leq \varepsilon^q \text{ and } \int_0^\delta |\xi^T \sum_{k=1}^d Z_t [\sigma_k, \sigma_j](t, X_t)|^2 dt < \frac{\varepsilon}{\delta}\right) \\
&\leq \mathbb{P}\left(\int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \leq \varepsilon^q \text{ and } \int_0^{\bar{t}} |\int_0^t \xi^T \sum_{k=1}^d Z_s [\sigma_k, \sigma_j](s, X_s) dW_s^k|^2 dt < \frac{\bar{t}}{12m^2\delta}\right) \\
&+ \mathbb{P}\left\{\int_0^{\bar{t}} |\int_0^t \xi^T \sum_{k=1}^d Z_s [\sigma_k, \sigma_j](s, X_s) dW_s^k|^2 dt \geq \frac{\bar{t}}{12m^2\delta} \text{ and } \int_0^{\bar{t}} |\xi^T \sum_{k=1}^d Z_t [\sigma_k, \sigma_j](t, X_t)|^2 dt < \frac{\varepsilon}{\delta}\right\}
\end{aligned}$$

Now remark that from the exponential martingale inequality, setting $u_s = (\xi^T Z_s [\sigma_k, \sigma_j](s, X_s))_{k=1,\dots,d}$,

$$\begin{aligned}
&\mathbb{P}\left\{\int_0^{\bar{t}} |\sum_{k=1}^d \int_0^t u_s^k dW_s^k|^2 dt \geq \frac{\bar{t}}{12m^2\delta}, \quad \int_0^{\bar{t}} |u_t|^2 dt < \frac{\varepsilon}{\delta}\right\} \\
&\leq \mathbb{P}\left\{\sup_{0\leq t\leq \bar{t}} |\sum_{k=1}^d \int_0^t u_s^k dW_s^k|^2 \geq \frac{1}{\bar{t}} \frac{\bar{t}}{12m^2\delta}, \quad \int_0^{\bar{t}} |u_t|^2 dt < \frac{\varepsilon}{\delta}\right\} \\
&\leq 2\exp\left(-\frac{1}{12m^2\delta}\frac{\delta}{2\varepsilon}\right) = 2\exp\left(-\frac{1}{24m^2\varepsilon}\right).
\end{aligned} \tag{4.5.7}$$

So $\forall p > 1$,

$$\mathbb{P}\{\int_0^{\bar{t}} |\int_0^t \xi^T \sum_{k=1}^d Z_s[\sigma_k, \sigma_j](s, X_s)dW_s^k|^2 dt \geq \frac{\bar{t}}{12m^2\delta}$$
$$\text{and } \int_0^{\bar{t}} |\xi^T \sum_{k=1}^d Z_t[\sigma_k, \sigma_j](t, X_t)|^2 dt < \frac{\varepsilon}{\delta}\} < \varepsilon^p.$$

Now we define $D := \left\{\int_0^{\bar{t}} | \int_0^t \xi^T \sum_{k=1}^\delta Z_s[\sigma_k, \sigma_j](s, X_s)dW_s^k|^2 dt < \frac{\bar{t}}{12m^2\delta}\right\}$ and prove

$$\mathbb{P}\left(\left\{\int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \leq \frac{\varepsilon^q}{4m^2}\right\} \cap D\right) \leq \varepsilon^p$$

(equivalent to the desired estimate $\mathbb{P}\left(\left\{\int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \leq \varepsilon^q\right\} \cap D\right) \leq \varepsilon^p$). From representation (4.5.4), for $\phi = \sigma_j$ we find

$$Z_t \sigma_j(t, X_t) = \sigma_j(0, X_0) + \int_0^t \sum_{k=1}^d Z_s[\sigma_k, \sigma_j](s, X_s)dW_s^k + R_t,$$

with

$$R_t = \int_0^t Z_s \left\{[b, \sigma_j] + \frac{1}{2}\sum_{k=1}^d [\sigma_k, [\sigma_k, \sigma_j]] + \frac{d\phi}{ds}\right\}(s, X_s)\, ds.$$

From $(a + b + c)^2 \geq a^2/3 - b^2 - c^2$ and $|\xi \cdot \sigma_j(x)| \geq \frac{1}{m\delta^{1/2}}$

$$\begin{aligned}
&\int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \\
&\quad \geq \frac{\bar{t}|\xi^T \sigma_j(0, X_0)|^2}{3} - \int_0^{\bar{t}} |\sum_{k=1}^d \int_0^t \xi^T Z_s[\sigma_k, \sigma_j](s, X_s)dW_s^k|^2 dt - \int_0^{\bar{t}} |\xi^T R_t|^2 dt \\
&\quad \geq \frac{\bar{t}}{3\delta m^2} - \int_0^{\bar{t}} |\sum_{k=1}^d \int_0^t \xi^T Z_s[\sigma_k, \sigma_j](s, X_s)dW_s^k|^2 dt - \int_0^{\bar{t}} |\xi^T R_t|^2 dt
\end{aligned}$$

and on D

$$\int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \geq \frac{\bar{t}}{4m^2\delta} - \int_0^{\bar{t}} |\xi^T R_t|^2 dt,$$

so

$$4m^2 \int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \geq \frac{\bar{t} - 4m^2\delta \int_0^{\bar{t}} |\xi^T R_t|^2 dt}{\delta}.$$

We set

$$a_{\bar{t}} = 4m^2 \int_0^{\bar{t}} |\xi^T Z_t \sigma_j(t, X_t)|^2 dt \,\vee\, \frac{\bar{t} - 4m^2\delta \int_0^{\bar{t}} |\xi^T R_t|^2 dt}{\delta},$$

and we have

$$\left\{\int_0^{\bar t} |\xi^T Z_t \sigma_j(t,X_t)|^2 dt \le \frac{\varepsilon^q}{4m^2}\right\} \cap D = \{a_{\bar t} \le \varepsilon^q\} \cap D$$

Standard computations, considering also $|\xi| = |T_\alpha^* v| \le |v|C/\delta = C/\delta$, give $\mathbb{E}(\int_0^{\bar t} |\xi^T R_t|^2 dt)^q \le C\bar t^{3q}/\delta^{2q}$, so $\mathbb{E}(4\delta m^2 \int_0^{\bar t} |\xi^T R_t|^2 dt)^q \le C\bar t^{2q}$. Setting

$$b_{\bar t} = 4\delta m^2 \int_0^{\bar t} |\xi^T R_t|^2 dt,$$

we can apply lemma 4.11, and we have

$$\mathbb{P}\left(\left\{\int_0^{\bar t} |\xi^T Z_t \sigma_j(t,X_t)|^2 dt \le \frac{\varepsilon^q}{4m^2}\right\} \cap D\right) = \mathbb{P}(\{a_{\bar t} \le \varepsilon^q\} \cap D) \le \mathbb{P}(a_{\bar t} \le \varepsilon^q) \le \varepsilon^p.$$

We obtain $I_1 < \varepsilon^p$ for any $p > 1$, for $\delta \le \delta_p$.
Case B) $|\xi \cdot [\sigma_j, \sigma_l](t,x)| \ge \frac{1}{m\delta}$ for some $j, l = 1 \dots d$, $j \ne l$.

$$I_1 \le \mathbb{P}\left(\int_0^{\bar t} |\xi^T Z_t [\sigma_j,\sigma_l](t,X_t)|^2 dt \le \frac{\varepsilon}{\delta}\right)$$

From representation (4.5.4) with $\phi = [\sigma_j, \sigma_l]$ we find

$$Z_t[\sigma_j,\sigma_l](t,X_t) = [\sigma_j,\sigma_l](0,X_0) + R_t,$$

with

$$\begin{aligned} R_t = &\int_0^t Z_s \sum_{k=1}^d [\sigma_k,[\sigma_j,\sigma_l]](s,X_s) dW_s^k \\ &+ \int_0^t Z_s \left\{[b,[\sigma_j,\sigma_l]] + \frac{1}{2}\sum_{k=1}^d [\sigma_k,[\sigma_k,[\sigma_j,\sigma_l]]] + \frac{d\phi}{ds}\right\}(s,X_s)\, ds. \end{aligned} \tag{4.5.8}$$

From $(a+b)^2 \ge a^2/2 - b^2$ and $|\xi \cdot [\sigma_j,\sigma_l](x)| \ge \frac{1}{m\delta}$

$$\int_0^{\bar t} |\xi^T Z_t[\sigma_j,\sigma_l](t,X_t)|^2 dt \ge \frac{\bar t |\xi^T[\sigma_j,\sigma_l](0,X_0)|^2}{2} - \int_0^{\bar t} |\xi^T R_t|^2 dt = \frac{\bar t}{2\delta^2 m^2} - \int_0^{\bar t} |\xi^T R_t|^2 dt.$$

Since

$$\mathbb{P}\left(\int_0^{\bar t} |\xi^T Z_s[\sigma_j,\sigma_l](s,X_s)|^2 ds \le \frac{\varepsilon}{\delta}\right) \le \mathbb{P}\left(\delta \int_0^{\bar t} |\xi^T Z_s[\sigma_j,\sigma_l](s,X_s)|^2 ds \le \varepsilon\right),$$

and

$$2m^2\delta \int_0^{\bar t} |\xi^T Z_s[\sigma_j,\sigma_l](s,X_s)|^2 ds = \frac{\bar t - 2m^2\delta^2 \int_0^{\bar t} |\xi^T R_t|^2 dt}{\delta}.$$

We apply lemma 4.11 with $b_{\bar{t}} = 2m^2\delta^2 \int_0^{\bar{t}} |\xi^T R_t|^2 dt$. Indeed from (4.5.8) and $|\xi| \leq C/\delta$,

$$\mathbb{E}|b_{\bar{t}}|^q \leq C\bar{t}^2.$$

So we find $I_1 < \varepsilon^p$, for $\delta \leq \delta_p$, $\varepsilon \leq \varepsilon_p$ uniform in δ. We estimate now

$$I_2 = \mathbb{P}\left(\sum_{i=1}^d \int_0^\delta |\xi^T Z_s \sigma_i(s, X_s)|^2 ds \leq \varepsilon \text{ and } \sum_{i,j=1}^d \int_0^\delta |\xi^T Z_s[\sigma_i, \sigma_j](s, X_s)|^2 ds > \frac{\varepsilon}{\delta}\right).$$

We use again (4.5.4) and find

$$\begin{aligned} \xi^T Z_t \sigma_i(t, X_t) &= \sum_{j=1}^n \int_0^t \xi^T Z_s[\sigma_i, \sigma_j](s, X_s) dW_s^j \\ &+ \int_0^t \xi^T Z_s \left\{[b, \sigma_i] + \frac{1}{2}\sum_{j=1}^n [\sigma_j, [\sigma_j, \sigma_i]] + \frac{d\phi}{ds}\right\}(s, X_s) ds. \end{aligned}$$

We can apply tha variant of Norris Lemma given in Lemma 4.13. Indeed for $t_0 = \delta$, from the fact that $|\xi| \leq \frac{C}{\delta}$, we have

$$\begin{aligned} &\mathbb{E}[\sup_{0\leq s\leq\delta} |\xi^T Z_s[\sigma_i, \sigma_j](s, X_s)|^p] \leq \frac{C}{\delta^p} \qquad \text{and} \\ &\mathbb{E}\left[\sup_{0\leq s\leq\delta} |\xi^T Z_s \left\{[b, \sigma_i] + \frac{1}{2}\sum_{k=1}^d [\sigma_k, [\sigma_k, \sigma_i]]\right\}(s, X_s)|^p\right] \leq \frac{C}{\delta^p}. \end{aligned}$$

Thus

$$\mathbb{P}\left(\sum_{i=1}^d \int_0^\delta |\xi^T Z_s \sigma_i(s, X_s)|^2 ds \leq \varepsilon^q \text{ and } \sum_{i,j=1}^d \int_0^\delta |\xi^T Z_s[\sigma_i, \sigma_j](s, X_s)|^2 ds > \frac{\varepsilon}{\delta}\right) \leq \varepsilon^p$$

for any $p > 0$, $\delta \leq \delta_p$ for $\varepsilon \leq \varepsilon_p$ uniform in δ. Both the estimate of I_1 and I_2 do not depend on v, so it also holds

$$\sup_{|v|=1} \mathbb{P}(\langle \bar{\gamma}_F v, v\rangle \leq \varepsilon^q) \leq \varepsilon^p$$

$p > 0$, $\delta \leq \delta_p$ for $\varepsilon \leq \varepsilon_p$ uniform in δ. □

4.5.3 A specific version of Norris lemma

Lemma 4.13. *Suppose $u(t) = (u_1(t), \ldots, u_d(t))$ and $a(t)$ are adapted processes such that for some $p \geq 1$*

$$\mathbb{E}\left[\sup_{0\leq s\leq t_0} |u_s|^p\right] \leq \frac{C}{t_0^p}, \qquad \mathbb{E}\left[\sup_{0\leq s\leq t_0} |a_s|^p\right] \leq \frac{C}{t_0^p}. \tag{4.5.9}$$

for $t_0 \leq 1$*. Set*

$$Y(t) = y + \int_0^t a(s)ds + \sum_{k=1}^d \int_0^t u_k(s)dW_s^k$$

Then, for any $q > 4$*, for any* $r > 0$ *such that* $6r + 4 < q$*, there exists* $\varepsilon_0(q, r, p)$ *such that* $\forall t_0 \leq 1$*,* $\varepsilon \leq \varepsilon_0(q, r, p)$

$$\mathbb{P}\left\{\int_0^{t_0} Y_t^2 dt < \varepsilon^q, \int_0^{t_0} |u(t)|^2 dt \geq \frac{\varepsilon}{t_0}\right\} \leq \varepsilon^{rp}.$$

Proof. 114 Set $\theta_t = |a_t| + |u_t|$, and

$$T = \inf\left\{s \geq 0 : \sup_{0 \leq u \leq s} \theta_u > \frac{\varepsilon^{-r}}{t_0}\right\} \wedge t_0.$$

We have

$$\mathbb{P}\left\{\int_0^{t_0} Y_t^2 dt < \varepsilon^q, \int_0^{t_0} |u(t)|^2 dt \geq \frac{\varepsilon}{t_0}\right\} \leq A_1 + A_2.$$

where $A_1 = \mathbb{P}[T < t_0]$ and

$$A_2 = \mathbb{P}\left\{\int_0^{t_0} Y_t^2 dt < \varepsilon^q, \int_0^{t_0} |u_t|^2 dt \geq \frac{\varepsilon}{t_0}, T = t_0\right\}$$

An upper bound for A_1 easily follows from (4.5.9). Indeed

$$\mathbb{P}[T < t_0] \leq \mathbb{P}\left[\sup_{0 \leq s \leq t_0} \theta_u > \frac{\varepsilon^{-r}}{t_0}\right] \leq t_0^p \varepsilon^{rp} \mathbb{E}\left[\sup_{0 \leq s \leq t_0} \theta_s^p\right] \leq c\varepsilon^{rp}.$$

for $\varepsilon \leq \varepsilon_0$ and c not dependent upon t_0. To estimate A_2 we introduce

$$N_t = \int_0^t Y_s \sum_{k=1}^d u_s^k dW_s^k.$$

For

$$\delta = \frac{\varepsilon^{2r+2}}{t_0}, \quad \rho = \frac{\varepsilon^{q-2r}}{t_0^2}$$

define

$$B = \left[\langle N \rangle_T < \rho, \sup_{0 \leq s \leq T} |N_s| \geq \delta\right],$$

By the exponential martingale inequality

$$\mathbb{P}(B) \leq \exp(\frac{-\delta^2}{2\rho}) \leq \exp\left(-\varepsilon^{2r+4-q}\right)$$

We show

$$\left\{\int_0^{t_0} Y_t^2 dt < \varepsilon^q, \int_0^{t_0} |u_t|^2 dt \geq \frac{\varepsilon}{t_0}, T = t_0\right\} \subset B, \tag{4.5.10}$$

and this finishes the proof. We suppose $\omega \notin B$, $\int_0^{t_0} Y_t^2 dt < \varepsilon^q$ and $T = t_0$ and show $\int_0^{t_0} |u_t|^2 dt < \varepsilon/t_0$. With these assumption,

$$\langle N \rangle_T = \int_0^T Y_t^2 |u_t|^2 dt < \frac{\varepsilon^{q-2r}}{t_0^2} = \rho.$$

We are also supposing $\omega \notin B$, so $\sup_{0\le t\le T} |\int_0^t Y_s \sum_{k=1}^d u_s^k dW_s^k| < \delta = \varepsilon^{2r+2}/t_0$ must hold. From $6r+4 \le q$,

$$\sup_{0\le t\le T} \left| \int_0^t Y_s a_s ds \right| \le \left(t_0 \int_0^T Y_s^2 a_s^2 ds \right)^{1/2} < t_0^{1/2} \frac{\varepsilon^{-r+q/2}}{t_0} \le \frac{\varepsilon^{2r+2}}{t_0}$$

Thus

$$\sup_{0\le t\le T} \left| \int_0^t Y_s dY_s \right| < \sup_{0\le t\le T} \left| \int_0^t Y_s a_s ds + \int_0^t Y_s u_s dW_s \right| < \frac{2\varepsilon^{2r+2}}{t_0}.$$

By Ito's formula $Y_t^2 = y^2 + 2\int_0^t Y_s dY_s + \langle M \rangle_t$, with

$$M_t = \sum_{k=1}^d \int_0^t u_k(s) dW_s^k.$$

So

$$\begin{aligned} \int_0^T \langle M \rangle_t dt &= \int_0^T Y_t^2 dt - Ty^2 - 2\int_0^T \int_0^t Y_s dY_s dt \\ &< \varepsilon^q + 2\varepsilon^{2r+2} < 3\varepsilon^{2r+2}, \end{aligned}.$$

Since $\langle M \rangle_t$ is increasing in t, for $0 < \gamma < T$

$$\gamma \langle M \rangle_{T-\gamma} < 3\varepsilon^{2r+2}.$$

Using also the fact that

$$\langle M \rangle_T - \langle M \rangle_{T-\gamma} = \int_{T-\gamma}^T |u_s|^2 ds \le \gamma \frac{\varepsilon^{-2r}}{t_0^2}$$

we have

$$\langle M \rangle_T < \frac{3\varepsilon^{2r+2}}{\gamma} + \gamma \frac{\varepsilon^{-2r}}{t_0^2}.$$

With $\gamma = t_0 \varepsilon^{2r+1}$, this gives

$$\langle M_T \rangle = \int_0^T |u_s|^2 ds < \frac{4\varepsilon}{t_0}.$$

□

4.6 Tube estimates

We consider the diffusion (4.1.1) and, for $\phi \in L^2[0,T]$, the skeleton (4.1.2) on $[0,T]$. We denote in this section, for fixed $t \in [0,T]$,

$$\mathcal{C}_t = \{C_t = K\,(n_t/\lambda_t)^q\,, \exists K, q \geq 1 \text{ universal constants}\}\,.$$

Recall (H_3), and remark that the function $t \to C_t = K\,(n_t/\lambda_t)^q$ is in $L(\mu^{2q}, h)$, for p large enough. Denote, for K_* and q_* constants,

$$R_*(\phi) = \inf_{0\leq t\leq T}\left(\frac{1}{K_*}\frac{\lambda_t}{\mu n_t}\right)^{q_*}\left(h \wedge \inf\left\{\delta/\int_t^{t+\delta}|\phi_s|^2 ds\ :\quad t\in[0,T], \delta\in[0,h]\right\}\right). \tag{4.6.1}$$

Theorem 4.14. *There exist universal positive constants K, q such that for $R \in]0,1]$*

$$\exp\left(-K\int_0^T\left(\frac{\mu n_t}{\lambda_t}\right)^q\left(\frac{1}{h}+\frac{1}{R}+|\phi_t|^2 dt\right)\right) \leq \mathbb{P}\left(\sup_{t\leq T}|X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} \leq 1\right).$$

Moreover, exist K_*, q_*, K, q such that for $R \leq R_*(\phi)$

$$\mathbb{P}\left(\sup_{t\leq T}|X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} \leq 1\right) \leq \exp\left(-\int_0^T\frac{1}{K}\left(\frac{\lambda_t}{\mu n_t}\right)^q\left(\frac{1}{R}+|\phi_t|^2\right)dt\right)$$

Remark 4.15. For $R \leq R_*(\phi) \leq h$ the lower bound holds as in (4.2.6)

We prove first that moving along a control $\phi \in L^2[0,T]$ for a small time, the trajectory remains close to the initial point in the A_δ-norm. Recall (4.1.2), ad take initial condition x_0.

$$x_t(\phi) = x_0 + \int_0^t \sigma(s, x_s(\phi))\phi_s ds + \int_0^t b(s, x_s(\phi))ds.$$

We write here x_t for $x_t(\phi)$ to have a more readable notation. Define, for $\delta > 0$,

$$\varepsilon = \left(\int_0^\delta |\phi_s|^2 ds\right)^{1/2}.$$

Set

$$\mathcal{C}_{(0,x_0)} = \left\{C_{(0,x_0)} = K\,(n(0,x_0)/\lambda(0,x_0))^q\,, \exists K, q \geq 1 \text{ universal constants}\right\}.$$

Lemma 4.16. *There exist $\delta_*, \varepsilon_* \in 1/\mathcal{C}$ such that for $\delta \leq \delta_*$, $\varepsilon \leq \varepsilon_*$, for every $0 \leq t \leq \delta$ and $z \in \mathbb{R}^n$,*

$$|z|^2_{A_\delta(0,x_0)} \leq 4\,|z|^2_{A_\delta(t,x_t)} \leq 16\,|z|^2_{A_\delta(0,x_0)}\,. \tag{4.6.2}$$

Moreover $\exists\, C \in \mathcal{C}$ such that

$$\sup_{0\leq t\leq\delta}|x_t(\phi) - (x_0 + b(0,x_0)t)|_{A_\delta(0,x_0)} \leq C(\varepsilon \vee \delta). \tag{4.6.3}$$

Proof. Since $\delta \le \delta_*$, $\varepsilon \le \varepsilon_*$, we can choose δ_*, ε_* such that

$$|x_s - x_0| + |s| \le C\sqrt{\delta}(\varepsilon + \sqrt{\delta}) \le \sqrt{\delta}/C. \tag{4.6.4}$$

(4.6.2) follows from Lemma 4.18. For $t \le \delta$ we write

$$J_t := x_t - x_0 - b(0, x_0)t = \int_0^t \partial_s x_s - b(s, x_s)ds + \int_0^t b(s, x_s) - b(0, x_0)ds.$$

Using (4.7.4) we get

$$|J_t|^2_{A_\delta(0,x_0)} \le 2t \int_0^t |\partial_s x_s - b(s, x_s)|^2_{A_\delta(0,x_0)} ds + 2t \int_0^t |b(s, x_s) - b(0, x_0)|^2_{A_\delta(0,x_0)} ds.$$

For $s \le t \le \delta$, from (4.6.2), we have

$$|\partial_s x_s - b(s, x_s)|^2_{A_\delta(0,x_0)} \le 4|\partial_s x_s - b(s, x_s)|^2_{A_\delta(s,x_s)}.$$

Moreover, for $i = 1, \dots, m$, we set $\psi^{(j-1)d+j} = \frac{1}{\sqrt{\delta}}\phi^j$ for $j = 1, \dots, d$, $\psi^i = 0$ otherwise. We can write

$$\partial_s x_s - b(s, x_s) = \sum_{j=1}^d \sigma_j(s, x_s)\phi_s^j = A_\delta(s, x_s)\psi_s$$

so that

$$|\partial_s x_s - b(s, x_s)|_{A_\delta(s,x_s)} \le |\psi_s| = \frac{1}{\sqrt{\delta}} |\phi_s| .$$

Then, for $t \le \delta$

$$2t \int_0^t |\partial_s x_s - b(s, x_s)|^2_{A_\delta(0,x_0)} ds \le 8\delta \int_0^\delta |\partial_s x_s - b(s, x_s)|^2_{A_\delta(s,x_s)} ds \le 8 \int_0^\delta |\phi_s|^2 ds = 8\varepsilon^2.$$

With the following estimate the statement is proved:

$$\begin{aligned} 2t \int_0^t |b(s, x_s) - b(0, x_0)|^2_{A_\delta(0,x_0)} ds &\le C\delta \int_0^\delta |b(s, x_s) - b(0, x_0)|^2 ds \\ &\le C \int_0^\delta (|s| + |x_s - x_0|)^2 ds \le C\delta^2. \end{aligned}$$

□

Proof. (of Theorem 4.14)
STEP 1: We first prove the lower bound. We set, for large q_1, K_1 to be fixed in the sequel,

$$f_R(t) = K_1 \left(\frac{\mu n_t}{\lambda_t}\right)^{q_1} \left(\frac{1}{h} + \frac{1}{R} + |\phi_t|^2\right).$$

Recall (H_3): $|\phi_.|^2, n_., \lambda_. \in L(\mu, h)$, $\exists \mu \ge 1$, $0 < h \le 1$, where

$$L(\mu, h) = \{f : f(t) \le \mu f(s) \quad \text{for} \quad |t - s| \le h\}.$$

This implies $f_R \in L(\mu^{2q_1+1}, h)$. We also define

$$\delta(t) = \inf_{\delta>0} \left\{ \int_t^{t+\delta} f_R(s)ds \geq \frac{1}{\mu^{2q_1+1}} \right\}. \quad (4.6.5)$$

Clearly $\delta(t) \leq h$, so we can use on the intervals $[t, t+\delta(t)]$ the fact that our bounds are in $L(\mu, h)$. If $0 < t - t' \leq h$,

$$\mu^{2q_1+1} f_R(t)\delta(t) \geq \int_t^{t+\delta(t)} f_R(s)ds = 1 = \int_{t'}^{t'+\delta(t')} f_R(s)ds \geq \mu^{-(2q_1+1)} f_R(t)\delta(t'),$$

so $\delta(t')/\delta(t) \leq \mu^{4q_1+2}$.The converse holds as well, so $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$. We also set the energy

$$\varepsilon(t) = \left(\int_t^{t+\delta(t)} |\phi_s|^2 ds \right)^{1/2}.$$

We have

$$\frac{1}{\mu^{2q_1+1}} = \int_t^{t+\delta(t)} f_R(s)ds \geq \int_t^{t+\delta(t)} \frac{f_R(t)}{\mu^{2q_1+1}} ds \geq \delta(t) \frac{f_R(t)}{\mu^{2q_1+1}},$$

so

$$\delta(t) \leq \frac{1}{f_R(t)} \leq \frac{R}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}. \quad (4.6.6)$$

Similarly,

$$\frac{1}{\mu^{2q_1+1}} \geq \int_t^{t+\delta(t)} K_1 \left(\frac{\mu n_s}{\lambda_s} \right)^{q_1} |\phi_s|^2 ds = \frac{1}{\mu^{2q_1}} K_1 \left(\frac{\mu n_t}{\lambda_t} \right)^{q_1} \varepsilon(t)^2,$$

and we can write both

$$\delta(t) \leq \frac{1}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}, \quad \text{and} \quad \varepsilon(t)^2 \leq \frac{1}{K_1} \left(\frac{\lambda_t}{\mu n_t} \right)^{q_1}. \quad (4.6.7)$$

We set our time grid as

$$t_0 = 0; \quad t_k = t_{k-1} + \delta(t_{k-1}),$$

and introduce the following notation on the grid:

$$\delta_k = \delta(t_k); \quad \varepsilon_k = \varepsilon(t_k); \quad n_k = n(t_k, x_k); \quad \lambda_k = \lambda(t_k, x_k); \quad X_k = X_{t_k}; \quad x_k = x_{t_k}.$$

We also define

$$\hat{X}_k = X_k + b(t_k, X_k)\delta_k; \quad \hat{x}_k = x_k + b(t_k, x_k)\delta_k,$$

and for $t_k \leq t \leq t_{k+1}$,

$$\hat{X}_k(t) = X_k + b(t_k, X_k)(t - t_k); \quad \hat{x}_k(t) = x_k + b(t_k, x_k)(t - t_k).$$

Moreover we denote

$$|\xi|_k = |\xi|_{A_{\delta_k}(t_k, x_k)}; \quad \mathcal{C}_k = \mathcal{C}_{t_k},$$

and $r_*^k \in \mathcal{C}_k$ the ray r_* of Theorem 4.5 associated to x_k. Lemma 4.16 holds for δ_k and ε_k small enough, and in this case $|x_{k+1} - \hat{x}_k|_k \le C_k(\varepsilon_k \vee \delta_k)$. Moreover, for all $t_k \le t \le t_{k+1}$, applying also (4.7.1), $|x_t - \hat{x}_k(t)|_{A_R(t,x_t)} \le C_k(\varepsilon_k \vee \delta_k)\sqrt{\delta_k/R}$. Recall (4.6.7), and we fix q_3, K_3 such that, for $q_1 \ge q_3, K_1 \ge K_3$, Lemma 4.16 holds and

$$|x_{k+1} - \hat{x}_k|_k \le r_*^k/4 \tag{4.6.8}$$

$$|\hat{x}_k(t) - x_t|_{A_R(t,x_t)} \le \frac{1}{4} \text{ for all } t_k \le t \le t_{k+1}, \tag{4.6.9}$$

and moreover the theorem in short time 4.5 holds. Also (4.6.2) holds and

$$\frac{1}{2}|\xi|_{A_{\delta_k}(t_k,x_k)} \le |\xi|_{A_{\delta_k}(t_{k+1},x_{k+1})} \le 2|\xi|_{A_{\delta_k}(t_k,x_k)}.$$

now, $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$ implies $\delta_k/\delta_{k+1} \le \mu^{4q_1+2}$ and $\delta_{k+1}/\delta_k \le \mu^{4q_1+2}$. Together with (4.7.1) this gives

$$\frac{1}{2\mu^{2q_1+1}}|\xi|_k \le |\xi|_{k+1} \le 2\mu^{2q_1+1}|\xi|_k \tag{4.6.10}$$

We now set, for K_2, q_2 to be fixed in the sequel,

$$r_k = \frac{1}{K_2\mu^{2q_1+2q_2+1}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}, \tag{4.6.11}$$

and define

$$\Gamma_k = \{|X_k - x_k|_k \le r_k\}, \quad D_k = \{\sup_{t_k \le t \le t_{k+1}} |X_t - x_t|_{A_R(t,x_t)} \le 1\},$$

and $\mathbb{P}_k$ as the conditional probability

$$\mathbb{P}_k(\cdot) = \mathbb{P}\left(\cdot | W_t, t \le t_k; X_k \in \Gamma_k\right).$$

We denote p_k the density of X_{k+1} with respect to this probability. We prove that on $\{|\cdot - x_{k+1}|_{k+1} \le r_{k+1}\}$ we can apply Theorem 4.5 to p_k and so there exists $\underline{C}_k \in \mathcal{C}_k$ such that

$$\frac{1}{\underline{C}_k \delta_k^{n - \frac{dim\langle\sigma(t_k,X_k)\rangle}{2}}} \le p_k(y)$$

or, because of (4.4.21),

$$\frac{1}{\underline{C}_k\sqrt{\det A_{\delta_k}A_{\delta_k}^T(t_k,X_k)}} \le p_k(y) \tag{4.6.12}$$

We estimate

$$|y - \hat{X}_k|_k \le |y - x_{k+1}|_k + |x_{k+1} - \hat{x}_k|_k + |\hat{x}_k - \hat{X}_k|_k. \tag{4.6.13}$$

We already have (4.6.8). Since we are on $|y-x_{k+1}|_{k+1}\le r_{k+1}$, from (4.6.10) and the fact that $r_{k+1}/r_k\le\mu^{2q_2}$

$$|y-x_{k+1}|_k\le 2\mu^{2q_1+1}|y-x_{k+1}|_{k+1}\le 2\mu^{2q_1+1}r_{k+1}\le 2\mu^{2q_1+2q_2+1}r_k\le\frac{2}{K_2}\left(\frac{\lambda_k}{n_k}\right)^{q_2}.$$

It also holds $|\hat{x}_k-\hat{X}_k|_k\le C_k|x_k-X_k|_k\le C_k r_k$, for some $C_k\in\mathcal{C}_k$. Similarly, $|\hat{x}_k(t)-\hat{X}_k(t)|_{A_R(t,x_t)}\le C_k r_k$, for all $t_k\le t\le t_{k+1}$. Recalling (4.6.11), we can fix K_2,q_2 such that $|y-x_{k+1}|_k\le r_*^k/8$, $|\hat{x}_k-\hat{X}_k|_k\le r_*^k/8$, and

$$|\hat{X}_k(t)-\hat{x}_k(t)|_{A_R(t,x_t)}\le 1/4,\qquad\text{for all } t_k\le t\le t_{k+1}. \tag{4.6.14}$$

From (4.6.13), this implies $|y-\hat{X}_k|_k\le r_*^k/2$. We also have, from (4.7.2), $|x_k-X_k|\le|x_k-X_k|_k\sqrt{\lambda_k}\sqrt{\delta_k}$, so we can also fix K_2,q_2 such that $r_k\sqrt{\lambda_k}\le 1/C$ in (4.7.5). Therefore

$$\frac{1}{2}|\xi|_k\le|\xi|_{A_{\delta_k}(t_k,X_k)}\le 2|\xi|_k.$$

So $|y-\hat{X}_k|_{A_{\delta_k}(t_k,X_k)}\le r_*^k$. Now, also from (4.6.10)

$$\{|\cdot-x_{k+1}|_{A_{\delta_k}(t_k,X_k)}\le r_{k+1}/(4\mu^{2q_1+1})\}\subset\{|\cdot-x_{k+1}|_k\le r_{k+1}/(2\mu^{2q_1+1})\}\subset\{|\cdot-x_{k+1}|_{k+1}\le r_{k+1}\},$$

and $r_{k+1}/(4\mu^{2q_1+1})\ge r_k/(4\mu^{2q_1+2q_2+1})=\frac{1}{4K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}$. So

$$Leb(|\cdot-x_{k+1}|_{k+1}\le r_{k+1})\ge\sqrt{\det(A_{\delta_k}A_{\delta_k}^T(t_k,X_k))}\left(\frac{1}{4K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^n.$$

So, from (4.6.12),

$$\mathbb{P}_k(\Gamma_{k+1})\ge\frac{1}{\underline{C}_k}\left(\frac{1}{4K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^n$$

where $\underline{C}_k\in 1/\mathcal{C}_k$ is the constant in (4.4.1). This implies

$$2\mu^{-4q_1}\exp(-K_4(\log\mu+\log n_k-\log\lambda_k))\le P_k(\Gamma_{k+1})$$

for some constant K_4 (depending on K_2,K_3,q_2,q_3, remark that q_1 is not a constant, since we have not fixed it yet, and that is why we keep the explicit dependence on q_1 in the expression above).

STEP 2: Consider now $t_k\le t\le t_{k+1}$. Recall the definition

$$D_k=\left\{\sup_{t_k\le t\le t_{k+1}}|X_t-x_t|_{A_R(t,x_t)}\le 1\right\},$$

and introduce

$$E_k=\left\{\sup_{t_k\le t\le t_{k+1}}|X_t-\hat{X}_k(t)|_{A_R(t,x_t)}\le\frac{1}{2}\right\}.$$

We decompose

$$|X_t - x_t|_{A_R(t,x_t)} \le |X_t - \hat{X}_k(t)|_{A_R(t,x_t)} + |\hat{X}_k(t) - \hat{x}_k(t)|_{A_R(t,x_t)} + |\hat{x}_k(t) - x_t|_{A_R(t,x_t)},$$

and, from the previous part of the proof, (4.6.9) gives $|\hat{x}_k(t) - x_t|_{A_R(t,x_t)} \le 1/4$, and (4.6.14) gives $|\hat{X}_k(t) - \hat{x}_k(t)|_{A_R(t,x_t)} \le 1/4$. So $|X_t - x_t|_{A_R(t,x_t)} \le |X_t - \hat{X}_k(t)|_{A_R(t,x_t)} + 1/2$, and therefore $E_k \subset D_k$. Since, passing from Stratonovic to Ito integrals,

$$\begin{aligned}|X_t - \hat{X}_k(t)|_{A_R(t,x_t)} \le & \frac{1}{\sqrt{R}}|\sigma(t_k, X_{t_k})(W_t - W_{t_k})|_{A(t,x_t)} \\ & + |A_R(t,x_t)^{-1}\int_{t_k}^t \sigma(s,X_s) - \sigma(t_k,X_k)dW_s| \\ & + |A_R(t,x_t)^{-1}\int_{t_k}^t b(s,X_s) - b(t_k,X_k)ds| \\ & + \sum_{l=1}^d |A_R(t,x_t)^{-1}\int_{t_k}^t \nabla\sigma_l(s,X_s)(\sigma_l(s,X_s) - \sigma_l(t_k,X_k))ds|,\end{aligned}$$

using the exponential martingale inequality we find that

$$\mathbb{P}_k(E_k^c) \le \exp\left(-\frac{1}{K_5}\left(\frac{\lambda_k}{\mu n_k}\right)^{q_5}\frac{R}{\delta_k}\right)$$

for some constants K_5, q_5. From (4.6.6), $R/\delta_k \ge K_1(\mu n_k/\lambda_k)^{q_1}$, so choosing and fixing now q_1, K_1 large enough we conclude

$$\mathbb{P}_k(E_k^c) \le \mu^{-4q_1}\exp(-K_4(\log\mu + \log n_k - \log\lambda_k)) \le \frac{1}{2}\mathbb{P}_k(\Gamma_{k+1}),$$

so

$$\begin{aligned}\mathbb{P}_k(\Gamma_{k+1}\cap D_k) \ge \mathbb{P}_k(\Gamma_{k+1}\cap E_k) \ge \mathbb{P}_k(\Gamma_{k+1}) - \mathbb{P}_k(E_k^c) \ge \frac{1}{2}\mathbb{P}_k(\Gamma_{k+1}) \\ \ge \exp\left(-K_6(\log\mu + \log n_k - \log\lambda_k)\right),\end{aligned} \tag{4.6.15}$$

for some constant K_6. Let now $N(T) = \max\{k : t_k \le T\}$. From Definition 4.6.5

$$\int_0^T f_R(t)dt \ge \sum_{k=1}^{N(T)}\int_{t_{k-1}}^{t_k} f_R(t)dt = \frac{N(T)}{\mu^{2q_1+1}}.$$

From (4.6.15),

$$\begin{aligned}\mathbb{P}\left(\sup_{t\le T}|X_t - x_t|_{A_R(t,x_t))} \le 1\right) & \ge \mathbb{P}\left(\bigcap_{k=1}^{N(T)}\Gamma_{k+1}\cap D_k\right) \\ & \ge \prod_{k=1}^{N(T)}\exp(-K_6(\log\mu + \log n_k - \log\lambda_k)) \\ & = \exp\left(-K_6\sum_{k=1}^{N(T)}\log\mu + \log n_k - \log\lambda_k\right).\end{aligned}$$

Since

$$\sum_{k=1}^{N(T)}(\log\mu+\log n_k-\log\lambda_k)=\mu^{2q_1+1}\sum_{k=1}^{N(T)}\int_{t_k}^{t_k+1}f_R(s)ds(\log\mu+\log n_k-\log\lambda_k)$$
$$\leq\mu^{2q_1+1}\int_0^T f_R(t)\log\left(\frac{\mu^3 n_t}{\lambda_t}\right)dt,$$

the lower bound follows.

STEP 3: We now prove the upper bound. Now recall (4.6.1), and $R\leq R_*(\phi)$. We define, with the same K_1,q_1 as in STEP 1 and 2,

$$g_R(t)=\frac{1}{h}+\frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7}+K_1\mu^{2q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}|\phi_t|^2.$$

for some constants $K_7>K_1, q_7>q_1+1$ to be fixed in the sequel. We define a new $\delta(t)$

$$\delta(t)=\inf_{\delta>0}\left\{\int_t^{t+\delta}g_R(s)ds\geq 1\right\}.$$

Clearly $\delta(t)\leq h$, so we can use on the intervals $[t,t+\delta(t)]$ the property of being in $L(\mu,h)$. If $0<t-t'\leq h$,

$$\mu^{2q_7}g_R(t)\delta(t)\geq\int_t^{t+\delta(t)}g_R(s)ds=1=\int_{t'}^{t'+\delta(t')}g_R(s)ds\geq\mu^{-2q_7}g_R(t)\delta(t'),$$

so $\delta(t')/\delta(t)\leq\mu^{4q_7}$. Taking q_* and K_* in (4.6.1) large enough such that $q_*>5q_1+1+q_7$, $K_*>K_1K_7$,

$$\int_t^{t+\delta(t)}\frac{1}{h}ds\leq\int_t^{\delta(t)}\frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7}ds$$

and again from (4.6.1)

$$\int_t^{t+\delta(t)}K_1\mu^{2q_1+1}\left(\frac{\mu n_s}{\lambda_s}\right)^{q_1}|\phi_s|^2ds\leq K_1\mu^{4q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\int_t^{t+\delta(t)}|\phi_s|^2ds$$
$$\leq K_1\mu^{4q_1+1}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\frac{\delta(t)}{R}\frac{1}{K_*}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_*}$$
$$\leq\int_t^{t+\delta(t)}\frac{1}{R}\frac{1}{\mu^{2q_7}K_7}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7}ds.$$

Therefore, since $1=\int_t^{t+\delta(t)}g_R(s)ds$ and

$$\int_t^{t+\delta(t)}\frac{1}{R}\frac{1}{K_7\mu^{2q_7}}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7}ds\leq\int_t^{t+\delta(t)}g_R(s)ds\leq 3\int_t^{t+\delta(t)}\frac{1}{R}\frac{1}{K_7\mu^{2q_7}}\left(\frac{\lambda_s}{\mu n_s}\right)^{q_7}ds,$$

we find that forall t

$$\frac{1}{K_7\mu^{4q_7}}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7}\leq\frac{R}{\delta(t)}\leq\frac{3}{K_7}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7}$$

For q_*, K_* large enough this also implies

$$\delta(t)\leq\frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}.$$

We set

$$\varepsilon(t)=\left(\int_t^{t+\delta(t)}|\phi_s|^2ds\right)^{1/2}.$$

We find, with the same computations as before,

$$\varepsilon(t)^2\leq\frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}.$$

This implies that (4.6.7) also holds with this new grid, and also lemma 4.16. Since we are taking the same K_1 and q_1 as before (4.6.10) holds. For the same reason, the theorem in short time 4.5 also holds. We define

$$\Delta_k=\{|X_k-x_k|_{A_R(t_k,x_k)}\leq 1\},$$

$\tilde{\mathbb{P}}_k$ as the conditional probability $\tilde{\mathbb{P}}_k(\cdot)=\mathbb{P}\left(\cdot|W_t,t\leq t_k;X_k\in\Delta_k\right)$. As we did in STEP 1, if q_*,K_* are large enough, R is small enough and the upper bound for the density holds on Δ_{k+1}. Because of (4.6.2),

$$Leb(|\cdot-x_k|_{A_R(t_{k+1},x_{k+1})}\leq 1)\leq 2^n Leb(|\cdot-x_k|_{A_R(t_k,x_k)}\leq 1)=2^n\det(A(t_k,x_k))R^{n-\frac{\dim\langle\sigma(t_k,x_k)\rangle}{2}}.$$

Now, using the short time density estimate,

$$\begin{aligned}\tilde{\mathbb{P}}_k(\Delta_{k+1})&\leq 2^n\det(A(t_k,x_k))R^{n-\frac{\dim\langle\sigma(t_k,x_k)\rangle}{2}}\,\overline{C}_k\,\delta_k^{n-\frac{\dim\langle\sigma(t_k,x_k)\rangle}{2}}\\&\leq 2^n\det(A(t_k,x_k))\overline{C}_k\left(\frac{R}{\delta}\right)^{n-\frac{\dim\langle\sigma(t_k,x_k)\rangle}{2}}.\end{aligned}$$

where $\overline{C}_k$ is the constant of theorem 4.5. Recall

$$\frac{R}{\delta(t)}\leq\frac{3}{K_7}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_7},$$

so we fix now K_7,q_7 large enough to have

$$\tilde{\mathbb{P}}_k(\Delta_{k+1})\leq\exp(-K_{10})$$

for a $K_{10} > 0$. (We also fix now q_*, K_*, whose size depend on q_7, K_7). From the definition of $N(T)$

$$\int_0^T g_R(t)dt = \sum_{k=1}^{N(T)} \int_{t_{k-1}}^{t_k} g_R(t)dt = N(T).$$

As before

$$\begin{aligned}\mathbb{P}\left(\sup_{t\le T} |X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} \le 1\right) &\le \prod_{k=1}^{N(T)} \tilde{\mathbb{P}}_k(\Delta_{k+1}) \\ \le \prod_{k=1}^{N(T)} \exp(-K_{10}) = \exp(-K_{10}\, N(T)) &\le \exp\left(-K_{10} \int_0^T g_R(t)\right),\end{aligned}$$

and we have the upper bound. □

4.7 Norms and distances

4.7.1 Matrix norms

We write $A_R = A_R(0, X_0)$ and we work with the norm $|y|^2_{A_R} = \langle (A_R A_R^T)^{-1} y, y\rangle$, $y \in \mathbb{R}^n$.

Lemma 4.17. *i) For every $y \in \mathbb{R}^n$ and $0 < R \le R' \le 1$*

$$\sqrt{\frac{R}{R'}}\,|y|_{A_R} \ge |y|_{A_{R'}} \ge \frac{R}{R'}\,|y|_{A_R} \tag{4.7.1}$$

$$\frac{1}{\sqrt{R}\sqrt{\lambda^{\#}(A)}}\,|y| \le |y|_{A_R} \le \frac{1}{R\sqrt{\lambda_{\#}(A)}}\,|y|\,. \tag{4.7.2}$$

ii) For every $z \in \mathbb{R}^m$ and $R > 0$

$$|A_R z|_{A_R} \le |z|\,. \tag{4.7.3}$$

iii) For every $\mu \in L^2([0,T];\mathbb{R}^m)$ and $R > 0$

$$\Big|\int_0^t \mu_s\, ds\Big|^2_{A_R} \le t \int_0^t |\mu_s|^2_{A_R}\, ds, \quad t \in [0,T]. \tag{4.7.4}$$

Proof. 115 *i)* It is easy to check that

$$\frac{R'}{R} A_R A_R^T \le A_{R'} A_{R'}^T \le \left(\frac{R'}{R}\right)^2 A_R A_R^T$$

which is equivalent to (4.7.1). This also implies (taking $R'=1$ so $A_{R'}=A$) that

$$\frac{1}{R}\lambda_{\#}(A_R)\leq\lambda_{\#}(A)\leq\frac{1}{R^2}\lambda_{\#}(A_R)\quad and\quad \frac{1}{R}\lambda^{\#}(A_R)\leq\lambda^{\#}(A)\leq\frac{1}{R^2}\lambda^{\#}(A_R)$$

which immediately gives (4.7.2).
ii) For $z\in R^m$, we write $z=A_R^Ty+w$ with $y\in R^n$ and $w\in(\mathrm{Im}A_R^T)^{\perp}=KerA_R$. Then $A_Rz=A_RA_R^Ty$ so that

$$\begin{aligned}|A_Rz|^2_{A_R} &= \left|A_RA_R^Ty\right|^2_{A_R}=\left\langle (A_RA_R^T)^{-1}A_RA_R^Ty,A_RA_R^Ty\right\rangle\\ &= \left\langle z,A_RA_R^Tz\right\rangle=\left\langle A_R^Ty,A_R^Ty\right\rangle=\left|A_R^Ty\right|^2\leq|z|^2\end{aligned}$$

and (4.7.3) holds.
iii) For $\mu\in L^2([0,T];\mathbb{R}^m)$ and $t\in[0,T]$

$$\begin{aligned}\Big|\int_0^t\mu_sds\Big|^2_{A_R}=&\langle {A_R}^{-1}\int_0^t\mu_sds,\int_0^t\mu_sds\rangle=\int_0^t\int_0^t\langle {A_R}^{-1}\mu_s,\mu_u\rangle dsdu\\ =&\frac{1}{2}\int_0^t\int_0^t\Big(\langle {A_R}^{-1}(\mu_s-\mu_u),\mu_s-\mu_u\rangle-\langle {A_R}^{-1}\mu_s,\mu_s\rangle-\langle {A_R}^{-1}\mu_u,\mu_u\rangle\Big)dsdu\\ =&\frac{1}{2}\int_0^t\int_0^t\Big(|\mu_s-\mu_u|^2_{A_R}-2|\mu_s|^2_{A_R}\Big)dsdu\\ \leq&\int_0^t\int_0^t|\mu_u|^2_{A_R}dsdu=t\int_0^t|\mu_u|^2_{A_R}du.\end{aligned}$$

□

From now on we consider the specific situation when $a_i=\sigma_i(t,x),[a]_{i,j}=[\sigma_i,\sigma_j](t,x)$ and we denote by $A(t,x)$ respectively $A_R(t,x)$ the matrices associated to these coefficients.

Lemma 4.18. *Let $x,y\in\mathbb{R}^n$, $s,t\in[0,1]$, $\delta\leq1$. There exists $C\in\mathcal{C}$ such that if*

$$|x-y|+|t-s|\leq\sqrt{\delta}/C, \tag{4.7.5}$$

then for every $z\in\mathbb{R}^n$

$$\frac{1}{4}\left|z\right|^2_{A_\delta(t,x)}\leq\left|z\right|^2_{A_\delta(s,y)}\leq4\left|z\right|^2_{A_\delta(t,x)}. \tag{4.7.6}$$

Proof. (4.7.6) is equivalent to

$$4(A_\delta A_\delta^T)(t,x)\geq(A_\delta A_\delta^T)(s,y)\geq\frac{1}{4}(A_\delta A_\delta^T)(t,x).$$

Recall that A_δ, k the columns of A_δ. We use (4.7.5) and the fact that $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$:

$$\begin{aligned}
\langle A_\delta A_\delta^T(s,y)z, z\rangle &= \sum_{k=1}^m \langle A_{\delta,k}(s,y), z\rangle^2 \\
&= \sum_{k=1}^m (\langle A_{\delta,k}(t,x), z\rangle + \langle A_{\delta,k}(s,y) - A_{\delta,k}(t,x), z\rangle)^2 \\
&\geq \frac{1}{2}\sum_{k=1}^m \langle A_{\delta,k}(t,x), z\rangle^2 - \sum_{k=1}^m (\langle A_{\delta,k}(s,y) - A_{\delta,k}(t,x), z\rangle)^2 \\
&\geq \frac{1}{2}\sum_{k=1}^m \langle A_{\delta,k}(t,x), z\rangle^2 - \bar{C}\delta(|x-y|^2 + |t-s|^2)\,|z|^2
\end{aligned}$$

$\bar{C} \in \mathcal{C}$. From $\lambda_\#(A_\delta(t,x)) \geq \delta^2 \lambda_\#(A(t,x))$ follows

$$\bar{C}\delta(|x-y|^2 + |t-s|^2)\,|z|^2 \leq \frac{1}{4}\sum_{k=1}^m \langle A_{\delta,k}(t,x), z\rangle^2 .$$

So

$$\left\langle (A_\delta A_\delta^T)(s,y)z, z\right\rangle \geq \frac{1}{4}\sum_{k=1}^m \langle A_{\delta,k}(t,x), z\rangle^2 = \frac{1}{4}\left\langle (A_\delta A_\delta^T)(t,x)z, z\right\rangle .$$

The converse inequality follows from analogous computations and inequality $(a+b)^2 \leq 2a^2 + 2b^2$. □

4.7.2 The control distance

We establish the link between the norm $|\cdot|_{A_R(t,x)}$ and the control (Caratheodory) distance. We will use in a crucial way the alternative characterizations given in [78]. Since these results hold in the homogeneous case, we suppose now $\sigma_j(t,x) = \sigma_j(x)$. Consequently, $A_R(t,x) = A_R(x)$.

We first introduce a quasi-distance d which is naturally associated to the family of norms $|y|_{A_R(x)}$. We set $\Omega = \{x \in \mathbb{R}^n : \lambda_*(A(x)) > 0\} = \{x : \det(AA^T(x)) \neq 0\}$, which is open because the function $x \mapsto \det AA^T(x)$ is continuous. Notice that if $x \in \Omega$ then $\det(A_R A_R^T(x)) > 0$ for every $R > 0$. For $x, y \in \Omega$, we define $d(x,y)$ by

$$d(x,y) < \sqrt{R} \quad \Leftrightarrow \quad |y-x|_{A_R(x)} < 1.$$

The motivation for taking $\sqrt{R}$ is the following: in the elliptic case $|y-x|_{A_R(x)} \sim R^{-1/2}|y-x|$ so $|y-x|_{A_R(x)} \leq 1$ amounts to $|y-x| \leq \sqrt{R}$. It is straightforward to see that d is a quasi-distance on Ω, meaning that d verifies the following three properties (see [78]):

i) for every $r > 0$, the set $\{y \in \Omega : d(x,y) < r\}$ is open;

ii) $d(x,y) = 0$ if and only if $x = y$;

iii) for every compact set $K \Subset \Omega$ there exists $C > 0$ such that $d(x,y) \le C\big(d(x,z) + d(z,y)\big)$ holds for every $x, y, z \in K$.

We recall the definition of equivalence of semi-distances, given in section 3.4.2. Two quasi-distances $d_1 : \Omega \times \Omega \to \mathbb{R}^+$ and $d_2 : \Omega \times \Omega \to \mathbb{R}^+$ are equivalent if for every compact set $K \Subset \Omega$ there exists a constant C such that for every $x, y \in K$

$$\frac{1}{C} d_1(x,y) \le d_2(x,y) \le C d_1(x,y). \tag{4.7.7}$$

d_1 and d_2 are locally equivalent if for every $x_0 \in \Omega$ there exist a neighborhood V of x_0 and a constant C such that (4.7.7) holds for every $x, y \in V$.

We introduce now the control metric. For $x, y \in \mathbb{R}^n$ we denote by $C(x,y)$ the set of controls $\psi \in L^2([0,1];\mathbb{R}^d)$ such that the corresponding skeleton solution of

$$du_t(\psi) = \sum_{j=1}^{d} \sigma_j(u_t(\psi))\psi_t^j dt, \qquad u_0(\psi) = x \tag{4.7.8}$$

satisfies $u_1(\psi) = y$. Notice that the drift b does not appear in the equation of $u_t(\psi)$. We define the control (Caratheodory) distance as

$$d_c(x,y) = \inf\Big\{ \Big(\int_0^1 |\psi_s|^2 \, ds \Big)^{1/2} : \psi \in C(x,y) \Big\}.$$

We define $C_\infty(x,y)$ the set of controls $\psi \in L^\infty([0,1];\mathbb{R}^d)$ such that the corresponding skeleton solution of (4.7.8) satisfies $u_1(\psi) = y$, and

$$d_\infty(x,y) = \inf\Big\{ \|\psi\|_\infty : \psi \in C_\infty(x,y) \Big\}.$$

We also denote $C_\delta(x,y)$ the set of controls $\phi \in L^2([0,\delta];\mathbb{R}^d)$ such that the corresponding skeleton $u_t(\phi)$ with $u_0(\phi) = x$ verifies $u_\delta(\phi) = y$, and the associated energy

$$\varepsilon_\phi(\delta) = \Big(\int_0^\delta |\phi_s|^2 \, ds \Big)^{1/2}.$$

Theorem 4.19. *Let ε_*, δ_* the constants in lemma 4.16 $C_x \in \mathcal{C}_x$ be the constant in* (4.6.3)*. We also suppose, wlog, $1/C_x \le \varepsilon_* \wedge \delta_* \in 1/\mathcal{C}_x$.*
***A**. For every $x, y \in \Omega$ such that $d_c(x,y) \le 1/C_x^2$ it holds $d(x,y) \le C_x d_c(x,y)$.*
***B**. d is locally equivalent to d_c on Ω.*
***C**. In particular for every compact set $K \Subset \Omega$ there exists r_K and C_K such that for every $x, y \in K$ with $d(x,y) \le r_K$ one has $d_c(x,y) \le C_K d(x,y)$.*

Proof. **A.** Let $\delta > 0$, $x, y \in \Omega$ and $\psi \in C(x,y)$. Setting $x_t = u_{t/\delta}(\psi)$, we obtain $dx_t = \sum_{j=1}^d \sigma_j(x_t)\phi_t^j dt$ with $\phi(t) = \delta^{-1}\psi(t\delta^{-1})$, which means that $x_t = u_t(\phi)$. Notice also that $\int_0^1 |\psi_s|^2 \, ds = \delta \int_0^\delta |\phi_s|^2 \, ds$. This implies

$$d_c(x,y) = \sqrt{\delta} \inf\{\varepsilon_\phi(\delta) : \phi \in C_\delta(x,y)\}.$$

Consider now two points such that $d_c(x,y) \le 1/C_x^2$. Take $\delta = C_x^2 d_c(x,y)^2$. Then there exists a control $\phi \in C_\delta(x,y)$ such that $\varepsilon_\phi(\delta) \le 1/C_x$. Since $\varepsilon_\phi(\delta) \vee \delta \le 1/C_x$, from (4.6.3), we obtain $|y-x|_{A_\delta(x)} \le 1$, and this implies $d(x,y) \le \sqrt{\delta} \le C_x d_c(x,y)$.
B. We prove now the converse inequality. We use a result from [78], for which we need to recall the definition of the quasi-distance d_* (denoted by ρ_2 in [78]). The definition we give here is slightly different but clearly equivalent. For $\theta \in \mathbb{R}^m$ we consider the equation

$$v_t(\theta) = x + \int_0^t A(v_s(\theta))\theta ds \tag{4.7.9}$$

denote

$$\bar{C}_A(x,y) = \{\theta \in \mathbb{R}^m, \text{ satisfying (4.7.9) }, v_1(\theta) = y\}$$

Notice that θ is a constant vector, and not a time depending control as in the standard skeleton, and that in (4.7.9) are involved also the vector fields $[\sigma_i, \sigma_j]$, in contrast with (4.7.8). In both equations the drift term b does not appear. Recall that $\overline{R}$ is the diagonal $m \times m$ matrix with $\overline{R}_{l,l} = R$ for $i \neq p$ and $\overline{R}_{l,l} = \sqrt{R}$, and $A_R(x) = A(x)\overline{R}$.

Recall that, taking $l = (p-1)d + i \in \{1, \ldots, m\}$, with $p, i \in \{1, \ldots, d\}$.

$$\begin{aligned} A_l(t,x) &= [\sigma_i, \sigma_p](t,x) \quad if \quad i \neq p, \\ &= \sigma_i(t,x) \quad if \quad i = p, \end{aligned} \tag{4.7.10}$$

$\overline{R}$ is the diagonal $m \times m$ matrix with $\overline{R}_{l,l} = R$ for $i \neq p$ and $\overline{R}_{l,l} = \sqrt{R}$ for $i = p$, and

$$A_R(t,x) = A(t,x)\overline{R} = (\sqrt{R}\sigma_i(t,x), [\sqrt{R}\sigma_j, \sqrt{R}\sigma_p](t,x))_{i,j,p=1,\ldots,d, j\neq p}.$$

We define

$$d_*(x,y) = \inf\{R > 0 | \exists \theta \in \bar{C}_A(x,y), \, |\overline{R}^{-1}\theta| < 1\}$$

As a consequence of Theorem 2 and Theorem 4 from [78] d_* is locally equivalent with d_∞. Since $d_c(x,y) \le d_\infty(x,y)$ for every x and y, one gets that d_c is locally dominated from above by d_*. To conclude we need to prove that d_* is locally dominated by d.

Let us be more precise: we fix $x \in \Omega$ and we look for two constants $C_x, R_x > 0$ such that the following holds: if $0 < R \le R_x$ and $d(x,y) \le \sqrt{R}$, then exists a control $\theta \in \bar{C}_A(x,y)$ such that $|\overline{R}^{-1}\theta| < C_x$. This implies $d_*(x,y) \le C_x\sqrt{R}$, and the statement holds. Notice that we discuss local equivalence, and that is why we can take C_x, R_x depending on x.

$d(x,y) \le \sqrt{R}$ means $|x-y|_{A_R(x)} \le 1$, and this also implies $|x-y| \le C\sqrt{R}$. We look for θ such that $v_1(\theta)$ in (4.7.9) is y. We define

$$\Phi(\theta) = \int_0^t A(v_s(\theta))\theta ds = A(x)\theta + r(\theta)$$

with $r(\theta) = \int_0^t (A(v_s(\theta)) - A(x))\theta ds$. With this notation, we look for θ such that $\Phi(\theta) = y - x$. We introduce now the Moore-Penrose pseudoinverse of $A(x)$: $A(x)^+ = A(x)^T(AA^T(x))^{-1}$. The idea here is to use it as in the least squares problem, but we need some computations to overcome the fact that we are in a non-linear setting. We use the following properties: $AA(x)^+ = Id$, $|x-y|_{A(x)} = |A(x)^+(x-y)|$. Write $\theta = A(x)^+\gamma$, $\gamma \in \mathbb{R}^d$. This implies $A(x)\theta = \gamma$, and so

$$\Phi(A(x)^+\gamma) = \gamma + r(A(x)^+\gamma) = x - y.$$

Remark $r(0) = 0$, $\nabla r(0) = 0$, $|r(\theta)| \leq C_x|\theta|^2$. From local inversion theorem, there exists $l_x \in \mathcal{C}_x$ an a diffeomorphism from $B(0, l_x)$ to a neighborhood of 0 such that $|\gamma| \leq 2|x-y|$. Remark that $|x-y| \leq C_x\sqrt{R}$, and l_x is uniform in R for $R \downarrow 0$. Now, using (4.7.2)

$$|r(A(x)^+\gamma)|_{A_R(x)} \leq \frac{C_x|r(A(x)^+\gamma)|}{R} \leq C_x\frac{|A(x)^+\gamma|^2}{R} \leq C_x\frac{|x-y|^2}{R} \leq C_x|x-y|^2_{A_R(x)}.$$

Since $\gamma = x - y - r(A(x)^+\gamma)$,

$$|\gamma|_{A_R(x)} \leq |x-y|_{A_R(x)} + C_x|x-y|^2_{A_R(x)} \leq C_x,$$

(using $|x-y|_{A_R(x)} \leq 1$). We have $|\overline{R}^{-1}\theta| = |\overline{R}^{-1}A(x)^+\gamma|$. Since $A_R^+A_R(x) = A_R^+(x)A(x)\overline{R}$ is an orthogonal projection and $AA^+(x)$ is the identity,

$$|\overline{R}^{-1}\theta| = |A_R^+(x)A(x)\overline{R}\overline{R}^{-1}A(x)^+\gamma| = |A_R^+(x)\gamma| = |\gamma|_{A_R(x)}.$$

So $|\overline{R}^{-1}\theta| \leq C_x$, and as we said before this implies $d_*(x,y) \leq C_x\sqrt{R}$.
C. Standard □

Part II

Scaling properties of stochastic volatility models

Chapter 5

A multivariate model for financial indices

5.1 Introduction

This chapter is based on [22], with Bonino and Camelia. We consider a process for the detrended log-price given by $dX_t = \sigma_t dB_t$, where B is a Brownian motion and the volatility σ is the square root of the stationary solution of a SDE of the following form:

$$d\sigma_t^2 = -f(\sigma_t^2)dt + dL_t, \tag{5.1.1}$$

and in this chapter L is a pure jump process (see [50], [67], [12]). We work with the model presented in [1], (inspired by [4] and [93]), considering a bivariate version and an algorithm for the detection of shocks in the market (peaks in the volatility profile). These two aspects are linked by the fact that the cross-correlation between two indices is in fact the correlation between the two time changes, and those are highly dependent on the jumps of the volatility process.

In section 5.2 we present the univariate model introduced in [1], recalling main properties and some proofs. In section 5.3 we consider the bivariate version, defining the joint process of shocks through correlated Poisson point processes. This is a main ingredient in our modeling since the long range dependence heavily relies on the jumps of the volatility process. Defining

$$\begin{aligned} dX_t &= \sigma_t^X dB_t^X, & d(\sigma_t^X)^2 &= -f((\sigma_t^X)^2)dt + dL_t^X, \\ dY_t &= \sigma_t^Y dB_t^Y, & d(\sigma_t^Y)^2 &= -f((\sigma_t^Y)^2)dt + dL_t^Y, \end{aligned}$$

it is easy to show under mild hypothesis on the volatility $(\sigma_t^X, \sigma_t^Y)_t$ that

$$\lim_{h\downarrow 0} corr(|X_h - X_0|, |Y_{t+h} - Y_t|) = corr(\sigma_0^X, \sigma_t^Y).$$

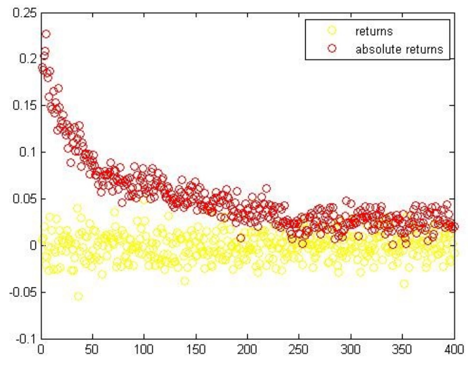
If the volatilities are of the precise form considered in [1], explicit computations are possible and the evolution of (σ^X, σ^Y) depends just on the jumps of L^X and L^Y. We consider the cross-correlation of absolute increments at different times, and compute how this correlation decays as the time difference increases. This issue has been addressed by Podobnik et al. in [83], [84], [97]. In our framework we find this explicit formula for the decay of cross-asset correlations between absolute returns depending on the time lag, analogous to the formula for the decay of autocorrelations (see Corollary 5.9):

$$\lim_{h\downarrow 0} \frac{Cov(|X_h - X_0|, |Y_{t+h} - Y_t|)}{h} = \frac{4}{\pi}\bar{\sigma}^X \bar{\sigma}^Y \sqrt{D^X D^Y} \left(\lambda^X\right)^{1/2-D^X} \left(\lambda^Y\right)^{1/2-D^Y} \times Cov\left((S^X)^{D^X-1/2}, (\lambda^Y t + S^Y)^{D^Y-1/2}\right) e^{-\lambda^Y t}$$

The quantities involved are constant parameters of the volatilities σ^X and σ^Y, except from S^X and S^Y which are correlated exponential variables coming from the jump process $L = (L^X, L^Y)$.

In section 5.4 we apply this result to the time series of the Dow Jones Industrial Average (DJIA) and the Financial Times Stock Exchange (FTSE) 100, from 1984 to 2013, finding an excellent agreement between predictions of the model and empirical findings. In particular we find that in both modeling and empirical data the decay of autocorrelations and cross-correlations is almost coincident, and it is slow over time, confirming that this is a long-memory processes. On the other hand, we empirically find a non-significant cross-correlation between returns of FTSE and DJIA, even for very small time lags, and this is consistent with the model as well. This is not surprising, since for both indices there are no long-range autocorrelations of returns, and this is easily seen to be consistent with our model. In contrast, as already said, the decay of cross-correlation of absolute returns is very slow (see Fig. 5.1).

Figure 5.1: Decay of DJA-FTSE cross-correlations



Since cross-correlation is highly dependent on the jumps of the volatility process, we propose here an algorithm for the detection of such jumps. The problem of finding shocks in financial time series is a classical one. For example, GARCH models (Generalized Autoregressive Conditional Heteroskedasticity, [21]) are widely used, but in practice "volatility seems to behave more like a jump process, where it fluctuates around some value for an extended period of time, before undergoing an abrupt change, after which if fluctuates around a new value" (see [87]). To adress this issue, regime-switching GARCH models have been developed (see [57], [60]), but they can be hard to implement. Therefore, a common approach is to use an approximate procedure, the so-called ICSS-GARCH algorithm, introduced in [65]. This algorithm is similar to the algorithm that we present because they both use squared returns to detect volatility shocks. However, the ICSS-GARCH algorithm works well under the assumption that the returns are normally distributed. Our algorithm, on the contrary, does not need any particular assumption on the distribution of the returns, but is simply based on geometrical considerations.

In section 5.4 we prove formally some results justifying the convergence of the algorithm for the detection of shocks in the volatility. We stick to this precise model for the linearity of the exposition but the same proof would give an analogous result for a wider class of volatilities solving (5.1.1). Some heuristic considerations on the output of the algorithm confirm its validity in the detection of jumps. We use it on the two empirical time series finding that the majority of the peaks of the volatility are shared by the two indices. This is a motivation to our choice to consider two processes of shocks with a common part.

5.2 Definition and properties of the univariate model

In this chapter we describe the model and state properties and results related to stylized facts.

5.2.1 Definition of the univariate model

Given three real numbers $D \in (0, 1/2]$, $\lambda \in (0, \infty)$, $\bar{\sigma} \in (0, \infty)$, the model is defined upon two sources of randomness:

- a Brownian motion $W = (W_t)_{t \geq 0}$;
- a Poisson point process $\mathcal{T} = (\tau_n)_{n \in \mathbb{Z}}$ of rate λ on $\mathbb{R}$.

We suppose W and $\mathcal{T}$ independent. By convention we label the points of $\mathcal{T}$ so that $\tau_0 < 0 < \tau_1$. For $t \geq 0$, we define

$$i(t) := \sup\{n \geq 0 : \tau_n \leq t\} = \#\{\mathcal{T} \cap (0, t]\}.$$

$i(t)$ is the number of positive times in the Poisson process before t, so that $\tau_{i(t)}$ is the location of the last point in $\mathcal{T}$ before t. We introduce the process $I = (I_t)_{t \geq 0}$ defining

$$I_t = \bar{\sigma}^2 \left[(t - \tau_{i(t)})^{2D} + \sum_{k=1}^{i(t)} (\tau_k - \tau_{k-1})^{2D} - (-\tau_0)^{2D} \right] \tag{5.2.1}$$

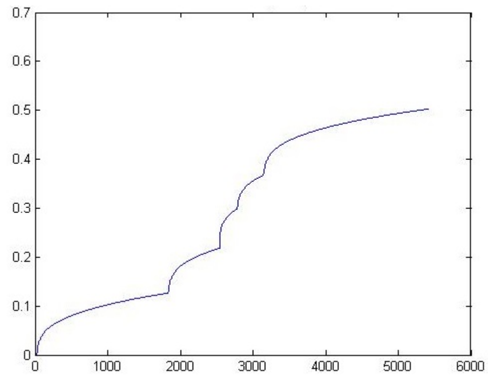
where we agree that the sum in the right hand side is 0 if $i(t) = 0$. Now we define the process which is the model for the detrended log price as

$$X_t = W_{I(t)}. \tag{5.2.2}$$

Observe that I is a strictly increasing process with absolutely continuous paths, and it is independent of the Brownian motion W. Thus this model may be viewed as an independent random time change of a Brownian motion.

We shortly give a motivation for this definition. Remark that for $D = 1/2$ the model reduces to Black & Scholes with volatility $\bar{\sigma}$. For $D < 1/2$, the introduction of a time inhomogeneity $t \to t^{2D}$ at times τ_n is meant to represent the *trading time* of a financial time series, where at "random" times there are shocks in the market, modeled by our Poisson point process. The reaction of the market is an acceleration of the dynamics immediately after the shock, and a gradual slowing down at later times, until a new shock accelerates the dynamics again. This behavior is due to the shape of the function $t \to t^{2D}$, $D \in (0, 1/2]$, which is steep for t close to 0 and bends down for increasing t.

Figure 5.2: Time inhomogeneity



The definition of the model as a time changed Brownian Motion implies that we can equivalently express it as a a stochastic volatility model, where the volatility is

$$\sigma_t = \sqrt{I'(t)} = \sqrt{2D}\bar{\sigma}(t - \tau_{i(t)})^{D-1/2},$$

and the evolution of X is given by $dX_t = \sigma_t dB_t$. To write the volatility as solution of a stochastic differential equation of the form (5.1.1), we can define think of it as the the stationary solution of

$$d(\sigma_t^2) = -\alpha(\sigma_t^2)^\gamma dt + \infty\, di(t),$$

where the constants are

$$\gamma = 2 + \frac{2D}{1 - 2D} > 2, \quad \alpha = \frac{1 - 2D}{(2D)^{1/(1-2D)}} \frac{1}{\bar{\sigma}^{2/(1-2D)}}.$$

This process is well defined, since after the infinite jumps the super-linear drift term instantaneously produces a finite pathwise solution. We refer to Chapter 6 for the details of the correspondence between time change and stochastic volatility in this framework, for a wider class of stochastic volatility models.

Remark 5.1. In the most general version of this model the parameter $\bar{\sigma}$ is not constant. A sequence of random variables $(\bar{\sigma}_n)_{n\in\mathbb{N}}$ is simulated, and each of them is associated to the corresponding jump. The results presented here are still valid in this case, with a slightly more complicated formulation. We have decided to assume $\bar{\sigma}$ constant in this work, since calibration on data coming from financial time series leads in any case to this type of choice.

5.2.2 Main properties

We briefly recall some properties of the process X. For proofs, more detailed statements and some additional considerations we refer to [1].

Proposition 5.2 (Basic Properties)**.** *Let X be the process defined in* (5.2.2)*. The following assertions are satisfied:*

(1) *X has stationary increments.*

(2) *X is a zero-mean, continuous, square-integrable martingale, with quadratic variation $\langle X\rangle_t = I_t$.*

(3) *The distributions of the increments of X is ergodic.*

(4) $\mathbb{E}(|X_t|^q) < \infty$ *for some (and hence any)* $t > 0$, $q \in [0, \infty)$.

We are now ready to state some results, important because they establish a link between our model and the stylized fact presented in the introduction. The process X defined in (5.2.2) is consistent with important facts empirically detected in many (financial) real time series, namely: diffusive scaling of returns, multiscaling of moments, slow decay of volatility autocorrelation.

The first result, proved in section 5.2.3, shows that the increments $(X_{t+h} - X_t)$ have an approximate diffusive scaling both when $h \downarrow 0$, with a heavy-tailed limit distribution, and when $h \uparrow \infty$, with a normal limit distribution. This is a precise mathematical formulation of a crossover phenomenon in the log-return distribution, from approximately heavy-tailed (for small time) to approximately Gaussian (for large time).

Theorem 5.3 (Diffusive scaling)**.** *The following convergences in distribution hold for any choice of the parameters $D, \lambda, \bar{\sigma}$.*

- *Small-time diffusive scaling:*

$$\frac{(X_{t+h} - X_t)}{\sqrt{h}} \xrightarrow[h\downarrow 0]{d} f(x)\,\mathrm{d}x \;:=\; \textit{law of}\;\; \bar{\sigma}\big(\sqrt{2D}\,\lambda^{\frac{1}{2}-D}\big)\, S^{D-\frac{1}{2}}\, W_1\,, \qquad (5.2.3)$$

where $S \sim Exp(1)$ and $W_1 \sim \mathcal{N}(0,1)$ are independent random variables:

$$f(x) = \int_0^\infty dt\, \lambda e^{-\lambda t} \frac{t^{1/2-D}}{\bar{\sigma}\sqrt{4D\pi}} exp\left(-\frac{t^{1-2D}x^2}{4D\bar{\sigma}^2}\right).$$

- *Large-time diffusive scaling:*

$$\frac{(X_{t+h} - X_t)}{\sqrt{h}} \xrightarrow[h\uparrow\infty]{d} \frac{e^{-x^2/(2c^2)}}{\sqrt{2\pi}c}\,\mathrm{d}x = \mathcal{N}(0,c^2), \qquad c^2 = \bar{\sigma}^2\,\lambda^{1-2D}\,\Gamma(2D+1), \tag{5.2.4}$$

where $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$ denotes Euler's Gamma function.

The density f, when $D < \frac{1}{2}$, has *power-law tails*:

$$\mathbb{E}_f(|x|^q) = \infty \Leftrightarrow q \geq q^* := (1/2 - D)^{-1}.$$

The function f, which describes the asymptotic law, for $h \downarrow 0$, of $\frac{X_{t+h}-X_t}{\sqrt{h}}$, has a different tail behavior from the density of $X_{t+h} - X_t$, for fixed h (cf. Proposition 5.2 point 4). This feature of f is linked to another property of our model: the multiscaling of moments. Let us define the $q - th$ moment of the log returns, at time scale h:

$$m_q(h) := \mathbb{E}\left(|X_{t+h} - X_t|^q\right) = \mathbb{E}\left(|X_h|^q\right)$$

the last equality holding for the stationarity of increments. Because of the diffusive scaling properties (Theorem 5.3), we would expect $m_q(h)$ to approximate in some sense $h^{\frac{q}{2}} \int x^q f(x)dx = C_q\, h^{\frac{q}{2}}$, for $h \downarrow 0$. This is actually true for $q < q^*$, that is, for q such that the $q - th$ moment of the limit distribution is finite. For $q \geq q^*$, the $q - th$ moment of the limit distribution is not finite, and it turns out that a faster scaling holds, namely $m_q(h) \approx h^{Dq+1}$. This transition in the scaling of $m_q(h)$ is known as multiscaling of moments, a property empirically detected in many time series, in particular in financial series. The following theorem states that for this model the multiscaling exponent is a piecewise linear function of q. In chapter 6 the problem of multiscaling in more general stochastic volatility models is considered, finding that an analogous behavior is common to a much wider class.

Theorem 5.4 (Multi-scaling of moments)**.** *For $q > 0$ the $q-th$ moment of log returns $m_q(h)$ has the following asymptotic behavior as $h \downarrow 0$:*

$$m_q(h) \sim \begin{cases} C_q h^{\frac{q}{2}}, & if\, q < q^* \\ C_q h^{\frac{q}{2}} \log\left(\frac{1}{h}\right), & if\, q = q^* \\ C_q h^{Dq+1}, & if\, q > q^* \end{cases}$$

The constant $C_q \in (0, \infty)$ *is given by*

$$C_q := \begin{cases} \mathbb{E}(|W_1|^q)\, \bar{\sigma}^q\, \lambda^{q/q^*}\, (2D)^{q/2}\, \Gamma(1 - q/q^*) & \text{if } q < q^* \\ \mathbb{E}(|W_1|^q)\, \bar{\sigma}^q\, \lambda\, (2D)^{q/2} & \text{if } q = q^* \\ \mathbb{E}(|W_1|^q)\, \bar{\sigma}^q\, \lambda \left[\int_0^\infty ((1+x)^{2D} - x^{2D})^{\frac{q}{2}}\, \mathrm{d}x + \frac{1}{Dq+1} \right] & \text{if } q > q^* \end{cases}, \tag{5.2.5}$$

where $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x$ *denotes Euler's Gamma function. As a consequence, the scaling exponent* $A(q)$ *is*

$$A(q) = \lim_{h \downarrow 0} \frac{\log m_q(h)}{\log h} = \begin{cases} \dfrac{q}{2} & \text{if } q \le q^* \\ Dq + 1 & \text{if } q \ge q^* \end{cases}$$

We now state a result concerning the volatility autocorrelation of the process X, that is the correlations of absolute values of returns at a given time distance. Recall that the correlation coefficient of two random variables X and Y is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}}.$$

For the process X, introduce $\xi = (\xi_t)_{t \ge 0}$, the process of absolute values of increments, for h fixed: $\xi_t = |X_{t+h} - X_t|$. Then the volatility autocorrelation of X is

$$\rho(t - s) = \lim_{h \downarrow 0} \rho(\xi_s, \xi_t) = \frac{Cov(\xi_s, \xi_t)}{\sqrt{Var(\xi_s) Var(\xi_t)}}$$

Indeed, being the process stationary, the quantity we have defined above depends just on the time difference $t - s$. Let's state our result, concerning as above the asymptotic behavior as $h \downarrow 0$.

Theorem 5.5 (Volatility autocorrelaton)**.** *For* $t \ge 0$*,*

$$\rho(t) = \frac{2}{\pi} \frac{Cov\left(S^{D-1/2}, (\lambda t + S)^{D-1/2}\right)}{Var(|N| S^{D-1/2})} e^{-\lambda t}$$

where S *is an exponential variable with parameter* 1*,* N *is a standard normal variable and they are mutually independent.*

This theorem is actually a special case of Corollary 5.9. It shows that the decay of volatility autocorrelation is between polynomial and exponential for $t = O(1/\lambda)$, exponential for $t >> 1/\lambda$.

5.2.3 Scaling and multiscaling: proof of Theorem 5.3 and 5.4

We prove the scaling properties of our model. Recall that for all fixed $t, h > 0$ we have the equality in law $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, as it follows by the definition of $(X_t)_{t \ge 0} = (W_{I_t})_{t \ge 0}$. We also observe that $i(h) = \#\{\mathcal{T} \cap (0, h]\} \sim Po(\lambda h)$, as it follows from the properties of the Poisson process.

Proof of Theorem 5.3: Since $\mathbb{P}(i(h) \ge 1) = 1 - e^{-\lambda h} \to 0$ as $h \downarrow 0$, we may focus on the event $\{i(h) = 0\} = \{\mathcal{T} \cap (0, h] = \emptyset\}$, on which we have $I_h = \bar{\sigma}^2((h - \tau_0)^{2D} - (-\tau_0)^{2D})$, with $-\tau_0 \sim Exp(\lambda)$. In particular,

$$\lim_{h \downarrow 0} \frac{I_h}{h} \;=\; I'(0) \;=\; 2D\bar{\sigma}^2\,(-\tau_0)^{2D-1} \qquad \text{a.s.}\,.$$

Since $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, the convergence in distribution (5.2.3) follows:

$$\frac{X_{t+h} - X_t}{\sqrt{h}} \xrightarrow{d} \sqrt{2D}\,\bar{\sigma}\,(-\tau_0)^{D-1/2}\, W_1 \qquad \text{as } h \downarrow 0\,.$$

Next we focus on the case $h \uparrow \infty$. The random variables $\{(\tau_k - \tau_{k-1})^{2D}\}_{k \ge 1}$ are independent and identically distributed with finite mean, hence by the strong law of large numbers

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (\tau_k - \tau_{k-1})^{2D} \;=\; \mathbb{E}((\tau_1)^{2D}) \;=\; \lambda^{-2D}\,\Gamma(2D+1) \qquad \text{a.s.}\,.$$

Plainly, $\lim_{h \to +\infty} i(h)/h = \lambda$ a.s., by the strong law of large numbers applied to the random variables $\{\tau_k\}_{k \ge 1}$. Recalling (5.2.1), it follows easily that

$$\lim_{h \uparrow \infty} \frac{I(h)}{h} \;=\; \bar{\sigma}^2 \lambda^{1-2D}\,\Gamma(2D+1) \qquad \text{a.s.}\,.$$

Since $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, we obtain the convergence in distribution

$$\frac{X_{t+h} - X_t}{\sqrt{h}} \xrightarrow{d} \sqrt{\bar{\sigma}^2\,\lambda^{1-2D}\,\Gamma(2D+1)}\, W_1 \qquad \text{as } h \uparrow \infty\,,$$

which coincides with (5.2.4). □

Proof of Theorem 5.4: Since $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, we can write

$$\mathbb{E}(|X_{t+h} - X_t|^q) = \mathbb{E}(|I_h|^{q/2} |W_1|^q) = \mathbb{E}(|W_1|^q)\,\mathbb{E}(|I_h|^{q/2}) = c_q\,\mathbb{E}(|I_h|^{q/2})\,, \tag{5.2.6}$$

where we set $c_q := \mathbb{E}(|W_1|^q)$. We therefore focus on $\mathbb{E}(|I_h|^{\frac{q}{2}})$, that we write as the sum of three terms, that will be analyzed separately:

$$\mathbb{E}(|I_h|^{\frac{q}{2}}) = \mathbb{E}(|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=0\}}) + \mathbb{E}(|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=1\}}) + \mathbb{E}(|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)\ge 2\}})\,. \tag{5.2.7}$$

For the first term in the right hand side of (5.2.7), we note that $\mathbb{P}(i(h) = 0) = e^{-\lambda h} \to 1$ as $h \downarrow 0$ and that $I_h = \bar{\sigma}^2((h - \tau_0)^{2D} - (-\tau_0)^{2D})$ on the event $\{i(h) = 0\}$. Setting $-\tau_0 =: \lambda^{-1} S$ with $S \sim Exp(1)$, we obtain as $h \downarrow 0$

$$\mathbb{E}(|I_h|^{\frac{q}{2}} \mathbf{1}_{\{i(h)=0\}}) = \bar{\sigma}^q \, \lambda^{-Dq} \, \mathbb{E}\big(((S + \lambda h)^{2D} - S^{2D})^{\frac{q}{2}}\big) \big(1 + o(1)\big) \,. \tag{5.2.8}$$

Recalling that $q^* := (\frac{1}{2} - D)^{-1}$, we have

$$q \gtreqless q^* \quad \iff \quad \frac{q}{2} \gtreqless Dq + 1 \quad \iff \quad -1 \gtreqless \left(D - \frac{1}{2}\right) q \,.$$

As $\delta \downarrow 0$ we have $\delta^{-1}((S + \delta)^{2D} - S^{2D}) \uparrow 2D \, S^{2D-1}$ and note that $\mathbb{E}\big(S^{(D-\frac{1}{2})q}\big) = \Gamma(1 - q/q^*)$ is finite if and only if $(D - \frac{1}{2})q > -1$, that is $q < q^*$. Therefore the monotone convergence theorem yields

$$\text{for } q < q^* : \quad \lim_{h \downarrow 0} \frac{\mathbb{E}\big(((S + \lambda h)^{2D} - S^{2D})^{\frac{q}{2}}\big)}{\lambda^{\frac{q}{2}} \, h^{\frac{q}{2}}} \;=\; (2D)^{q/2} \, \Gamma(1 - q/q^*) \;\in\; (0, \infty) \,. \tag{5.2.9}$$

Next observe that, by the change of variables $s = (\lambda h) x$, we can write

$$\begin{aligned} \mathbb{E}\big(((S + \lambda h)^{2D} - S^{2D})^{\frac{q}{2}}\big) &= \int_0^\infty ((s + \lambda h)^{2D} - s^{2D})^{\frac{q}{2}} \, e^{-s} \, \mathrm{d}s \\ &= (\lambda h)^{Dq+1} \int_0^\infty ((1 + x)^{2D} - x^{2D})^{\frac{q}{2}} \, e^{-\lambda h x} \, \mathrm{d}x \,. \end{aligned} \tag{5.2.10}$$

Note that $((1+x)^{2D} - x^{2D})^{\frac{q}{2}} \sim (2D)^{\frac{q}{2}} x^{(D-\frac{1}{2})q}$ as $x \to +\infty$ and that $(D - \frac{1}{2})q < -1$ if and only if $q > q^*$. Therefore, again by the monotone convergence theorem, we obtain

$$\text{for } q > q^* : \quad \lim_{h \downarrow 0} \frac{\mathbb{E}\big(((S + \lambda h)^{2D} - S^{2D})^{\frac{q}{2}}\big)}{\lambda^{Dq+1} \, h^{Dq+1}} \;=\; \int_0^\infty ((1 + x)^{2D} - x^{2D})^{\frac{q}{2}} \, \mathrm{d}x \;\in\; (0, \infty) \,. \tag{5.2.11}$$

Finally, in the case $q = q^*$ we have $((1+x)^{2D} - x^{2D})^{q^*/2} \sim (2D)^{q^*/2} \, x^{-1}$ as $x \to +\infty$ and we want to study the integral in the second line of (5.2.10). Fix an arbitrary (large) $M > 0$ and note that, integrating by parts and performing a change of variables, as $h \downarrow 0$ we have

$$\begin{aligned} \int_M^\infty \frac{e^{-\lambda h x}}{x} \, \mathrm{d}x \;&=\; -\log M e^{-\lambda h M} \;+\; \lambda h \int_M^\infty (\log x) \, e^{-\lambda h x} \, \mathrm{d}x \;=\; O(1) + \int_{\lambda h M}^\infty \log\left(\frac{y}{\lambda h}\right) e^{-y} \, \mathrm{d}y \\ &=\; O(1) + \int_{\lambda h M}^\infty \log\left(\frac{y}{\lambda}\right) e^{-y} \, \mathrm{d}y \;+\; \log\left(\frac{1}{h}\right) \int_{\lambda h M}^\infty e^{-y} \, \mathrm{d}y \;=\; \log\left(\frac{1}{h}\right) \big(1 + o(1)\big) \,. \end{aligned}$$

From this it is easy to see that as $h \downarrow 0$

$$\int_0^\infty ((1 + x)^{2D} - x^{2D})^{\frac{q^*}{2}} \, e^{-\lambda h x} \, \mathrm{d}x \;\sim\; (2D)^{\frac{q^*}{2}} \log\left(\frac{1}{h}\right) \,.$$

Coming back to (5.2.10), noting that $Dq+1 = \frac{q}{2}$ for $q = q^*$, it follows that

$$\lim_{h \downarrow 0} \frac{\mathbb{E}\big(((S+h)^{2D} - S^{2D})^{\frac{q^*}{2}}\big)}{\lambda^{Dq^*+1}\, h^{\frac{q^*}{2}} \, \log(\frac{1}{h})} = (2D)^{\frac{q^*}{2}} \,. \tag{5.2.12}$$

Recalling (5.2.6) and (5.2.8), the relations (5.2.9), (5.2.11) and (5.2.12) show that the first term in the right hand side of (5.2.7) has the same asymptotic behavior as in the statement of the theorem, except for the regime $q > q^*$ where the constant does not match (the missing contribution will be obtained in a moment).

We now focus on the second term in the right hand side of (5.2.7). Note that, conditionally on the event $\{i(h) = 1\} = \{\tau_1 \le h, \tau_2 > h\}$, we have

$$I_h = \bar\sigma^2\big((h-\tau_1)^{2D} + (\tau_1-\tau_0)^{2D} - (-\tau_0)^{2D}\big) \sim \bar\sigma^2\bigg((h-hU)^{2D} + \Big(hU + \frac{S}{\lambda}\Big)^{2D} - \Big(\frac{S}{\lambda}\Big)^{2D}\bigg)\,,$$

where $S \sim Exp(1)$ and $U \sim U(0,1)$ (uniformly distributed on the interval $(0,1)$). Since $\mathbb{P}(i(h) = 1) = \lambda h + o(h)$ as $h \downarrow 0$, we obtain

$$\mathbb{E}(|I_h|^{\frac{q}{2}}\, \mathbf{1}_{\{i(h)=1\}}) = \lambda \bar\sigma^q h^{Dq+1}\, \mathbb{E}\Bigg[\bigg((1-U)^{2D} + \bigg(\Big(U + \frac{S}{\lambda h}\Big)^{2D} - \Big(\frac{S}{\lambda h}\Big)^{2D}\bigg)\bigg)^{\frac{q}{2}}\Bigg]\,. \tag{5.2.13}$$

Since $(u+x)^{2D} - x^{2D} \to 0$ as $x \to \infty$, for every $u \ge 0$, by the dominated convergence theorem we have (for every $q \in (0,\infty)$)

$$\lim_{h \downarrow 0} \frac{\mathbb{E}(|I_h|^{\frac{q}{2}}\, \mathbf{1}_{\{i(h)=1\}})}{h^{Dq+1}} \;=\; \lambda \bar\sigma^q\, \mathbb{E}\big((1-U)^{Dq}\big) \;=\; \frac{\lambda \bar\sigma^q}{Dq+1}\,. \tag{5.2.14}$$

This shows that the second term in the right hand side of (5.2.7) gives a contribution of the order h^{Dq+1} as $h \downarrow 0$. This is relevant only for $q > q^*$, because for $q \le q^*$ the first term gives a much bigger contribution of the order $h^{q/2}$ (see (5.2.9) and (5.2.12)). Recalling (5.2.6), it follows from (5.2.14) and (5.2.11) that the contribution of the first and the second term in the right hand side of (5.2.7) matches the statement of the theorem (including the constant).

It only remains to show that the third term in the right hand side of (5.2.7) gives a negligible contribution. We begin by deriving a simple upper bound for I_h. Since $(a+b)^{2D} - b^{2D} \le a^{2D}$ for all $a, b \ge 0$ (we recall that $2D \le 1$), when $i(h) \ge 1$, i.e. $\tau_1 \le h$, we can write

$$\begin{aligned} I_h &= \bar\sigma^2 \left[(h - \tau_{i(h)})^{2D} + \sum_{k=2}^{i(h)} (\tau_k - \tau_{k-1})^{2D} + \big[(\tau_1 - \tau_0)^{2D} - (-\tau_0)^{2D}\big] \right] \\ &\le \bar\sigma^2 \left[(h - \tau_{i(h)})^{2D} + \sum_{k=2}^{i(h)} (\tau_k - \tau_{k-1})^{2D} + \tau_1^{2D} \right]\,, \end{aligned} \tag{5.2.15}$$

where we agree that the sum over k is zero if $i(h) = 1$. Since $\tau_k \leq h$ for all $k \leq i(h)$, by the definition of $i(h)$, relation (5.2.15) yields the bound $I_h \leq \bar{\sigma}^2 h^{2D}(i(h)+1)$, which holds clearly also when $i(h) = 0$. In conclusion, we have shown that for all $h, q > 0$

$$|I_h|^{q/2} \leq h^{Dq} \bar{\sigma}^q (i(h) + 1)^{q/2}. \tag{5.2.16}$$

For any fixed $a > 0$, by the Hölder inequality with $p = 3$ and $p' = 3/2$ we can write for $h \leq 1$

$$\begin{aligned} \mathbb{E}\big((i(h) + 1)^{q/2} \, \mathbf{1}_{\{i(h)\geq 2\}}\big) \; &\leq \; \mathbb{E}\big((i(h) + 1)^{3q/2}\big)^{1/3} \, \mathbb{P}(i(h) \geq 2)^{2/3} \\ &\leq \; \mathbb{E}\big((i(1) + 1)^{3q/2}\big)^{1/3} \, (1 - e^{-\lambda h} - e^{-\lambda h}\lambda h)^{2/3} \; \leq \; (const.)\, h^{4/3} \, , \end{aligned} \tag{5.2.17}$$

because $\mathbb{E}\big((i(1) + 1)^{3q/2}\big) < \infty$ (recall that $i(h) \sim Po(\lambda)$) and $(1 - e^{-\lambda h} - e^{-\lambda h}\lambda h) \sim \frac{1}{2}(\lambda h)^2$ as $h \downarrow 0$. Then it follows from (5.2.17) that

$$\mathbb{E}\big(|I_h|^{q/2} \, \mathbf{1}_{\{i(h)\geq 2\}}\big) \; \leq \; (const.')\, h^{Dq+4/3} \, .$$

This shows that the contribution of the third term in the right hand side of (5.2.7) is always negligible with respect to the contribution of the second term (recall (5.2.14)). □

5.3 The bivariate model

5.3.1 Definition of the bivariate model

We investigate here the decay of correlation of the absolute returns of a bivariate version $(X, Y) = (X_t, Y_t)_{t\geq 0}$ of the model defined in Section 5.2. We need the following quantities:

- two Brownian motions $W^X = \big(W_t^X\big)_{t\geq 0}$ and $W^Y = \big(W_t^Y\big)_{t\geq 0}$;
- two Poisson point processes on $\mathbb{R}$: $\mathcal{T}^X = (\tau_n^X)_{n\in\mathbb{Z}}$ and $\mathcal{T}^Y = (\tau_n^Y)_{n\in\mathbb{Z}}$, of rates respectively λ^X and λ^Y;
- positive constants D^X, D^Y, $\bar{\sigma}^X$ and $\bar{\sigma}^Y$.

The tricky point is the definition of the Poisson processes, that we want dependent but different. We introduce $\mathcal{T}^i$, $i = 1, 2, 3$ independent Poisson point processes with intensities λ_i, $i = 1, 2, 3$. Then we define $\mathcal{T}^X = \mathcal{T}^1 \cup \mathcal{T}^2$, $\mathcal{T}^Y = \mathcal{T}^1 \cup \mathcal{T}^3$. These are again Poisson processes, with intensity $\lambda_1 + \lambda_2$ and $\lambda_1 + \lambda_3$, and they are actually mutually dependent if $\mathcal{T}^1$ is non-degenerate.

We want to have a correlation coefficient $\rho \in [-1, 1]$ also between the Brownian motions, and this is a standard issue in financial modeling. We introduce two independent Brownian motions W^X, $\tilde{W}$, and define

$$W_t^Y = \rho W_t^X + \sqrt{1 - \rho^2}\tilde{W}_t.$$

The correlation between W^Y and W^X will play no role in this paper, but the parameter ρ is important for the correlation of the increments of X and Y at the same time, which could be an interesting aspect to consider.

We suppose that the two-dim Brownian $W = (W^X, W^Y)$ and the two-dim time change $\mathcal{T} = (\mathcal{T}^X, \mathcal{T}^Y)$ are independent. The requirements of section 5.2 on the marginal one-dim processes are satisfied and we can define X and Y as

$$X_t = W^X_{I^X_t}, \quad Y_t = W^Y_{I^Y_t}$$

where the random time changes I^X_t and I^Y_t are defined as in (5.2.1). This definition is motivated by the fact that in empirical data the occurrence of a shock in one of the two indices often coincides with a peak in the volatility of the other, as we will see in Section 5.4.2. So it is reasonable to suppose that part of the shock process is "common".

5.3.2 Covariance and correlations of absolute log-returns

For a given time h,

$$\xi_t = |X_{t+h} - X_t|, \quad \eta_t = |Y_{t+h} - Y_t|,$$

are the absolute values of the returns of X and Y at time t. We are interested in the correlations between these two variables, and we start computing their covariance. In fact, we are now going to state a result on the asymptotic behavior of the covariance of log-returns as the time scale goes to 0.

Theorem 5.6 (Covariance of absolute log-returns)**.** *Let the process* (X, Y) *be defined as above. Then, for any* $t \geq s \geq 0$*, the following holds:*

$$\lim_{h\downarrow 0} \frac{Cov(\xi_s, \eta_t)}{h} = \lim_{h\downarrow 0} \frac{Cov(\xi_0, \eta_{t-s})}{h} = \frac{4\,\bar\sigma^X \bar\sigma^Y \sqrt{D^X D^Y}}{\pi} Cov\left((-\tau_0^X)^{D^X-1/2}, (t-s-\tau_0^Y)^{D^Y-1/2}\right) e^{-\lambda^Y(t-s)}$$

Remark 5.7. Using the definition of $\mathcal{T}^X$ and $\mathcal{T}^Y$ and the properties of Poisson processes it is possible to rewrite this expression as

$$\lim_{h\downarrow 0} \frac{Cov(\xi_0, \eta_t)}{h} = \frac{4}{\pi} \bar\sigma^X \bar\sigma^Y \sqrt{D^X D^Y} \left(\lambda^X\right)^{1/2-D^X} \left(\lambda^Y\right)^{1/2-D^Y} \times Cov\left((S^X)^{D^X-1/2}, (\lambda^Y t + S^Y)^{D^Y-1/2}\right) e^{-\lambda^Y t}$$

where $S^X = \min\{S^{1,X}, S^2\}$ and $S^Y = \min\{S^{1,Y}, S^3\}$ are correlated exponential variables of parameter 1, and

$$(\lambda_1 + \lambda_2) S^{1,X} = (\lambda_1 + \lambda_3) S^{1,Y} \sim \exp(\lambda_1),$$
$$S^2 \sim \exp\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right), \quad S^3 \sim \exp\left(\frac{\lambda_3}{\lambda_1 + \lambda_3}\right)$$

are mutually independent.

Remark 5.8. If instead of taking absolute returns we consider simple returns, we find that $\lim_{h\downarrow 0} Cov(X_h - X_0, Y_{t+h} - Y_t) = 0$, for any $t > 0$. This is why we say that our model is consistent with the fact that empirical cross-correlations of returns are not significant even for very small time lags, in analogy with the autocorrelations.

From this theorem we obtain an asymptotic evaluation for correlations between log-returns, when the time scale goes to 0. Recall that the correlation coefficient between ξ_0 and η_t is defined as

$$\rho(\xi_0, \eta_t) = \rho(|X_h|, |Y_{t+h} - Y_t|) = \frac{Cov(\xi_0, \eta_t)}{\sqrt{Var(\xi_0)Var(\eta_t)}}.$$

Corollary 5.9 (Decay of cross-asset correlations)**.** *For the process (X, Y) defined above, for any $t \geq s \geq 0$, the following expression holds as $h \downarrow 0$:*

$$\lim_{h\downarrow 0} \rho(\xi_0, \eta_t) = \frac{2}{\pi} \frac{Cov\left((S^X)^{D^X-1/2}, (\lambda^Y t + S^Y)^{D^Y-1/2}\right)}{\sqrt{Var(|N|S^{D^X-1/2})Var(|N|S^{D^Y-1/2})}} e^{-\lambda^Y t}$$

where with S we denote an exponential variable of parameter 1 and with N a standard normal variable, they are mutually independent and both independent of all the other random variables. S^X and S^Y are defined in Remark 5.7.

Remark 5.10. Suppose we are dealing with X and Y produced by the same time change of two different Brownian motions, i.e $I^X = I^Y =: I$, or:

$$D^X = D^Y, \quad \mathcal{T}^X = \mathcal{T}^Y, \quad \bar{\sigma}^X = \bar{\sigma}^Y.$$

The expression for the decay of cross-asset correlation becomes in this case

$$\lim_{h\downarrow 0} \rho(\xi_0, \eta_t) = \frac{2}{\pi} \frac{Cov\left(\bar{\sigma} S^{D-1/2}, \bar{\sigma}(\lambda t + S)^{D-1/2}\right) e^{-\lambda t}}{Var(\bar{\sigma}|N|S^{D-1/2})},$$

which is exactly the expression for the decay of autocorrelation coefficients (cf. Theorem 5.5). An analysis of real data suggests that this property is very close to what we see in financial markets.

5.3.3 Proof of theorem 5.6

We start the computations on $Cov(\xi_s, \eta_t)$ writing more explicitly the quantities involved. Recall that the increments of W^X and W^Y are independent on disjoint time intervals, and W^X and $\tilde{W}$ are independent Brownian Motions. So for $h < t - s$

$$\begin{aligned} Cov(\xi_s, \eta_t) &= \mathbb{E}(|X_{s+h} - X_s||Y_{t+h} - Y_t|) - \mathbb{E}|X_{s+h} - X_s|\mathbb{E}|Y_{t+h} - Y_t| \\ &= \mathbb{E}\left(|W_1^X|\sqrt{I_{s+h}^X - I_s^X}|\tilde{W}_1|\sqrt{I_{t+h}^Y - I_t^Y}\right) \\ &\quad - \mathbb{E}\left(|W_1^X|\sqrt{I_{s+h}^X - I_s^X}\right)\mathbb{E}\left(|\tilde{W}_1|\sqrt{I_{t+h}^Y - I_t^Y}\right) \end{aligned}$$

and using independence

$$\begin{aligned} Cov(\xi_s, \eta_t) &= (\mathbb{E}|W_1^X|)^2 Cov\left(\sqrt{I_{s+h}^X - I_s^X}, \sqrt{I_{t+h}^Y - I_t^Y}\right) \\ &= \frac{2}{\pi} Cov\left(\sqrt{I_{s+h}^X - I_s^X}, \sqrt{I_{t+h}^Y - I_t^Y}\right). \end{aligned}$$

From our choice of $\mathcal{T}^X$ and $\mathcal{T}^Y$ we have the stationarity of the increments of (I^X, I^Y), therefore

$$Cov\left(\sqrt{I_{s+h}^X - I_s^X}, \sqrt{I_{t+h}^Y - I_t^Y}\right) = Cov\left(\sqrt{I_h^X}, \sqrt{I_{t-s+h}^Y - I_{t-s}^Y}\right).$$

Remark that the covariance of the absolute values of the returns actually depends just on $Cov\left(\sqrt{I_h^X}, \sqrt{I_{t+h}^Y - I_t^Y}\right)$, where t is the time difference. Recall

$$I_h = \bar{\sigma}^2 \left[(h - \tau_{i(h)})^{2D} + \sum_{k=1}^{i(h)} (\tau_k - \tau_{k-1})^{2D} - (-\tau_0)^{2D} \right]$$

Almost surely, for h small enough, $i(h) = i(0) = 0$, so the sum in the right hand vanishes and a.s.

$$\begin{aligned} \lim_{h \downarrow 0} \frac{I_h}{h} &= \lim_{h \downarrow 0} \bar{\sigma}^2 \frac{(h - \tau_{i(h)})^{2D} - (-\tau_0)^{2D}}{h} \\ &= \bar{\sigma}^2 \lim_{h \downarrow 0} \frac{(h - \tau_0)^{2D} - (-\tau_0)^{2D}}{h} = 2D\bar{\sigma}^2(-\tau_0)^{2D-1}, \end{aligned}$$

and analogously

$$\lim_{h \downarrow 0} \frac{I_{t+h} - I_t}{h} = 2D\bar{\sigma}^2 (t - \tau_{i(t)})^{2D-1}.$$

Lemma 5.11 implies the uniform integrability of the families

$$\left\{ \frac{I_h^X}{h} : h \in (0,1] \right\}, \quad \left\{ \frac{I_{t+h}^Y - I_t^Y}{h} : h \in (0,1] \right\},$$

therefore we first apply bi-linearity of covariance and then take the limit inside, obtaining

$$\begin{aligned} \lim_{h \downarrow 0} \frac{Cov\left(\sqrt{I_h^X}, \sqrt{I_{t+h}^Y - I_t^Y}\right)}{h} &= Cov\left(\lim_{h \downarrow 0} \sqrt{\frac{I_h^X}{h}}, \lim_{h \downarrow 0} \sqrt{\frac{I_{t+h}^Y - I_t^Y}{h}} \right) \\ &= 2\sqrt{D^X D^Y} \bar{\sigma}^X \bar{\sigma}^Y Cov\left((-\tau_0^X)^{D^X - 1/2}, (t - \tau_{i^Y(t)}^Y)^{D^Y - 1/2} \right). \end{aligned}$$

We can obtain a better representation of this quantity multiplying the right term in the covariance by the characteristic function of $\{i^Y(t) = 0\}$ plus the characteristic function of its complement:

$$\begin{aligned}Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_{i^Y(t)}^Y)^{D^Y-1/2}\right) \\ = Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_{i^Y(t)}^Y)^{D^Y-1/2}\mathbf{1}_{\{i^Y(t)=0\}}\right) \\ + Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_{i^Y(t)}^Y)^{D^Y-1/2}\mathbf{1}_{\{i^Y(t)>0\}}\right).\end{aligned}$$

The second covariance is 0 because $(t-\tau_{i^Y(t)}^Y)^{D^Y-1/2}\mathbf{1}_{\{i^Y(t)>0\}}$ is $\mathcal{G}_{>0}^Y$ measurable, where $\mathcal{G}_{>0}^Y = \bar{\sigma}(\tau_k^Y : k > 0)$, and $\mathcal{G}_{>0}^Y$ is independent of τ_0 (loss of memory property of Poisson processes). So, using the fact that $\mathbf{1}_{\{i^Y(t)=0\}}$ is $\mathcal{G}_{>0}^Y$ measurable, because so is $\mathbf{1}_{\{i^Y(t)>0\}}$, we have

$$\begin{aligned}Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_{i^Y(t)}^Y)^{D^Y-1/2}\right) \\ = Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_0^Y)^{D^Y-1/2}\mathbf{1}_{\{i^Y(t)=0\}}\right) \\ = Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_0^Y)^{D^Y-1/2}\right)\mathbb{E}\left(\mathbf{1}_{\{i^Y(t)=0\}}\right) \\ = Cov\left((-\tau_0^X)^{D^X-1/2}, (t-\tau_0^Y)^{D^Y-1/2}\right)e^{-\lambda^Y t},\end{aligned}$$

and the theorem is proved. □

We present now the technical lemma used in the proof of Theorem 5.6. Recall that $0 < D < 1/2$.

Lemma 5.11. *The class of random variables*

$$\left\{\frac{I_h^X}{h} : h \in (0,1]\right\}$$

is bounded in L^δ *for* $\delta < \frac{1}{1-2D}$.

Proof. Recall

$$I_t = \bar{\sigma}^2\left[(t-\tau_{i(t)})^{2D} + \sum_{k=1}^{i(t)}(\tau_k - \tau_{k-1})^{2D} - (-\tau_0)^{2D}\right]$$

and decompose $\mathbb{E}(I_t^\delta)$

$$\mathbb{E}(I_t^\delta) = \mathbb{E}(I_t^\delta | i(t) = 0)\mathbb{P}(i(t) = 0) + \sum_{k=1}^{\infty}\mathbb{E}(I_t^\delta | i(t) = k)\mathbb{P}(i(t) = k)$$

Conditioning on $i(t) = 0$ and using convexity,

$$I_t = \bar{\sigma}^2\left[(t-\tau_0)^{2D} - (-\tau_0)^{2D}\right] \le 2D\bar{\sigma}^2(-\tau_0)^{2D-1}t$$

in a right neighborhood of $t = 0$. So

$$\mathbb{E}(I_t^\delta | i(t) = 0) \leq (2D)^\delta \bar{\sigma}^{2\delta} \mathbb{E}\left((-\tau_0)^{\delta(2D-1)}\right) t^\delta \leq C_0 t^\delta$$

for $\delta < \frac{1}{1-2D}$, since $-\tau_0$ is an random variable with exponential distribution. Conditioning on $i(t) = k$, $k \geq 1$, and using convexity again,

$$\begin{aligned} I_t &= \bar{\sigma}^2 \left[(t - \tau_k)^{2D} + \sum_{j=1}^{k} (\tau_j - \tau_{j-1})^{2D} - (-\tau_0)^{2D} \right] \\ &\leq \bar{\sigma}^2 \left[(t - \tau_k)^{2D} + \sum_{j=2}^{k} (\tau_j - \tau_{j-1})^{2D} + (t - \tau_0)^{2D} - (-\tau_0)^{2D} \right] \\ &\leq \bar{\sigma}^2 \left[(t - \tau_k)^{2D} + \sum_{j=2}^{k} (\tau_j - \tau_{j-1})^{2D} + 2D(-\tau_0)^{2D-1} t \right]. \end{aligned}$$

By Jensen inequality and the fact that $2D < 1$,

$$(t - \tau_k)^{2D} + \sum_{j=2}^{k} (\tau_j - \tau_{j-1})^{2D} \leq k \left(\frac{(t - \tau_k) + \sum_{j=2}^{k} (\tau_j - \tau_{j-1})}{k} \right)^{2D} \leq k \left(\frac{t}{k} \right)^{2D}$$

Then

$$I_t \leq \bar{\sigma}^2 \left(2D(-\tau_0)^{2D-1} t + k \left(\frac{t}{k} \right)^{2D} \right).$$

Now, supposing $t \leq 1$, we have that for suitable positive constants C_1 and C_2

$$\mathbb{E}(I_t^\delta | i(t) = k) \mathbb{P}(i(t) = k) \leq C_1 \frac{\lambda^k}{k!} t^\delta + C_2 k^{\delta(1-2D)} \frac{\lambda^k}{k!} t^{1+2D\delta}.$$

Recall $\delta < \frac{1}{1-2D}$. Therefore $\delta < 1 + 2D\delta$, so $t^{1+2D\delta} \leq t^\delta$, and then

$$\mathbb{E}(I_t^\delta | i(t) = k) \mathbb{P}(i(t) = k) \leq \left(C_1 + C_2 k^{\delta(1-2D)} \right) \frac{\lambda^k}{k!} t^\delta.$$

Therefore

$$\mathbb{E}(I_t^\delta) \leq \left[C_0 + \sum_{k=1}^{\infty} C_3 \frac{\lambda^k}{k!} \right] t^\delta \leq C_4 t^\delta$$

where C_3 and C_4 are positive constants. So $\left\{ \frac{I_t^X}{t} : t \in (0, 1] \right\}$ is bounded in L^δ. $\square$

5.4 Empirical results

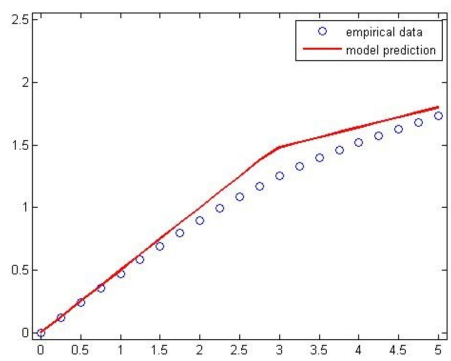
We consider the DJIA Index and FTSE Index, from April 2nd, 1984 to July 6th, 2013. For the data analysis we use the software MatLab [76]. What follows is justified by the ergodic properties of the increments X. We start considering the two series separately. We want to assign to the parameters some values such that the predictions of the model are as close as possible to real data. For this purpose, we choose some significant quantities (taking into account interesting features related to stylized facts), and use them for the calibration. Here we consider the multiscaling coefficients C_1 and C_2, the multiscaling exponent $A(q)$, the volatility autocorrelation function $\rho(t)$. The procedure for the calibration is described precisely in [1] for what concerns the one-dimensional model. Here we just outline the basic idea, which is to minimize an L^2 distance between predictions of the model and empirical estimations of these significant quantities. The details of the calibration of the bivariate version can be found in [82]. We find the following estimates for the parameters.

$$\text{FTSE:} \quad \overline{D} \approx 0.16; \quad \overline{\lambda} \approx 0.0019; \quad \overline{\overline{\sigma}} \approx 0.11.$$
$$\text{DJIA:} \quad \overline{D} \approx 0.14; \quad \overline{\lambda} \approx 0.0014; \quad \overline{\overline{\sigma}} \approx 0.127.$$

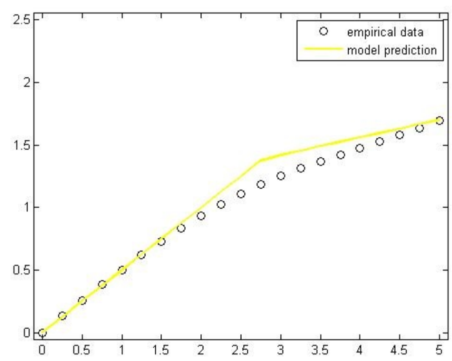
In Figure 5.3 we show the empirical multiscaling exponent versus the prediction of our model with this parameters. Our estimate for the multiscaling exponent looks smoothed out by the empirical curve. Since a simulation of daily increments of the model yields a graph analogous to the empirical one, this slight inconsistency is likely due to the fact that the theoretical line shows the limit for $h \downarrow 0$, whereas the empirical data come from a daily sample.

Analogously Figure 5.4 concerns volatility autocorrelation. The decay is between polynomial and exponential, and fits very well empirical data considering the fact that they are quite widespread. We conclude that the agreement is excellent for both multiscaling and volatility autocorrelation.

Figure 5.3: Multiscaling exponent



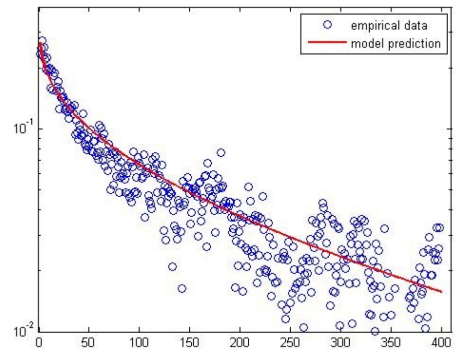
(a) FTSE



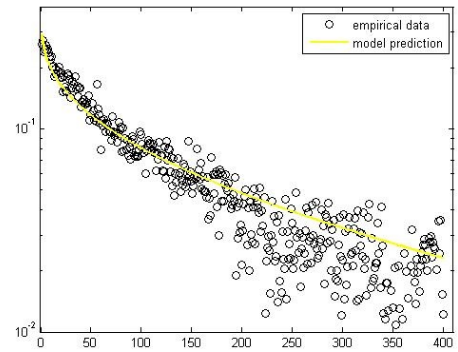
(b) DJIA

We display now the distribution of log returns for our model $p_t(\cdot) = \mathbb{P}(X_t \in \cdot) = \mathbb{P}(X_{n+t} - X_n \in \cdot)$ for $t = 1$ day, and the analogous empirical quantity. We do not have an explicit analytic expression for p_t, but we can easily obtain it numerically.

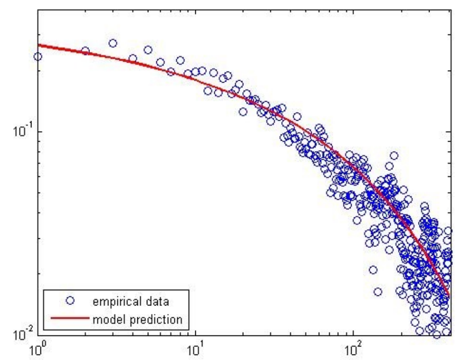
Figure 5.4: Volatility autocorrelation



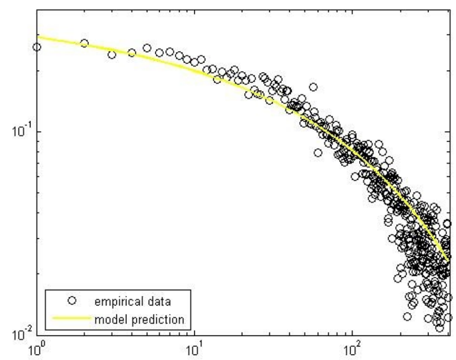
(a) FTSE log plot



(b) DJIA log plot



(c) FTSE loglog plot



(d) DJIA loglog plot

Figure (5.5) represent the bulks and the integrated tails of the distributions. We see that the agreement is remarkable, given that this curves are a test *a posteriori*, and no parameter has been estimated using these distributions!

5.4.1 Jumps and quadratic variation

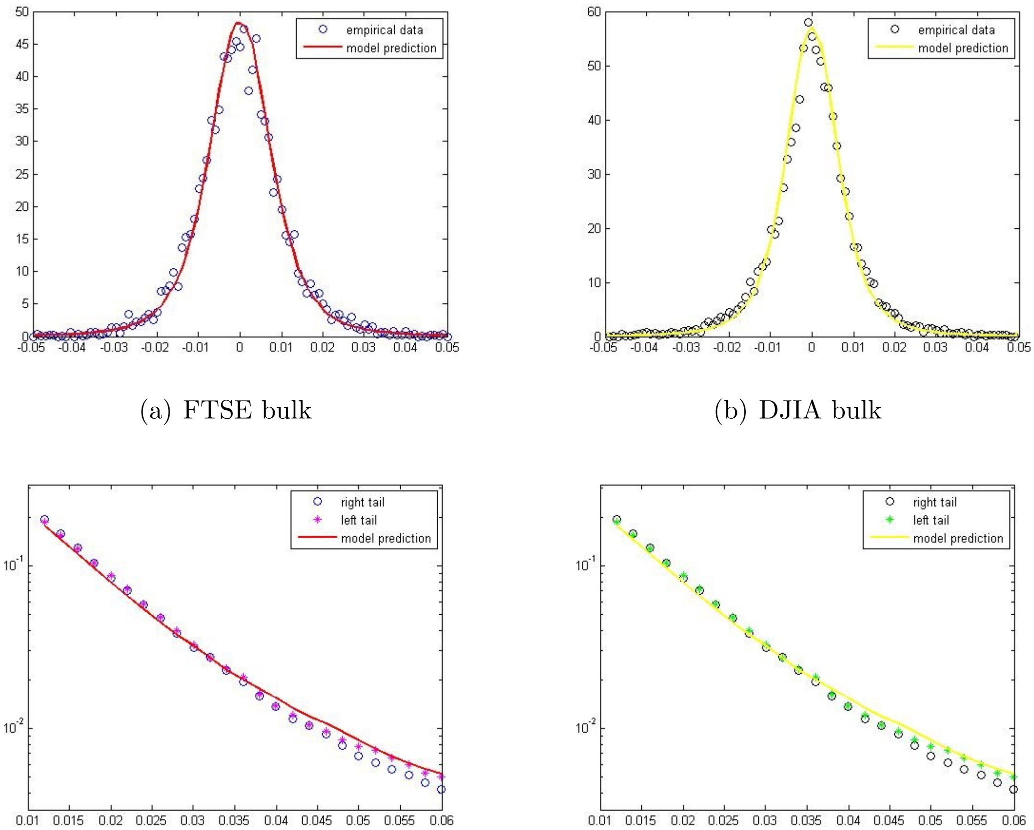
In this section we introduce the theoretical tools and results that have inspired the algorithm for finding the location of relevant big jumps in the volatility, that will be presented in the next section. This algorithm has appeared for the first time in [23].

On one hand we know that the quadratic variation of X is given by I (Proposition 5.2), i.e.

$$\langle X \rangle_t = I_t.$$

Therefore, since we know that the quadratic variation of X_t is the limit in probability of the squared increments on shrinking partitions, it seems natural to estimate I by evaluating the squared increments of a dense sampling of X.

Figure 5.5: Distribution of log returns



(a) FTSE bulk

(b) DJIA bulk

(c) FTSE integrated tails

(d) DJIA integrated tails

On the other hand, the process I is piecewise-concave; in fact, we recall that such process is defined by

$$I_t = \bar{\sigma}^2 \left[(t - \tau_{i(t)})^{2D} + \sum_{k=1}^{i(t)} (\tau_k - \tau_{k-1})^{2D} - (-\tau_0)^{2D} \right]$$

It is clear that between two consecutive shock times the process is concave. Therefore, if we are at time T and we consider the backward difference quotient defined by

$$Q_T(t) := \frac{I_T - I_{T-t}}{t}$$

this quantity is, conditional on $\mathcal{T}$, increasing up to the last shock time before T, therefore it has a local maximum in $t = T - \tau_{i(T)}$. Moreover, the derivative of I_t is very big after a shock but it quickly decays over time. Because of that, we expect that $Q_T(s) < Q_T(T - \tau_{i(T)})$ if $s \in (T - \tau_{i(T)}, T - \tau_{i(T)} - L)$, for some $L > 0$. We propose here an algorithm based on the following idea: if we choose a $M > 0$ such that $T-M < \tau_{i(T)} < T$ and $T-M$ is "closer to $\tau_{i(T)}$ than to $\tau_{i(T)-1}$", then the global maximum of $Q_T(t)$ in the interval $(T-M, T)$ should be attained in $t = T - \tau_{i(T)}$.

In view of these two observations, we introduce the following estimator

$$V_T(k) := \frac{1}{k}\sum_{i=1}^{k}(X_{T-i+1} - X_{T-i})^2$$

In our data analysis, the estimator is an average of daily squared increments of X_t (the densest sampling we could get), so it should be a good estimate of the quadratic variation of X. Moreover, it has the following property

$$\begin{aligned}
\mathbb{E}\{V_T(k)|\mathcal{T}\} &= \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\left\{(X_{T-i+1} - X_{T-i})^2|\mathcal{T}\right\} \\
&= \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\left\{\left(W_{I_{T-i+1}} - W_{I_{T-i}}\right)^2|\mathcal{T}\right\} \\
&= \frac{1}{k}\sum_{i=1}^{k}(I_T - I_{T-k}) \\
&= Q_T(k)
\end{aligned}$$

Since we can't observe Q_T directly, we can use V_T as an approximation of it and implement and algorithm for finding the realised jump time estimating V_T on historical data. We give now a more accurate mathematical explanation of these heuristics. This work has been developed in [23], [29], [82].

First we introduce the process $Q_T^{(N)}(t)$, which is an analogous of the estimator V_T, but its time argument is continuous. In order to simpify the notation we set $m = T - \tau_{i(T)}$ and $\alpha := \tau_{i(T)} - \tau_{i(T)-1}$. Moreover recall from subsection 5.4.1 that

$$Q_T(t) = \frac{I_T - I_{T-t}}{t}, \quad \text{and } I_t = \langle X\rangle_t.$$

Definition 5.12. Let X_t the stochastic process defined as W_{I_t}. We define the process $Q_T^{(N)}(t)$ as the discrete version of $Q_T(t)$.

$$Q_T^{(N)}(t) := \frac{1}{t}\left(\sum_{n=0}^{N}\left(\left(X_{\frac{(n+1)T}{N}} - X_{\frac{nT}{N}}\right)^2 - \left(X_{\frac{(n+1)(T-t)}{N}} - X_{\frac{n(T-t)}{N}}\right)^2\right)\right)$$

$Q_T(\cdot)$ has some nice geometrical properties which guarantee the existence of an "isolated" maximum point in m. However we cannot observe on real data the realization of $Q_T(\cdot)$ but we can observe the process $Q_T^{(N)}(\cdot)$ on the times where it coincides with V_T. Theorem 5.14 shows that in suitable settings maximum points observed through $Q_T^{(N)}(\cdot)$ converges to m.

The following lemma shows geometrical properties of $Q_T(\cdot)$. Given m small enough $Q_T(\cdot)$ attains its maximum in m and the peak attained in m is arbitrarily high, i.e. reducing m increases the distance between the maximum and the next minimum.

Lemma 5.13. *Let m and α as in above and $K := (\frac{1}{2D})^{\frac{1}{1-2D}}$. Then*

(1) *m is a local maximum point for $Q_T(t)$ iff $m < K\alpha$*

(2) *The following limit holds*

$$Q_T(m) - Q_T(\alpha + m) \xrightarrow{m \to 0+} +\infty$$

Proof. (1) $Q_T(t)$ is everywhere continuous and it is differentiable but in $\{T-\tau_n\}_{n\in\mathbb{N}}$. To prove that it attains the maximum in $T - \tau_{i(T)}$ we will prove that in m the left derivative is greater than 0 and the right one is less than 0.

The derivatives are

$$Q'_T(t) = \frac{\bar{\sigma}^2 2D(T - t - \tau_{i(T)-1})^{2D-1} t - (I_T - I_{T-t})}{t^2} \quad t \in (T - \tau_{i(T)}, T - \tau_{i(T)-1})$$

$$Q'_T(t) = \frac{\bar{\sigma}^2 2D(T - t - \tau_{i(T)})^{2D-1} t - (I_T - I_{T-t})}{t^2} \quad t \in (0, T - \tau_{i(T)}) \qquad (5.4.1)$$

I_s is concave then

$$I_T - I_{T-t} < I'(T - t)t$$

From (5.4.1) we get $Q'_T(t) > 0$ in $(0, m)$.

On the other hand

$$\lim_{t \to m} \frac{\bar{\sigma}^2 2D(T - t - \tau_{i(T)-1})^{2D-1} t - (I_T - I_{T-t})}{t^2} = \frac{\bar{\sigma}^2}{m^2}\left(2D\alpha^{2D-1} m - m^{2D}\right) =: L_{\bar{\sigma}}(\alpha, m) \qquad (5.4.2)$$

$L_{\bar{\sigma}}(\alpha, m)$ has the following trivial properties

$$L_{\bar{\sigma}}(\alpha, m) = 0 \Leftrightarrow m = \alpha \left(\frac{1}{2D}\right)^{\frac{1}{1-2D}}$$

$$\lim_{m \to 0^+} L_{\bar{\sigma}}(\alpha, m) = -\infty$$

$$\lim_{m \to +\infty} L_{\bar{\sigma}}(\alpha, m) = +\infty$$

which imply that the right derivative is less than zero if and only if $m < K\alpha$, thus $Q_T(t)$ attains a local maximum in m if and only if $m < K\alpha$.

(2) Note that $Q_T \in \mathcal{C}^\infty((T - \tau_{i(T)}, T - \tau_{i(T)-1}))$ a.s. The second order derivative on this interval is the following

$$Q_T^{(2)} = \frac{2D(2D - 1)(T - t - \tau_{i(T)-1})^{2D-2}}{t} - \frac{2Q'_T}{t}$$

Thus $Q'_T(t) = 0$ implies $Q_T^{(2)}(t) > 0$ then all stationary points are minimum points. Moreover Q_T can have only one minimum point which in fact exists, since from hypothesis and from (5.4.2) we get

$$\lim_{t \to T - \tau_{i(T)}} Q'_T(t) < 0$$

and

$$\lim_{t \to T - \tau_{i(T)-1}} Q'_T(t) = +\infty$$

Let $\gamma \in (T - \tau_{i(T)}, T - \tau_{i(T)-1})$ the point in which $Q_T(t)$ attains its minimum. By definition

$$Q_T(m) - Q_T(\gamma) > Q_T(m) - Q_T(\alpha + m)$$

Let $\xi = \tau_{i(T)-1} - \tau_{i(T)-2}$. We get

$$Q_T(m) - Q_T(\alpha + m) = \frac{I_T - I_{\tau_{i(T)}}}{T - \tau_{i(T)}} - \frac{I_T - I_{\tau_{i(T)-1}}}{T - \tau_{i(T)-1}} = \bar{\sigma}^2 \left(\frac{m^{2D} + \alpha^{2D}}{m} - \frac{m^{2D} + \alpha^{2D} + \xi^{2D}}{m + \alpha} \right)$$

Passing to the limit

$$Q_T(m) - Q_T(\alpha + m) = \frac{\bar{\sigma}^2(\alpha^{2D+1} + m^{2D}\alpha - m\xi^{2D})}{m(m + \alpha)} \xrightarrow{m \to 0+} +\infty$$

then $\lim_{m \to 0+} Q_T(m) - Q_T(\gamma) = +\infty$.

□

Theorem 5.14. *Let $Q_T(t)$, $Q_T^{(N)}(t)$, α and m as above. Let $K := (\frac{1}{2D})^{\frac{1}{1-2D}}$ and $m < K\alpha$. Then there exists an interval I , which contains m, and the sequence $\{\mu_N\}_{N \in \mathbb{N}}$ of absolute maximum points of $Q_T^{(N)}(t)$ in I such that the following limit holds in probability*

$$\mu_N \xrightarrow{N \to \infty} m$$

In order to prove this theorem we need to apply lemma 5.16. The most complicated part is to define a suitable interval I such that I contains m and excludes γ - the minimum point of Q_T between m and $m + \alpha$. Obviously, given such an interval we are sure that $Q_T|_I$ is increasing before m and decreasing after m because of lemma 5.13. Laboriousness comes up with the fact that we are able to observe only $Q_T^{(N)}$ realization, which means that I has to be defined starting from it. To proceed with our plan we need to find the maximum point of $Q_T^{(N)}$ which corresponds (in some sense) to m: therefore we have to get rid of all the maximum points caused by the irregular realization of $Q_T^{(N)}$. The idea is to find a maximum higher than the others: the following definition moves on this direction.

Definition 5.15. Let $\varepsilon > 0$. We define $\mu_{N,\varepsilon}$ as the minimum $t = \frac{kT}{N}$ such that t is the absolute maximum point of $Q_T^{(N)}$ on the connected component of $Q_T^{(N)\leftarrow}(Q_T^{(N)}(t) - 2\varepsilon, +\infty)$ which contains t. We define $A_{N,\varepsilon}$ as the connected component such that $\mu_{N,\varepsilon}$ is maximum of $Q_T^{(N)}$ on $A_{N,\varepsilon}$.

Proof. Recall Theorem 5.17. For all $\delta > 0, \varepsilon > 0$ there exists $\bar{N}$ such that for all $N \geq \bar{N}$ the following holds

$$\mathbb{P}\left[\left\{d_\infty\left(Q_T(\cdot), Q_T^{(N)}(\cdot)\right) > \varepsilon\right\}\right] < \delta \tag{5.4.3}$$

Let $C := \{\omega \in \Omega : d_\infty(Q_T(\cdot,\omega), Q_T^{(N)}(\cdot,\omega)) > \varepsilon\}$. We will consider only $\omega \in \Omega \backslash C$ e $N \geq \bar{N}$.

Let γ be the minimum point of Q_T on the interval $(T - \tau_{i(T)}, T - \tau_{i(T)-1})$. Lemma 5.13 shows that for all $\varepsilon > 0$, taking T close enough to $\tau_{i(T)}$ (i.e. taking m small enough) the following inequality holds [1]:

$$Q_T(m) - Q_T(\gamma) > 4\varepsilon \tag{5.4.4}$$

Let $\mu_{N,\varepsilon}$ and $A_{N.\varepsilon}$ defined in definition 5.15. (5.4.4), (5.4.3) imply $\mu_{N,\varepsilon} < \gamma$, thus $\mu_{N,\varepsilon} \in (0,\gamma)$. Moreover from Lemma 5.13 and from the hypothesis $T - \tau_{i(T)} < K(\tau_{i(T)} - \tau_{i(T)-1})$ follows that Q_T attains the absolute maximum on the interval $(0,\gamma)$ in m. Thus

$$Q_T^{(N)}(\mu_{N,\varepsilon}) - 2\varepsilon \leq Q_T(\mu_{N,\varepsilon}) - \varepsilon \leq Q_T(m) - \varepsilon \leq Q_T^{(N)}(m)$$

or equivalently $m \in A_{N,\varepsilon}$.

(5.4.4), definition 5.15 and $m \in A_{N,\varepsilon}$ implies

$$Q_T^{(N)}(\gamma) \leq Q_T(\gamma) + \varepsilon < Q_T(m) - 3\varepsilon \leq Q_T^{(N)}(m) - 2\varepsilon \leq Q_T^{(N)}(\mu_{N,\varepsilon}) - 2\varepsilon$$

or equivalently $\gamma \notin A_{N,\varepsilon}$.

Let $\varepsilon > 0$ fixed. Consider the interval

$$I := \bigcap_{N \geq \bar{N}} A_{N,\varepsilon}$$

taking T close enough to $\tau_{i(T)}$, $m \in I$, $\gamma \notin I$. From Lemma 5.13 follows that $Q_T|_I$ is increasing for $x < m$ and decreasing for $x > m$. Moreover I is closed since it is intersection of closed intervals.

From Lemma 5.16 follows that for $\omega \in \Omega \backslash C$

$$\mu_N(\omega) \xrightarrow{N\to\infty} m(\omega)$$

Thus arbitrary choice of $\delta > 0$ in (5.4.3) implies the thesis.

□

[1] Using the same argument of the proof of Lemma 5.13 we see that there is only a minimum point between two maximum points, thus $Q_T(m) - Q_T(\gamma) \xrightarrow{m\to 0+} +\infty$

In the proof above we have used the following results (see for instance [66] for Theorem 5.17).

Lemma 5.16. *Let I a closed interval. Let $f : I \to \mathbb{R}$ a continuous function with the following properties:*

- *f attains in m its unique local maximum in I*
- *f is strictly increasing for $x < m$, strictly decreasing for $x > m$,*
- *$\{f_n\}_{n\in\mathbb{N}}$ is a sequence of continuous function in I uniformly convergent to f. Moreover, for all $n \in \mathbb{N}$, m_n is the absolute maximum point of f_n.*

Then

$$m_n \xrightarrow{N\to\infty} m$$

Theorem 5.17. *Let $M \in \mathcal{M}^{c,loc}$. The process $S_n^{(2)}(M)$ converges to $\langle M\rangle$ in probability, uniformly on compact intervals $[0,T]$.*

5.4.2 An algorithm for the detection of jumps in the volatility

We explain now the empirical usage of the algorithm, justified by previous computations, in particular by theorem 5.14. We start introducing some notation. The financial index time series will be denoted by $(s_i)_{0\le i\le N}$, whereas the detrended logarithmic time series will be indicated by $(x_i)_{250\le i\le N}$, where

$$x_i := \log(s_i) - \bar{d}(i)$$

and $\bar{d}(i) := \frac{1}{250}\sum_{k=i-250}^{i-1}\log(s_i)$; we observe that it is not possible to define x_i for $i < 250$. We moreover define $(y(i))_{0\le i\le N}$ to be the corresponding series of the trading dates. We also introduce the empirical estimate of V_N as

$$\widehat{V_N}(k) := \frac{1}{k}\sum_{i=1}^{k}(x_{N-i+1} - x_{N-i})^2$$

Now, suppose that we want to know when the last shock time in the time series occurred. We recall that the idea is to choose an appropriate integer M such that $0 < M \le N$ and see where the sequence $(\widehat{V_N}(k))_{N-M\le k\le N}$ attains its maximum. This leads us to introduce the following definition.

Definition 5.18. Let $(s_i), (x_i), (y_i), N, M$ be as above; given as integer $\tilde{N}$ such that $M \le \tilde{N} \le N$, we define

$$\widehat{k}(\tilde{N}, M) := \text{argmax}_{\tilde{N}-M\le k\le\tilde{N}} \quad \frac{1}{k}\sum_{i=1}^{k}\left(x_{\tilde{N}-i+1} - x_{\tilde{N}-i}\right)^2.$$

This quantity is an estimate of the distance of the last shock time before $y_{\tilde{N}}$ from $\tilde{N}$. Using this we define also

$$\widehat{i}(\tilde{N}, M) := \tilde{N} - \widehat{k}(\tilde{N}, M) + 1,$$

our estimate of the index of the last shock time estimate, and consequently our estimate of the last shock time before $y_{\tilde{N}}$ is

$$\widehat{\tau}(\tilde{N}, M) := y\left(\widehat{i}(\tilde{N}, M)\right).$$

It is worth compare briefly our algorithm with the so called ICSS-GARCH algorithm. Following [87], we can describe the ICSS-GARCH algorithm as follows. Given a series of financial returns $r_1, \dots, r_n$, with mean 0 we define the cumulative sum of squares $C_k = \sum_{i=1}^{k} r_i^2$ and let

$$D_k = \frac{C_k}{C_n} - \frac{k}{n}, \qquad 1 \leq k \leq n, \qquad D_0 = D_n = 0$$

The idea is that if the sequence $r_1, \dots, r_n$ has constant variance, then the sequence $D_1, \dots, D_n$ should oscillate around 0. However, if there is a shock in the variance, the sequence should exhibit extreme behavior around that point.

We remark that both algorithms use squared returns to detect volatility shocks. However, the ICSS-GARCH algorithm works well under the assumption that the returns are normally distributed, but not with heavy-tailed distributions, as proved in [87]. Our algorithm, on the contrary, does not need any particular assumption on the distribution of the returns, but it is simply based on geometrical considerations. In fact it exploits the particular characteristics of a piecewise-concave Brownian motion time change to locate shocks. We point out the assumption of a piecewise-concave Brownian motion time change is very natural in the context of stochastic volatility models. In fact, to reproduce jumps in the volatility, one has to introduce a process that makes the volatility dramatically increase when a shock occurs, and then slowly decay over time. To reproduce such a behavior, it seems natural to introduce a piecewise-concave time change. Furthermore, we remark that this algorithm does not work just with the model that presented here. For instance it is possible to prove that it works with any model where the detrended log-price is given by W_{J_t}, where W is a Brownian motion and J_t is a time change such that

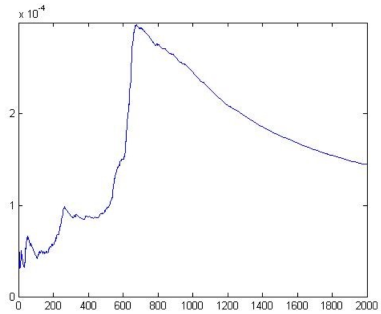
$$J_t = g(t - \tau_{i(t)}) + \sum_{k=1}^{i(t)} g(\tau_k - \tau_{k-1}),$$

with $\{\tau_i\}_{i\in\mathbb{Z}}$ and $i(t)$ as in the model in [1] and $g : [0, +\infty) \to [0, +\infty)$ is concave and satisfies $g(0) = 0$, $\lim_{h\to 0^+} g(h)/h = +\infty$. [2]

[2] In the model presented above, $g(h) = h^{2D}$

Recall that for the empirical discussion outlined here, we decided to use the DJIA Index and FTSE Index, from April 2nd, 1984 to July 6th, 2013, so that $N = 7368$. A similar data analysis has been done on the Standard & Poor's 500 Index, from January 3rd, 1950 to July 23th, 2013, finding analogous result and confirming the validity of the method we present here on aggregate indices. All the calculations and pictures presented here have been obtained using the software MatLab [76]. An example of the empirical procedure to estimate the last shock time is given in Figure 5.6.

Figure 5.6: Plot of the quantity $\widehat{V_{\tilde{N}}}(k)$ for $k = 1, \ldots, 2000$ ($M = 2000$). $\tilde{N}$ has been choosen so that $y_{\tilde{N}}$ is the 10th of May 2011. The peak corresponds to the the 15th of September 2008, the day of the Lehman Brothers bank bankruptcy.



However, we are not sure whether the choice of M that we made is good or not. Therefore, to confirm that the shock time estimate is good, we may repeat the estimate approaching the shock time, for example dropping the last observation, or dropping a particular number of the last observations. Then we can repeat this procedure many times and if we see that the last shock time estimate is confirmed, then we have a clear indication of the presence of a shock there (see Figure 5.7–(a)). We remark that when more than one shock is present on the time interval we consider, the most recent is always found as the maximum peak of $\hat{V}$ if we take $y_{\tilde{N}}$ close enough to it. When we get further, the chosen peak is not necessarily the most recent, as we can see in Figure 5.7–(b). This is exactly what we expect from Theorem 5.14.

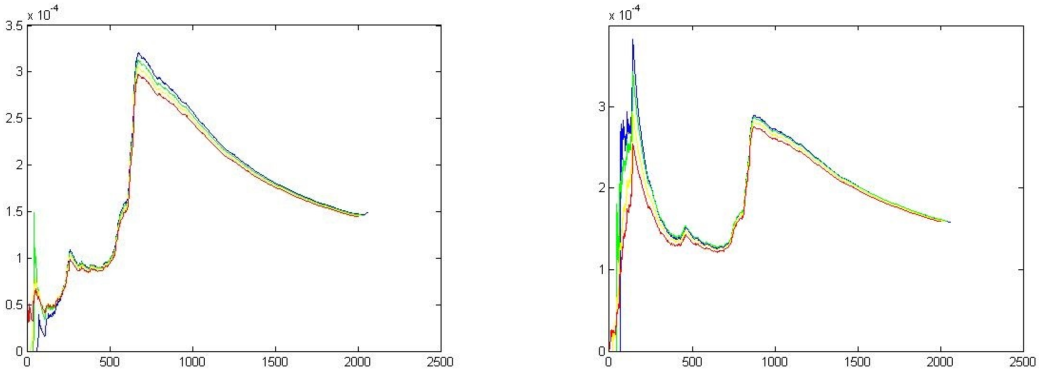
We can now apply the algorithm to try to locate all the past shocks in a given financial index time series. To do so, we simply calculate the quantity $\widehat{\imath}(\tilde{N}, M)$ for $\tilde{N} = N, \ldots, M$. We introduce the following sequence.

Definition 5.19. Given the quantities defined in definition 5.18, we introduce the past shock time sequence as

$$\widehat{h}((x_i)_{250 \leq i \leq N}, M) := \left(\widehat{\imath}(\tilde{N}, M)\right)_{M \leq \tilde{N} \leq N}$$

However, we slightly tweak the procedure in order to get a clearer result. When calculating $\widehat{k}(\tilde{N}, M)$ we ignore the last 20 elements of the sum, in order words, instead

Figure 5.7: Plot of the quantities $\widehat{V_{\tilde{N}}}(k)$ for $k = 1, \ldots, 2000$ ($M = 2000$), for the DJIA. In each figure we shift $\tilde{N}$ 4 times of 20 working days. In (a) $\tilde{N}$ has been chosen so that $y_{\tilde{N}}$ is the 10/05/11(red), the 11/04/11(yellow), the 14/03/11(green) and the 11/02/11(blue). The four maxima are all located the 15/09/08, the day of the Lehman Brothers bank bankruptcy, confirming the presence of a shock there. In (b) $\tilde{N}$ has been chosen so that $y_{\tilde{N}}$ is the 27/02/12(red), the 27/01/12(yellow), the 28/12/11(green) and the 29/11/11(blue). We can see that when $y_{\tilde{N}}$ is close to the 05/08/11 (European sovereign debt crisis), this date corresponds to the maximum of $\hat{V}$, whereas when we move further the maximum is again on the 15/09/08.

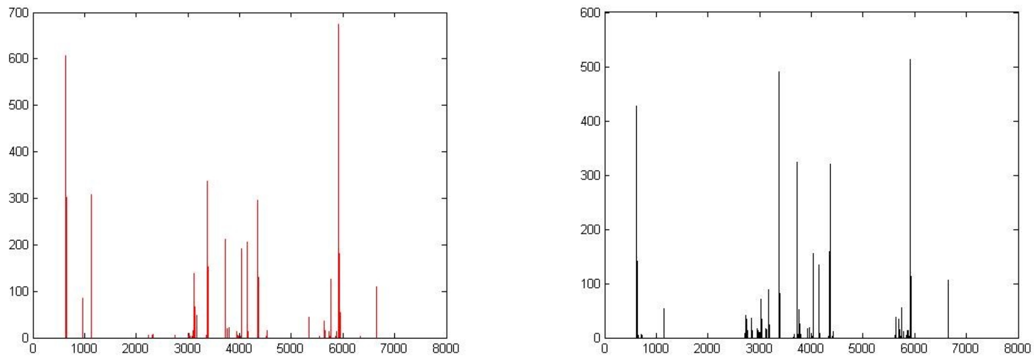


(a) (b)

of calculating $\widehat{k}(\tilde{N}, M)$ as the argmax for $\tilde{N}-M \leq k \leq \tilde{N}$, we drop the last 20 elements of the series. This leads us to recognize shocks that are at least 20 days old, removing the noise due to the excess volatility. We do this because when $\tilde{N}$ is very close to a shock, the procedure becomes unstable since near a shock the volatility is very high, so it is not always clear where the maximum is.

Finally, to get a clear picture of which are the big shocks in the time series, we can plot the number of the occurrences of each element of the sequence $\widehat{h}((x_i)_{250 \leq i \leq N}, M)$. We may choose to consider an element of the sequence of dates a shock if its numbers of occurrences exceeds a certain threshold. Table 5.1 contains our estimated shock-dates. In Figure 5.8 you can see the graphical evidence that maxima are concentrated on a small set of days for both FTSE and DJIA, supporting the validity of the method. The choice of the threshold is not completely determined, and we have based it on two criteria. Firstly, the number of estimated shocks should be consistent with the number of expected jumps of the Poisson process (whose rate will be calibrated in section 5.3). Secondly, we see that in both series it is possible to find a big interval in $\mathbb{N}$ for which almost no date has a number of occurrences contained in that interval. More explicitly, for the DJIA there are just 3 dates found approximately 50 times, whereas all the others are found more than 80 times or less than 25. Analogously, for the FTSE there are just 2 dates found approximately 50 times, whereas all the others are found more than 80 times or less than 20. It is therefore reasonable to consider true shocks the ones occurred more than 80 times, whereas it is not that clear how to consider the dates with approximately 50 occurrences. In any case, these choices are consistent with the number of expected jumps of the Poisson process. Another issue in the choice of shock dates is the fact that sometimes there are two or more very close dates which are found a considerable number of times. In this case we consider them as related to the same shock. These dates are marked with the word "sparse" in table 5.1, where we have reported our estimated dates.

Figure 5.8: Shock times; x-axis: increasing time index; y-axis: y(i)=number of times the maximum of $\hat{V}$ is realized at i



(a) FTSE shock times

(b) DJIA shock times

It is natural at this point to wonder if there is a relation between the shocks in the two indices, and a straightforward experiment is to try to superimpose the two graphics (see Figure 5.9). What we get is a clear indication that the shock times of the two series are almost coincident, only the magnitude (or evidence) being different and having very few shocks which are present just in one of the two indices. This is a validation of our choice of modeling the joint process of jumps $\mathcal{T} = (\mathcal{T}^X, \mathcal{T}^Y)$ as explained in section 5.3, taking $\mathcal{T}^X = \mathcal{T}^1 \cup \mathcal{T}^2$ and $\mathcal{T}^Y = \mathcal{T}^1 \cup \mathcal{T}^3$.

5.4.3 Application to cross asset correlation

As a consequence of previous section, a first idea to try a rough modeling of cross asset correlations is to suppose $\mathcal{T}^f$, jump process for FTSE, and $\mathcal{T}^d$, jump process for DJIA, to be the same process. But if this is true from Remark 5.10, and from the fact that D and $\bar{\sigma}$ are very similar for FTSE and DJIA, we would expect the decay of volatility autocorrelation in the DJIA, the decay of volatility autocorrelation in the FTSE and the decay of cross-asset correlation of absolute returns to display a similar behavior. In fact, this is exactly what happens if we plot these quantities (see Figure 5.10), in agreement with the empirical findings of [83].

Under this rough hypothesis our estimate for cross-asset correlations is therefore our prediction for the decay of volatility autocorrelation in FTSE or DJIA, or a mean between the two.

Table 5.1: Estimated dates of shock times

FTSE	*DJIA*
14/10/87	15/09/87
24/01/89	
26/09/89	11/10/89 (questionable, 53)
	09/01/96
	01/07/96 (questionable, 54)
	13/03/97 (sparse)
08/08/97	
22/10/97 (questionable, 48)	16/10/97
04/08/98	31/07/98
30/12/99	04/01/00
09/03/01	09/03/01
06/09/01	06/09/01
12/06/02	05/07/02
12/05/07 (questionable, 43)	
24/07/07	24/07/07 (questionable, 57)
15/01/08	04/01/08 (sparse)
03/09/08	15/09/08
05/08/11	05/08/11

Figure 5.9: Common jumps: overlap of Figures 5.8 (a) and (b)

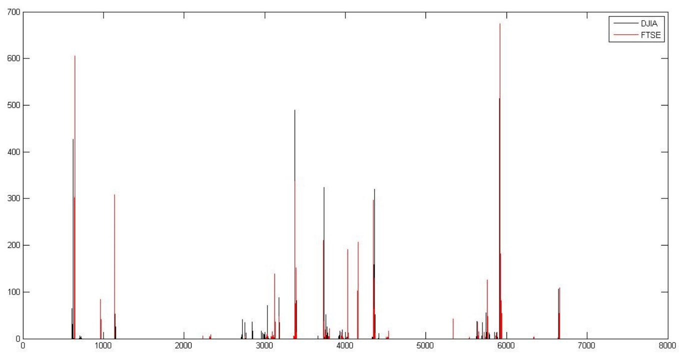
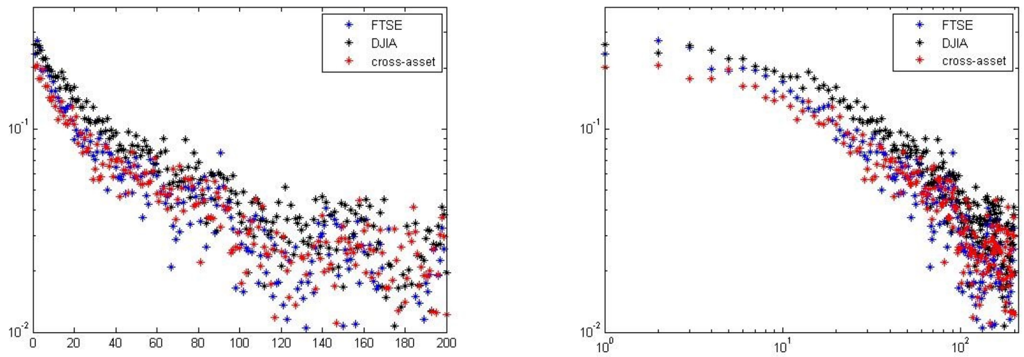


Figure 5.10: Comparison of empirical correlations



(a) log plot; one point out of three is plotted

(b) loglog plot; for $t \geq 20$, one point out of three is plotted

We can do better using the bivariate jump process $I = (I^X, I^Y)$ described at the beginning of section 5.3. We just have to estimate the intensities $\lambda_1, \lambda_2, \lambda_3$, subject to the constraints coming from the estimates of the one dimensional models. The set of feasible λs is in fact a segment in $\mathbb{R}^3$.

Define $\widehat{\gamma}_h(t)$ as the empirical correlation coefficient over h days:

$$\widehat{\gamma}_h(t) = corr(|x^f_{\cdot+h} - x^f_{\cdot}|, |x^d_{\cdot+t+h} - x^d_{\cdot+t}|).$$

where x^f and x^d are the FTSE and DJIA series of detrended log returns.

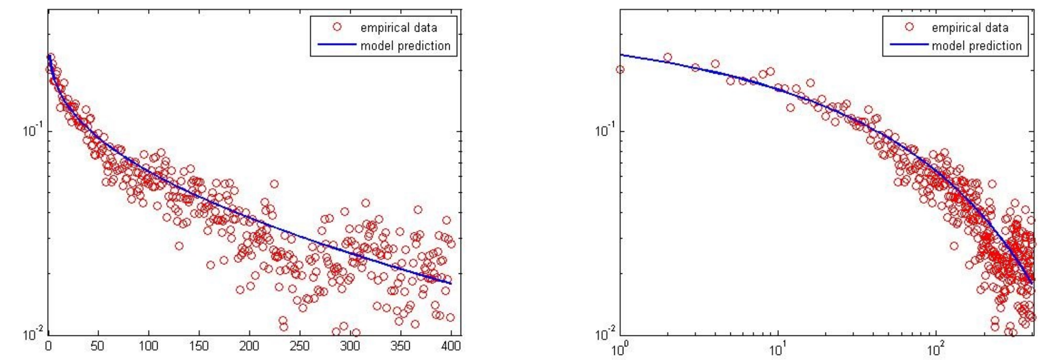
Minimizing a suitable L^2 distance between this quantity and the theoretical cross-correlation (Theorem 5.5) we obtain

$$\lambda_1 = 0.0014; \quad \lambda_2 = 0.0005; \quad \lambda_3 = 0.$$

In Figure 5.11 we can see the excellent agreement of the prediction of our model and the empirical decay of the cross-asset correlations, for $t = 1, .., 400$ days.

The fact that our estimate is $\lambda_3 = 0$ means that our best fitting with real data is obtained when the shocks for FTSE are given by the shocks of the DJIA plus some additional ones, given by a sparser and independent Poisson process. These estimates, due to the small sample size, are too rough to allow more quantitative considerations. In any case, if we want to see a reason for the situation above, we can suppose that shocks in the DJIA index always determine a shock in the FTSE index, whereas it is possible to see a shock in the FTSE which does not imply a significant increment in the empirical variance of DJIA.

Figure 5.11: FTSE and DJIA cross-asset correlations



(a) log plot (b) loglog plot

Chapter 6

Multi-scaling of moments in stochastic volatility models

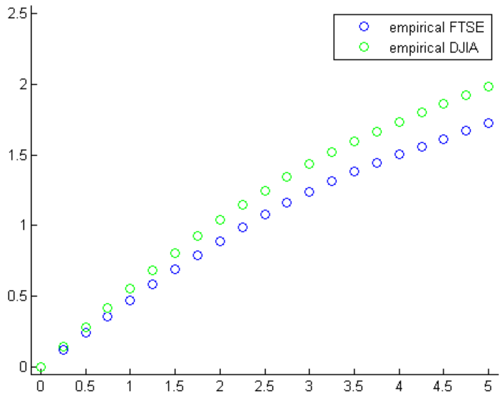
6.1 Introduction

Let $(X_t)_{t\geq 0}$ be a continuous-time martingale representing the log-price of an asset. We say that the *multi-scaling of moments* occurs if the limit

$$\limsup_{h\downarrow 0} \frac{\log \mathbb{E}\left(|X_{t+h} - X_t|^q\right)}{\log h} =: A(q) \tag{6.1.1}$$

is *non-linear* on the set $\{q \geq 1 : |A(q)| < +\infty\}$. More intuitively, (6.1.1) says that $\mathbb{E}\left(|X_{t+h} - X_t|^q\right)$ scales, in the limit as $h \downarrow 0$, as $h^{A(q)}$, with $A(q)$ non-linear. In the case X_t is a Brownian martingale (i.e. a stochastic integral w.r.t. a Brownian motion), one would expect $A(q) = \frac{q}{2}$, at least for q sufficiently small. In this case, multi-scaling of moments can be identified with deviations from this *diffusive* scaling, occurring for q above a given threshold; this type of multi-scaling is indeed widely observed in financial data ([95, 56, 54, 44, 43]), as we can see in fig. 6.1. *Multifractal models* are

Figure 6.1: Scaling exponent A(q) of FTSE and DJIA



a large class of stochastic processes that exhibit multi-scaling for a rather arbitrary scaling function $A(q)$ ([28, 27, 26]). In these models X_t is defined as a *time changed* Brownian motion:

$$X_t := W_{I(t)}, \tag{6.1.2}$$

but it cannot be expressed as a *stochastic volatility model*, i.e. in the form $dX_t = \sigma_t dB_t$, for a Brownian motion B_t. Some results on the scaling properties of stochastic volatility models are contained in [55], where the scaling of the log-volatility is investigated.

In chapter 5 a simple stochastic volatility model exhibiting a bi-scaling behavior has been constructed: (6.1.1) holds with a function $A(q)$ which is piecewise linear and the slope $A'(q)$ takes two different values. In this chapter, following [41], we analyze multi-scaling in a more general class of stochastic volatility models, namely those of the form $dX_t = \sigma_t dB_t$, with a volatility process σ_t independent of the Brownian motion B_t; these processes are exactly those that can be written in the form (6.1.2) with a trading time $I(t)$ independent of W_t, and with *absolutely continuous* trajectories. Remark that considering models with no drift does not affect our analysis, since in short time the drift has a smaller scale than the Brownian Motion, and therefore its contribution to $A(q)$ is negligible. We focus in particular on models in which $V_t := \sigma_t^2$ is a *stationary* solution of a stochastic differential equation of the form

$$dV_t \;=\; -f(V_t)dt + dL_t, \tag{6.1.3}$$

for a *Lévy subordinator* L_t whose characteristic measure has *power law tails at infinity*, and a function $f(\cdot)$ such that a stationary solution exists, and it is unique in law. We show that multi-scaling is *not* possible if $f(\cdot)$ has *linear growth*, but if $f(\cdot)$ behaves as Cx^γ as $x \to +\infty$, with $C > 0$ and $\gamma > 1$, then the stochastic volatility process whose volatility is a stationary solution of (6.1.3), exhibits multi-scaling. In this class of models multi-scaling comes from the combination of heavy tails of L_t and superlinear mean reversion; technically speaking, as will be seen later, the key point is that the distribution of V_t has *lighter* tails than those of L_t. We remark that the class of processes introduced in [1] can be seen as limiting cases of those considered here, with $\gamma > 2$ and the characteristic measure of the Lévy process L_t concentrated on $+\infty$. For all of these models the scaling function $A(q)$ is piecewise linear, with two values for the slope. We discuss in Section 6.4 how this is compatible with empirical data coming, for instance, from DJIA or FTSE indices or from currency exchange rate data. We also briefly discuss how the fact that the processes considered in this chapter are also of the form (6.1.2) can be useful in financial applications such as option pricing.

The chapter is organized as follows. In Section 6.2 we give some basic facts on stochastic volatility models, and provide some necessary conditions for multi-scaling. Section 6.3 contains more specific results for models whose volatility is given by (6.1.3). Section 6.4 deals with possible financial applications, and in section 6.5 we prove that for these models volatility autocorrelation decays exponentially, finding some estimates for the rate of decay.

6.2 Multi-scaling in stochastic volatility models

We consider a stochastic process $(X_t)_{t\geq 0}$ that can be expressed in the form

$$dX_t = \sigma_t dW_t, \tag{6.2.1}$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion, and $(\sigma_t)_{t\geq 0}$ is a *stationary*, $[0, +\infty)$-valued process, *independent* of $(X_t)_{t\geq 0}$, that we will call the *volatility process.* We assume the following weak continuity assumption on the volatility process.

Assumption A. As $h \downarrow 0$, the limit

$$\frac{1}{h}\int_0^h (\sigma_s - \sigma_0)^2 ds \;\to\; 0$$

holds in probability.

We begin with a basic result on the scaling function $A(q)$ defined in (6.1.1). It states that under a uniform integrability condition on the *integrated squared volatility*, the diffusive scaling holds. Thus a necessary condition for multi-scaling is the loss of this uniform integrability.

Proposition 6.1. *Assume that, $p > 1$,*

$$\limsup_{h\downarrow 0} \mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{p/2}\right] < +\infty. \tag{6.2.2}$$

Then, under Assumption A, $A(q) = \frac{q}{2}$ for every $q < p$.

Proof. Note that

$$\frac{X_h - X_0}{\sqrt{h}} = \frac{1}{\sqrt{h}}\int_0^h \sigma_s dW_s = \int_0^1 \sigma_{uh} dB_u^h,$$

where $B_u^h := \frac{1}{\sqrt{h}} W_{hu}$ is also a standard Brownian motion. Thus, $\frac{X_h - X_0}{\sqrt{h}}$ has the same law of $\int_0^1 \sigma_{uh} dB_u$, where B is *any* Brownian motion independent of the volatility process $(\sigma_t)_{t\geq 0}$. It follows from Assumption A and the isometry property of the stochastic integral, that

$$\int_0^1 \sigma_{uh} dB_u \;\to\; \sigma_0 B_1 \tag{6.2.3}$$

in L^2 and therefore in probability, as $h \downarrow 0$. By (6.2.2) and the Burkholder-Davis-Gundy inequality (see [86]),

$$\mathbb{E}\left[\left|\int_0^1 \sigma_{uh} dB_u\right|^p\right] \leq C_p \mathbb{E}\left[\left(\int_0^1 \sigma_{uh}^2 du\right)^{p/2}\right] = \mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{p/2}\right],$$

so the family of random variables $\left\{\int_0^1 \sigma_{uh} dB_u : h > 0\right\}$ is bounded in L^p. This implies that the convergence in (6.2.3) is also in L^q, for every $q < p$. Thus

$$\mathbb{E}\left[\left|\frac{X_h - X_0}{\sqrt{h}}\right|^q\right] = \mathbb{E}\left[\left|\int_0^1 \sigma_{uh} dB_u\right|^q\right] \to \mathbb{E}\left(\sigma_0^q\right)\mathbb{E}\left[|B_1|^q\right]$$

as $h \downarrow 0$ (in particular $\mathbb{E}\left(\sigma_0^q\right) < +\infty$). Taking the logarithms in the limit above, one obtains $A(q) = \frac{q}{2}$.

□

Remark 6.2. Suppose $1 \leq q < p$. Then $\frac{A(p)}{p} \leq \frac{A(q)}{q}$. This follows immediately from the fact that, for every $h > 0$,

$$\frac{\log \mathbb{E}\left(|X_{t+h} - X_t|^q\right)}{q} = \log \|X_{t+h} - X_t\|_q$$

is increasing in q, so that $\frac{\log \mathbb{E}(|X_{t+h} - X_t|^q)}{q \log h}$ is decreasing in q for all $0 < h < 1$.

In what follows, for models of the form (6.2.1), we assume the following further conditions.

Assumption B. $\mathbb{E}\left(\sigma_0^2\right) < +\infty$.

Under Assumption B, (6.2.2) holds true for $p = 2$. By Proposition 6.1 and Remark 6.2, we have that $A(q) = \frac{q}{2}$ for $1 \leq q < 2$, while $\frac{q}{2} \geq A(q) \geq -\infty$ for $q \geq 2$. This suggests the following formal definition of multi-scaling.

Definition 6.3. Under Assumptions A and B, we say that multi-scaling occurs if $\{q : -\infty < A(q) < \frac{q}{2}\}$ has a nonempty interior.

In what follows, Assumptions A and B will be assumed implicitly. Note now that, by the Burkholder-Davis-Gundy inequality, there are constants c_p, C_p such that for each $h > 0$

$$c_p \mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{p/2}\right] \leq \mathbb{E}\left[|X_h - X_0|^p\right] = \mathbb{E}\left[\left|\int_0^h \sigma_s dW_s\right|^p\right] \leq C_p \mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{p/2}\right]. \tag{6.2.4}$$

Thus, the condition

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{q/2}\right] < +\infty$$

for each $h > 0$ is necessary for $A(q) > -\infty$. Note also that, by Jensen's inequality,

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{q/2}\right] \leq \frac{1}{h}\int_0^h \mathbb{E}\left[\sigma_s^q\right] ds = \mathbb{E}\left[\sigma_0^q\right], \tag{6.2.5}$$

for $q \geq 2$. Thus, whenever $\mathbb{E}\left[\sigma_0^q\right] < +\infty$, the assumption of Proposition 6.1 holds.

These remarks, together with Proposition 6.1, yields the following statement.

Corollary 6.4. *A necessary condition for multi-scaling in* (6.2.1) *is that there exists* $p > 2$ *such that*

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{p/2}\right] < +\infty$$

for each $h > 0$, *but*

$$\mathbb{E}\left[\sigma_0^p\right] = +\infty.$$

From the result above we derive an alternative necessary condition for multi-scaling, which has sometimes the advantage to be more easily checked in specific models.

Corollary 6.5. *A necessary condition for multi-scaling in* (6.2.1) *is that, for some* $h > 0$, *there exists* $p \geq 2$ *such that*

$$\mathbb{E}\left[\sigma_0^p\right] < +\infty \qquad \mathbb{E}\left[\sup_{0 \leq t \leq h} \sigma_t^p\right] = +\infty.$$

Proof. Assume multi-scaling holds, and define

$$q^* := \inf\{q : \mathbb{E}\left[\sigma_0^q\right] = +\infty\}.$$

By Corollary 6.4, $q^* < +\infty$ while, by Assumption B, $q^* \geq 2$. Moreover, by Proposition 6.1, $A(q) = q/2$ for $q < q^*$. Thus, by Definition 6.3, $A(q)$ has to be finite for some $q > q^*$; in particular, as observed above,

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{q/2}\right] < +\infty$$

for $h > 0$. Consider l, r with $q^* < l < r < q$. Setting $M_h := \sup_{0 \leq t \leq h} \sigma_t$, we have

$$\frac{1}{h}\int_0^h \sigma_s^l ds \leq M_h^{l-2} \frac{1}{h}\int_0^h \sigma_s^2 ds.$$

By stationarity of σ_t, and by applying Hölder inequality with conjugate exponents $\frac{r}{2}$ and $\frac{r/2}{r/2-1}$, we obtain

$$\mathbb{E}\left(\sigma_0^l\right) \leq \left[\mathbb{E}\left(M_h^{r\frac{l/2-1}{r/2-1}}\right)\right]^{1-\frac{2}{r}} \left[\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{r/2}\right]\right]^{2/r}.$$

Since $l > q^*$, it follows that $\mathbb{E}\left(\sigma_0^l\right) = +\infty$. Moreover, as $r < q$,

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \sigma_s^2 ds\right)^{r/2}\right] < +\infty.$$

Thus, necessarily,

$$\mathbb{E}\left(M_h^{r\frac{l/2-1}{r/2-1}}\right) = +\infty.$$

It is easily checked that, choosing l and q^* sufficiently close, one gets

$$\tilde{r} := r\frac{l/2-1}{r/2-1} < q^*,$$

which implies

$$\mathbb{E}\left(\sigma_0^{\tilde{r}}\right) < +\infty.$$

Setting $p := \max(\tilde{r}, 2)$, the proof is completed.

□

We conclude this section by showing a further property of the scaling function $A(q)$

Remark 6.6. Assume that, for each $h > 0$, the integrated volatility has moments of all orders, i.e.

$$\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^q\right] < +\infty \quad \text{for every } q \geq 1. \tag{6.2.6}$$

The following argument shows that, under this assumption, $A(q)$ is increasing in q. We will see later an example in which the integrated volatility has heavy tails, so it violates (6.2.6), and $A(\cdot)$ is decreasing in an interval. We begin by observing that, by (6.2.4),

$$A(q) = \limsup_{h\downarrow 0} \frac{\log \mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{\log h}. \tag{6.2.7}$$

From this it easily follows that

$$\liminf_{h\downarrow 0} \frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{h^\lambda} = 0 \implies \lambda \leq A(q), \tag{6.2.8}$$

and

$$\lambda < A(q) \implies \liminf_{h\downarrow 0} \frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{h^\lambda} = 0. \tag{6.2.9}$$

Consider $p > q \geq 1$. Moreover, let $\epsilon > 0$, and take $l < q$ such that $[A(q) - \epsilon]\frac{q}{l} < A(q)$. Set

$$a_h := \left(\int_0^h \sigma_t^2 dt\right)^{1/2}.$$

We now use Young's inequality $\alpha\beta \le \frac{\alpha^r}{r} + \frac{\beta^{r'}}{r'}$, valid for $\alpha, \beta \ge 0$, $r, r' > 0$, $\frac{1}{r} + \frac{1}{r'} = 1$. Choosing $\alpha = \frac{a_h^l}{h^{A(q)-\epsilon}}$, $\beta = a_h^{p-l}$, $r = \frac{q}{l}$, we get

$$\frac{a_h^p}{h^{A(q)-\epsilon}} \le \frac{l}{q}\frac{a_h^q}{h^{(A(q)-\epsilon)\frac{q}{l}}} + \frac{q-l}{q} a_h^{q\frac{p-l}{q-l}}.$$

Taking expectations:

$$\frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{p/2}\right]}{h^{A(q)-\epsilon}} \le \frac{l}{q}\frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{h^{(A(q)-\epsilon)\frac{q}{l}}} + \frac{q-l}{q}\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q\frac{p-l}{2(q-l)}}\right]. \tag{6.2.10}$$

Since $[A(q)-\epsilon]\frac{q}{l} < A(q)$, by (6.2.9)

$$\liminf_{h\downarrow 0} \frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{h^{(A(q)-\epsilon)\frac{q}{l}}} = 0. \tag{6.2.11}$$

Moreover,

$$\lim_{h\downarrow 0} \mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q\frac{p-l}{2(q-l)}}\right] = 0 \tag{6.2.12}$$

by (6.2.6) and dominated convergence. It follows from (6.2.10), (6.2.11) and (6.2.12), that

$$\liminf_{h\downarrow 0} \frac{\mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{p/2}\right]}{h^{A(q)-\epsilon}} = 0$$

which, together with (6.2.8), yields $A(p) \ge A(q) - \epsilon$. Since ϵ is arbitrary, the conclusion follows.

6.3 Superlinear Ornstein-Uhlenbeck volatility

In this section we devote our attention to a specific class of stochastic volatility models, namely those of the form

$$\begin{aligned} dX_t &= \sigma_t dB_t \\ dV_t &= -f(V_t)dt + dL_t \\ V_t &= \sigma_t^2, \end{aligned} \tag{6.3.1}$$

where:

- $(B_t)_{t\ge 0}$ is a standard Brownian motion.

- $(L_t)_{t\geq 0}$ is a Lévy process with increasing paths (*subordinator*) independent of $(B_t)_{t\geq 0}$. More precisely $(L_t)_{t\geq 0}$ is a real-valued process, with independent increments, $L_0 = 0$ and
$$\mathbb{E}\left[\exp(-\lambda L_t)\right] = \exp[-t\Psi(\lambda)],$$
with
$$\Psi(\lambda) = m\lambda + \int_{(0,+\infty)} \left(1 - e^{-\lambda x}\right) \nu(dx),$$
where $m \geq 0$ is the *drift* of the process, and ν is a positive measure on $(0, +\infty)$, called *characteristic measure*, satisfying the condition
$$\int_{(0,+\infty)} (1 \wedge x)\nu(dx) < +\infty.$$
For generalities on Lévy Processes see [19, 38, 91].

- $f(\cdot)$ is a locally Lipschitz, nonnegative function such that $f(0) = 0$ (which guarantees $V_t \geq 0$ if $V_0 \geq 0$).

Some conditions on $f(\cdot)$ are needed for (6.3.1) to have a stationary solution. We will address this point later. We will always assume that V_0 is independent of $(L_t)_{t\geq 0}$. We note now that for many "natural" choices of f, multi-scaling is not allowed. In particular, multiscaling is not present in Ornstein-Uhlenbeck models (see e.g. [51, 68]).

Proposition 6.7. *Suppose $f(\cdot)$ satisfies the* linear growth condition
$$|f(v)| \leq Av + B$$
for some $A, B > 0$ and all $v > 0$. Moreover, assume (6.3.1) *has a solution for which $(V_t)_{t\geq 0}$ is stationary, nonnegative and integrable, such that Assumptions A and B hold. Then multi-scaling does not occur.*

Proof. By Remark 6.2, $A(q) \leq q/2$, so we need to show the converse inequality. Let $(V'_t)_{t\geq}$ be a solution of
$$\begin{aligned} dV'_t &= -(AV'_t + 2B)dt + dL_t \\ V'_0 &= V_0. \end{aligned} \tag{6.3.2}$$
Note that
$$d(V_t - V'_t) = -\left[f(V_t) - AV'_t - 2B\right]dt.$$
In particular $V_t - V'_t$ is continuously differentiable, and $V_0 - V'_0 = 0$. It follows that $V_t - V'_t \geq 0$ for every $t \geq 0$: indeed the path of $V_t - V'_t$ cannot downcross the value zero, since whenever $\bar{t}$ is such that $V_{\bar{t}} = V'_{\bar{t}} = v$, then
$$\frac{d}{dt}(V_{\bar{t}} - V'_{\bar{t}}) = -f(v) + Av + 2B \geq B > 0.$$

Thus for every $t \geq 0$

$$V_t \geq V_t' = V_0 e^{-At} + \frac{2B}{A}\left(e^{-At} - 1\right) + \int_0^t e^{-A(t-s)} dL_s \geq V_0 e^{-At} + e^{-tA/2} L_{t/2} - \frac{2B}{A}.$$

On the other hand

$$V_t = V_0 - \int_0^t f(V_s)ds + L_t \leq V_0 + L_t,$$

which yields

$$\sup_{t\in[0,h]} V_t \leq V_0 + L_h.$$

Putting all together

$$V_0 e^{-Ah} + e^{-Ah/2} L_{h/2} - \frac{2B}{A} \leq V_h \leq \sup_{t\in[0,h]} V_t \leq V_0 + L_h. \tag{6.3.3}$$

Since

$$V_0 e^{-Ah} + e^{-Ah/2} L_{h/2} - \frac{2B}{A} \in L^p \iff V_0 + L_h \in L^p,$$

the conclusion now follows from (6.3.3) and Corollary 6.5.

□

Proposition 6.7 shows that, for models of the form (6.3.1) to exhibit multi-scaling, one needs to consider a drift $f(\cdot)$ with a superlinear growth.

Definition 6.8. We say that a function $f : (0, +\infty) \to (0, +\infty)$ is *regularly varying at infinity* with exponent $\alpha \in \mathbb{R}$ if, for every $x > 0$,

$$\lim_{t\to+\infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

In the case $\alpha = 0$ we say that f is *slowly varying at infinity.* Note that f is regularly varying at infinity with exponent α if and only if $f(u) = u^\alpha l(u)$ where l is slowly varying at infinity. In what follows we consider models of the form (6.3.1) for which the following assumptions hold:

A1 $(B_t)_{t\geq 0}$ is a standard Brownian motion.

A2 $(L_t)_{t\geq 0}$ is a Lévy subordinator with characteristic measure ν. Moreover $(B_t)_{t\geq 0}$ and $(L_t)_{t\geq 0}$ are independent.

A3 The function $u \mapsto \nu((u, +\infty))$ is regularly varying at infinity with exponent $-\alpha < 0$.

A4 $f : [0, +\infty) \to [0, +\infty)$ is increasing, locally Lipschitz, $f(0) = 0$, and it is regularly varying at infinity with exponent $\gamma > 1$.

The following result has been proved in [89] (see also [69], [49] for related results).

Theorem 6.9. *Under assumption A2-A4, the equation* $dV_t = -f(V_t)dt + dL_t$ *admits an unique stationary distribution* μ. *Moreover* $\mu((u, +\infty))$ *is regularly varying at infinity with exponent* $-\alpha - \gamma + 1$.

In what follows we assume V_0 is independent of $(B_t)_{t \geq 0}$ and $(L_t)_{t \geq 0}$, and it has distribution μ. Theorem 6.9 shows that, if $\gamma > 1$, V_t has a distribution with lighter tails than those of the Lévy process L_t.

We are now ready to state the main result of this paper.

Theorem 6.10. *Assume A1-A4 are satisfied, and that* $\alpha + \gamma > 2$ *(which, in particular, implies Assumption B). Then the following statements hold.*

(1) *If* $\gamma \geq 2$ *then*

$$A(q) = \begin{cases} \frac{q}{2} & \text{for } 1 \leq q < 2(\alpha + \gamma - 1) \\ \frac{\gamma - 2}{2(\gamma - 1)} q + \frac{\alpha + \gamma - 1}{\gamma - 1} & \text{for } q > 2(\alpha + \gamma - 1). \end{cases}$$

(2) *If* $1 < \gamma < 2$ *then*

$$A(q) = \begin{cases} \frac{q}{2} & \text{for } 1 \leq q < 2(\alpha + \gamma - 1) \\ \frac{\gamma - 2}{2(\gamma - 1)} q + \frac{\alpha + \gamma - 1}{\gamma - 1} & \text{for } 2(\alpha + \gamma - 1) < q < \frac{2\alpha}{2 - \gamma} \\ -\infty & \text{for } q > \frac{2\alpha}{2 - \gamma}. \end{cases}$$

Moreover, for $q \neq 2(\alpha + \gamma - 1), \frac{2\alpha}{2-\gamma}$, *the scaling exponent* $A(q)$ *in* (6.1.1) *can be defined as a limit rather that a* lim sup.

We remark that, in the case $1 < \gamma < 2$, $A(\cdot)$ is decreasing for $2(\alpha + \gamma - 1) < q < \frac{2\alpha}{2-\gamma}$. This is not in contradiction with Remark 6.6, since assumption (6.2.6) is not satisfied.

Remark 6.11. A simple consequence of Theorem 6.10, is that, by a comparison argument, Proposition 6.7 can be extended to any f which is regularly varying at infinity with exponent 1.

The proof of Theorem 6.10 will be divided into several steps. We begin by dealing with the case $f(v) = Cv^{\gamma}$, with $C > 0$, and L_t is a *compound Poisson process*.

Proposition 6.12. *The conclusion of Theorem 6.10 hold if* $f(v) = Cv^{\gamma}$, *with* $C > 0$, L_t *is a Lévy subordinator with zero drift and* finite *characteristic measure* ν.

Proof. Note that, for $q < 2(\alpha + \gamma - 1)$, by Theorem 6.9, we have $\mathbb{E}\left[V_0^{q/2}\right] < +\infty$ so that, by Proposition 6.1 and (6.2.5), $A(q) = \frac{q}{2}$. Thus it is enough to consider the case $q > 2(\alpha + \gamma - 1)$. In what follows we also write $a_h \sim h^u$ for

$$\lim_{h \to 0} \frac{\log a_h}{\log h} = u. \tag{6.3.4}$$

We will repeatedly use the simple fact that (6.3.4) follows if we show that for every $\epsilon > 0$ there exist $C_\epsilon > 1$ such that

$$\frac{1}{C_\epsilon} h^{u+\epsilon} < a_h < C_\epsilon h^{u-\epsilon}.$$

In what follows all estimates on $A(q)$ are based on the fact (see (6.2.7)) that the limit

$$\lim_{h \downarrow 0} \frac{\log \mathbb{E}\left[\left(\int_0^h \sigma_t^2 dt\right)^{q/2}\right]}{\log h}$$

exists if and only if the limit

$$\lim_{h \downarrow 0} \frac{\log \mathbb{E}\left(|X_{t+h} - X_t|^q\right)}{\log h}$$

exists, and in this case they coincide.

Part 1: $\gamma > 2$

By the assumption of finiteness of ν, (L_t) jumps finitely many times in any compact interval. Denote by $(\tau_k)_{k \geq 1}$ the (ordered) set of positive jump times, and $\tau_0 = 0$. Given $h > 0$, we denote by $i(h)$ the random number of jump times in the interval $(0, h]$.

Case $i(h) = 0$. When $i(h) = 0$, V_t solves, for $t \in [0, h]$, the differential equation $\frac{d}{dt} V_t = -C V_t^\gamma$, whose solution is

$$V_t = \left(V_0^{1-\gamma} + (\gamma - 1)Ct\right)^{\frac{1}{1-\gamma}}.$$

Integrating, we get

$$\int_0^h V_t dt = \frac{\gamma - 2}{\gamma - 1}\left[(V_0^{1-\gamma} + (\gamma - 1)Ch)^{\frac{\gamma-2}{\gamma-1}} - (V_0^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right]. \tag{6.3.5}$$

Note that, setting $\lambda := \nu([0, +\infty))$,

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] = \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \Big| i(h) = 0\right] e^{-\lambda h}.$$

The factor $e^{-\lambda h}$ gives no contribution to the behavior of $\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right]$ as $h \to 0$, and it can be neglected. Moreover, by (6.3.5), and using the fact that V_0 and $\{i(h) = 0\}$ are independent,

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \Big| i(h) = 0\right] =$$

$$= \left(\frac{\gamma - 2}{\gamma - 1}\right)^{q/2} ((\gamma-1)Ch)^{\frac{\gamma-2}{2(\gamma-1)}q} \mathbb{E}\left[\left[\left(\frac{V_0^{1-\gamma}}{(\gamma - 1)Ch} + 1\right)^{\frac{\gamma-2}{\gamma-1}} - \left(\frac{V_0^{1-\gamma}}{(\gamma - 1)Ch}\right)^{\frac{\gamma-2}{\gamma-1}}\right]^{q/2}\right]. \tag{6.3.6}$$

Since, for $0 < a < 1$ and $z > 0$,

$$a(z+1)^{a-1} \leq (z+1)^a - z^a \leq (z+1)^{a-1}, \tag{6.3.7}$$

for computing the limit $\lim_h \frac{\log \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\right]}{\log h}$, the right hand side of (6.3.6) can be replaced by (using the previous inequality for $a = \frac{\gamma-2}{\gamma-1}$; recall that $\gamma > 2$)

$$h^{\frac{\gamma-2}{2(\gamma-1)}q}\mathbb{E}\left[\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{\frac{\gamma-2}{\gamma-1}-1}\right]^{q/2}\right] = h^{\frac{\gamma-2}{2(\gamma-1)}q}\mathbb{E}\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{-\frac{q}{2(\gamma-1)}}\right]. \tag{6.3.8}$$

In other words:

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \sim h^{\frac{\gamma-2}{2(\gamma-1)}q}\mathbb{E}\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{-\frac{q}{2(\gamma-1)}}\right]. \tag{6.3.9}$$

To estimate the r.h.s. of (6.3.9), we observe that for $y > 0$ and $0 < u < r$, the following inequalities can be easily checked

$$\frac{1}{2^r}\mathbf{1}_{\{y<1\}} \leq (1+y)^{-r} \leq (1+y)^{-u} \leq y^{-u}. \tag{6.3.10}$$

Setting $r := \frac{q}{2(\gamma-1)}$ and $Y := \frac{V_0^{1-\gamma}}{(\gamma-1)Ch}$, using (6.3.10) we obtain

$$\frac{1}{2^r}\mathbb{P}(Y<1) \leq \mathbb{E}\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{-\frac{q}{2(\gamma-1)}}\right] \leq E\left(Y^{-u}\right) \tag{6.3.11}$$

for every $u < \frac{q}{2(\gamma-1)}$. Set $\xi := \frac{\alpha+\gamma-1}{\gamma-1}$. Note that $\xi < r$ for $q > 2(\alpha+\gamma-1)$. By Theorem 6.9:

$$\mathbb{P}(Y<1) = \mathbb{P}\left(V_0 > \left(\frac{1}{(\gamma-1)Ch}\right)^{\frac{1}{\gamma-1}}\right) \sim \left[\left(\frac{1}{(\gamma-1)Ch}\right)^{\frac{1}{\gamma-1}}\right]^{\alpha+\gamma-1} \sim h^{\xi}, \tag{6.3.12}$$

Moreover, take $u < \xi$. We have

$$\mathbb{E}\left(Y^{-u}\right) = [(\gamma-1)Ch]^u\,\mathbb{E}\left[V_0^{u(\gamma-1)}\right] \leq Ah^u, \tag{6.3.13}$$

for some $A > 0$ that may depend on u but not on h, where we have used the fact that $\mathbb{E}\left[V_0^{u(\gamma-1)}\right] < +\infty$, since $u(\gamma-1) < \alpha+\gamma-1$. Since u can be taken arbitrarily close to ξ, by (6.3.11), (6.3.12) and (6.3.13) we obtain

$$\mathbb{E}\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{-\frac{q}{2(\gamma-1)}}\right] \sim h^{\xi}, \tag{6.3.14}$$

which yields

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}}. \tag{6.3.15}$$

Note that (6.3.15) has the right order, according to the statement of Theorem 6.10. Therefore, in order to complete the proof for $\gamma > 2$, it is enough to show that for each $u < \frac{\alpha+\gamma-1}{\gamma-1}$

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)\geq 1\}}\right] \leq A h^{\frac{\gamma-2}{2(\gamma-1)}q+u} \tag{6.3.16}$$

for some $A > 0$ that may depend on u but not on h.

Case $i(h) = 1$. Now

$$V_t = \begin{cases} \left(V_0^{1-\gamma} + (\gamma-1)Ct\right)^{\frac{1}{1-\gamma}} & \text{for } 0 \leq t \leq \tau_1 \\ \left(V_{\tau_1}^{1-\gamma} + (\gamma-1)C(t-\tau_1)\right)^{\frac{1}{1-\gamma}} & \text{for } \tau_1 \leq t \leq h, \end{cases}$$

which yields

$$\begin{aligned} \int_0^h V_t dt &= \frac{\gamma-2}{\gamma-1}\left[(V_0^{1-\gamma} + (\gamma-1)C\tau_1)^{\frac{\gamma-2}{\gamma-1}} - (V_0^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right] \\ &\quad + \frac{\gamma-2}{\gamma-1}\left[(V_{\tau_1}^{1-\gamma} + (\gamma-1)C(h-\tau_1))^{\frac{\gamma-2}{\gamma-1}} - (V_{\tau_1}^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right] \\ &=: P(h) + Q(h), \end{aligned} \tag{6.3.17}$$

and therefore

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=1\}}\right] \leq 2^{q-1}\left[\mathbb{E}\left(P^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right) + \mathbb{E}\left(Q^{q/2}(h)\right)\mathbf{1}_{\{i(h)=1\}}\right] \tag{6.3.18}$$

We now show that $\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=1\}}\right]$ can be *bounded above* as in (6.3.16):

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=1\}}\right] \leq A h^{\frac{\gamma-2}{2(\gamma-1)}q+u}, \tag{6.3.19}$$

for every $u < \frac{\alpha+\gamma-1}{\gamma-1}$. By (6.3.18) it suffices to show that both $\mathbb{E}\left(P^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right)$ and $\mathbb{E}\left(Q^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right)$ have an upper bound of the same form.
Note first that

$$P(h) \leq \frac{\gamma-2}{\gamma-1}\left[(V_0^{1-\gamma} + (\gamma-1)Ch)^{\frac{\gamma-2}{\gamma-1}} - (V_0^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right],$$

which coincides with (6.3.5), whose scaling has already been obtained. Since $\mathbb{P}(i(h) = 1) \sim h$, we have that $\mathbb{E}\left(P^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right)$ scales as the term studied in the case $i(h) = 0$, but with an extra factor h, i.e.

$$\mathbb{E}\left(P^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right) \leq A h^{1+\frac{\gamma-2}{2(\gamma-1)}q+u} \leq A h^{\frac{\gamma-2}{2(\gamma-1)}q+u}. \tag{6.3.20}$$

For the term $\mathbb{E}\left(Q^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right)$ we repeat the steps of the case $i(h) = 0$ (note that all inequalities used there held pointwise) with V_{τ_1} in place of V_0 and $h - \tau_1$ in place of h (see (6.3.9)), obtaining

$$\begin{aligned}\mathbb{E}\left(Q^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right) &\leq \mathbb{E}\left[(h-\tau_1)^{\frac{\gamma-2}{2(\gamma-1)}q}\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)C(h-\tau_1)}+1\right)^{-\frac{q}{2(\gamma-1)}}\mathbf{1}_{\{i(h)=1\}}\right] \\ &\leq h^{\frac{\gamma-2}{2(\gamma-1)}q}\mathbb{E}\left[\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)C(h-\tau_1)}+1\right)^{-\frac{q}{2(\gamma-1)}}\mathbf{1}_{\{i(h)=1\}}\right]\end{aligned} \tag{6.3.21}$$

This last term can be bounded from above as follows, for $u < \frac{q}{2(\gamma-1)}$ and using the trivial bound $V_{\tau_1} \leq V_0 + L_h$

$$\begin{aligned}\mathbb{E}\left[\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)C(h-\tau_1)}+1\right)^{-\frac{q}{2(\gamma-1)}}\mathbf{1}_{\{i(h)=1\}}\right] &\leq \mathbb{E}\left[\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)C(h-\tau_1)}+1\right)^{-u}\mathbf{1}_{\{i(h)=1\}}\right] \\ &\leq \mathbb{E}\left[\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)C(h-\tau_1)}\right)^{-\mathbf{u}}\mathbf{1}_{\{i(h)=1\}}\right] \\ &\leq \mathbb{E}\left[\left(\frac{V_{\tau_1}^{1-\gamma}}{(\gamma-1)Ch)}\right)^{-\mathbf{u}}\mathbf{1}_{\{i(h)=1\}}\right] \\ &\leq Ah^u\mathbb{E}\left[(V_0+L_h)^{u(\gamma-1)}\mathbf{1}_{\{i(h)=1\}}\right]\end{aligned} \tag{6.3.22}$$

for a constant $A > 0$. Now observe that V_0 is independent of $\mathbf{1}_{\{i(h)=1\}}$, that L_h has distribution ν conditioned to $\{i(h) = 1\}$, and that $\mathbb{P}(i(h) = 1) \leq \lambda h$. It follows that, for a suitable constant $C > 0$,

$$\begin{aligned}\mathbb{E}\left[(V_0+L_h)^{u(\gamma-1)}\mathbf{1}_{\{i(h)=1\}}\right] &\leq C\mathbb{P}(i(h)=1)\left[\mathbb{E}\left(V_0^{u(\gamma-1)}\right)+\mathbb{E}\left(L_h^{u(\gamma-1)}|i(h)=1\right)\right] \\ &= C\mathbb{P}(i(h)=1)\left[\int v^{u(\gamma-1)}\mu(dv)+\int l^{u(\gamma-1)}\nu(dl)\right].\end{aligned} \tag{6.3.23}$$

Since, by Theorem 6.9, the tails of μ are lighter that those of ν, the above integrals are both finite if an only if $\int v^{u(\gamma-1)}\nu(dv) < +\infty$, which holds true for $u < \alpha/(\gamma-1)$ (assumption A3). Thus, for every $u < \alpha/(\gamma-1)$,

$$\mathbb{E}\left[(V_0+L_h)^{u(\gamma-1)}\mathbf{1}_{\{i(h)=1\}}\right] \leq Ah \tag{6.3.24}$$

for some $A > 0$. By (6.3.21), (6.3.22), (6.3.23) and (6.3.24), we have that $\mathbb{E}\left(Q^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right)$ is bounded from above by

$$Bh^{\frac{\gamma-2}{2(\gamma-1)}q+u+1}$$

for every $u < \alpha/(\gamma-1)$ and some $B > 0$ possibly depending on u. Equivalently,

$$\mathbb{E}\left(Q^{q/2}(h)\mathbf{1}_{\{i(h)=1\}}\right) \le Bh^{\frac{\gamma-2}{2(\gamma-1)}q+u} \tag{6.3.25}$$

for all $u < \frac{\alpha+\gamma-1}{\gamma-1}$. Therefore, by (6.3.20) and (6.3.25), (6.3.19) is established.
Case $i(h) \ge 2$. To prove (6.3.16) and thus to complete the whole proof, we are left to show that

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)\ge 2\}}\right] \le Ah^{\frac{\gamma-2}{2(\gamma-1)}q+u} \tag{6.3.26}$$

for all $u < \frac{\alpha+\gamma-1}{\gamma-1}$ and some $A > 0$.

Let $n \ge 2$, and restrict to the event $\{i(h) = n\}$. We have

$$V_t = \begin{cases} \left(V_0^{1-\gamma} + (\gamma-1)Ct\right)^{\frac{1}{1-\gamma}} & \text{for } 0 \le t \le \tau_1 \\ \left(V_{\tau_1}^{1-\gamma} + (\gamma-1)C(t-\tau_1)\right)^{\frac{1}{1-\gamma}} & \text{for } \tau_1 \le t \le \tau_2 \\ \vdots & \\ \left(V_{\tau_{n-1}}^{1-\gamma} + (\gamma-1)C(t-\tau_{n-1})\right)^{\frac{1}{1-\gamma}} & \text{for } \tau_{n-1} \le t \le \tau_n \\ \left(V_{\tau_n}^{1-\gamma} + (\gamma-1)C(t-\tau_n)\right)^{\frac{1}{1-\gamma}} & \text{for } \tau_n \le t \le h, \end{cases}$$

so that (6.3.17) becomes

$$\begin{aligned} \int_0^h V_t dt &= \frac{\gamma-2}{\gamma-1}\sum_{k=1}^{n}\left[(V_{k-1}^{1-\gamma} + (\gamma-1)C(\tau_k-\tau_{k-1}))^{\frac{\gamma-2}{\gamma-1}} - (V_{k-1}^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right] \\ &+ \frac{\gamma-2}{\gamma-1}\left[(V_{\tau_n}^{1-\gamma} + (\gamma-1)C(h-\tau_n))^{\frac{\gamma-2}{\gamma-1}} - (V_{\tau_n}^{1-\gamma})^{\frac{\gamma-2}{\gamma-1}}\right] \\ &=: \sum_{k=1}^{n} P_k(h) + P_{n+1}(h). \end{aligned} \tag{6.3.27}$$

Each term $\mathbb{E}\left[P_k^{q/2}(h)\mathbf{1}_{\{i(h)=n\}}\right]$ can be estimated as in (6.3.21) and (6.3.22), obtaining

$$\begin{aligned} \mathbb{E}\left[P_k^{q/2}(h)\mathbf{1}_{\{i(h)=n\}}\right] &\le Ch^{\frac{\gamma-2}{2(\gamma-1)}q+u}\mathbb{E}\left[(V_0+L_h)^{u(\gamma-1)}\mathbf{1}_{\{i(h)=n\}}\right] \\ &\le C'h^{\frac{\gamma-2}{2(\gamma-1)}q+u}\mathbb{P}(i(h)=n)\left[\mathbb{E}\left(V_0^{u(\gamma-1)}\right) + \mathbb{E}\left(L_h^{u(\gamma-1)}|i(h)=n\right)\right] \end{aligned} \tag{6.3.28}$$

for $u < \frac{q}{2(\gamma-1)}$ and some constant C, C' that may depend on u but not on n and h. The distribution of L_h *given* $\{i(h) = n\}$ is given by the n-fold convolution ν^{*n}. In other words, if $X_1, X_2, \ldots, X_n$ are independent random variables with law ν,

$$\mathbb{E}\left(L_h^{u(\gamma-1)}|i(h)=n\right) = \mathbb{E}\left[(X_1+X_2+\cdots+X_n)^{u(\gamma-1)}\right] \le n^{u(\gamma-1)-1}\mathbb{E}\left[X_1^{u(\gamma-1)}\right].$$

For $u < \alpha/(\gamma-1)$, $\mathbb{E}\left[X_1^{u(\gamma-1)}\right] < +\infty$ as well as $\mathbb{E}\left(V_0^{u(\gamma-1)}\right) < +\infty$. Thus

$$\mathbb{E}\left[P_k^{q/2}(h)\mathbf{1}_{\{i(h)=n\}}\right] \le Ch^{\frac{\gamma-2}{2(\gamma-1)}q+u} n^{u(\gamma-1)-1}\mathbb{P}(i(h)=n), \tag{6.3.29}$$

for some constant C independent of n, h and k. By (6.3.27), (6.3.28) and (6.3.29) we obtain, for $u < \frac{q}{2(\gamma-1)}$,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=n\}}\right] &\le n^{q/2}\sum_{k=1}^{n+1}\mathbb{E}\left[P_k^{q/2}(h)\mathbf{1}_{\{i(h)=n\}}\right] \\ &\le Ch^{\frac{\gamma-2}{2(\gamma-1)}q+u} n^{u(\gamma-1)+q/2}\mathbb{P}(i(h)=n).\end{aligned} \tag{6.3.30}$$

We can now sum over $n \ge 2$, observing that $\mathbb{P}(i(h)=n) \le \frac{\lambda^n h^n}{n!}$:

$$\begin{aligned}\sum_{n\ge 2}\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=n\}}\right] &\le Ch^{\frac{\gamma-2}{2(\gamma-1)}q+u+2}\sum_{n\ge 2} n^{u(\gamma-1)+q/2}\frac{\lambda^n h^{n-2}}{n!} \\ &\le C'h^{\frac{\gamma-2}{2(\gamma-1)}q+u+2}.\end{aligned} \tag{6.3.31}$$

Since $\frac{q}{2(\gamma-1)} + 2 > \frac{\alpha+\gamma-1}{\gamma-1}$ (recall that $q > 2(\alpha+\gamma-1)$), we have that

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)\ge 2\}}\right]$$

is negligible with respect to (6.3.15).
This completes the proof for the case $\gamma > 2$.
Part 2: $1 < \gamma < 2$

Case $i(h) = 0$. Formula (6.3.5) still hold, but now $\gamma - 2 < 0$. So (6.3.6) becomes

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=0\}}\right] =$$

$$\left(\frac{2-\gamma}{\gamma-1}\right)^{q/2}((\gamma-1)Ch)^{\frac{\gamma-2}{2(\gamma-1)}q}\mathbb{E}\left[\left[\left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}\right)^{\frac{\gamma-2}{\gamma-1}} - \left(\frac{V_0^{1-\gamma}}{(\gamma-1)Ch}+1\right)^{\frac{\gamma-2}{\gamma-1}}\right]^{q/2}\right]e^{-\lambda h}. \tag{6.3.32}$$

To estimate this last expression we need, letting $a := \frac{2-\gamma}{\gamma-1}$, the following modifications of (6.3.7), valid for $z > 0$:

$$\begin{aligned}&a(z+1)^{-1}z^{-a} \le z^{-a} - (z+1)^{-a} \le (z+1)^{-1}z^{-a} \quad \text{for } 0 < a \le 1 \\ &(z+1)^{-1}z^{-a} \le z^{-a} - (z+1)^{-a} \le a(z+1)^{-1}z^{-a} \quad \text{for } a > 1.\end{aligned} \tag{6.3.33}$$

Using these inequalities as in (6.3.6) we obtain, for some $C > 1$

$$\frac{1}{C}\mathbb{E}\left[\left(\frac{V_0^{\gamma-1}h}{1+V_0^{\gamma-1}h}V_0^{2-\gamma}\right)^q\right] \le \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \le C\mathbb{E}\left[\left(\frac{V_0^{\gamma-1}h}{1+V_0^{\gamma-1}h}V_0^{2-\gamma}\right)^q\right]. \tag{6.3.34}$$

We now observe that

$$\begin{aligned}\mathbb{E}\left[\left(\frac{V_0^{\gamma-1}h}{1+V_0^{\gamma-1}h}V_0^{2-\gamma}\right)^{q/2}\right] &= \mathbb{E}\left[\left(\frac{V_0^{\gamma-1}h}{1+V_0^{\gamma-1}h}V_0^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{V_0^{\gamma-1}h\le 1\}}\right] \\ &\quad + \mathbb{E}\left[\left(\frac{V_0^{\gamma-1}h}{1+V_0^{\gamma-1}h}V_0^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{V_0^{\gamma-1}h> 1\}}\right] \\ &\sim h^{q/2}\mathbb{E}\left[V_0^{q/2}\mathbf{1}_{\{V_0^{\gamma-1}h\le 1\}}\right] + \mathbb{E}\left[V_0^{\frac{q}{2}(2-\gamma)}\mathbf{1}_{\{V_0^{\gamma-1}h> 1\}}\right].\end{aligned} \tag{6.3.35}$$

In order to estimate the two summand of the left hand side of (6.3.35) we use the following fact, whose simple proof follows from simple point wise bounds, and it is omitted. Let μ be a probability on $[0, +\infty)$ such that $\mu((u, +\infty))$ is regularly varying with exponent $-\xi < 0$. Then

$$\int_0^x u^p \mu(du) \sim x^{p-\xi} \qquad \text{for } p > \xi \tag{6.3.36}$$

$$\int_x^{+\infty} u^p \mu(du) \sim x^{p-\xi} \qquad \text{for } p < \xi. \tag{6.3.37}$$

Let μ be the law of V_0, so that, by Theorem 6.9, $\xi = \alpha+\gamma-1$. Since $q > 2(\alpha+\gamma-1)$, by (6.3.36) we have $\mathbb{E}\left[V_0^{q/2}\mathbf{1}_{\{V_0^{\gamma-1}h\le 1\}}\right] \sim h^{-\frac{1}{\gamma-1}(\frac{q}{2}-\alpha-\gamma+1)}$, and therefore

$$h^{q/2}\mathbb{E}\left[V_0^{q/2}\mathbf{1}_{\{V_0^{\gamma-1}h\le 1\}}\right] \sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}}. \tag{6.3.38}$$

Moreover, by (6.3.37), also

$$\mathbb{E}\left[V_0^{\frac{q}{2}(2-\gamma)}\mathbf{1}_{\{V_0^{\gamma-1}h> 1\}}\right] \sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}}. \tag{6.3.39}$$

for $\frac{q}{2}(2-\gamma) < \alpha+\gamma-1$, while

$$\mathbb{E}\left[V_0^{\frac{q}{2}(2-\gamma)}\mathbf{1}_{\{V_0^{\gamma-1}h> 1\}}\right] = +\infty \tag{6.3.40}$$

for $\frac{q}{2}(2-\gamma) > \alpha+\gamma-1$.

Summing up, we have shown that

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=0\}}\right] &\sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}} \quad \text{for } 2(\alpha+\gamma-1) < q < \frac{2(\alpha+\gamma-1)}{2-\gamma} \\ \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=0\}}\right] &= +\infty \quad \text{for } q > \frac{2(\alpha+\gamma-1)}{2-\gamma}.\end{aligned} \tag{6.3.41}$$

Case $i(h) = 1$. This case is dealt with as for $\gamma > 2$, and, using the same argument leading to (6.3.34), one sees that the crucial term to estimate is

$$\mathbb{E}\left[\left(\frac{V_{\tau_1}^{\gamma-1}(h-\tau_1)}{1+V_{\tau_1}^{\gamma-1}(h-\tau_1)}V_{\tau_1}^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right]. \tag{6.3.42}$$

Since $V_{\tau_1} \geq L_{\tau_1}$, (6.3.42) can be bounded from below by

$$\mathbb{E}\left[\left(\frac{L_{\tau_1}^{\gamma-1}(h-\tau_1)}{1+L_{\tau_1}^{\gamma-1}(h-\tau_1)}L_{\tau_1}^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right]$$

which takes the value infinity as soon as $\mathbb{E}\left[\left(L_{\tau_1}^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right] = +\infty$. Recalling that L_{τ_1} independent of $\{i(h) = 1\}$ and it has law ν, this holds as $q > \frac{2\alpha}{2-\gamma}$. This implies that

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right] = +\infty \quad \text{for } q > \frac{2\alpha}{2-\gamma}. \tag{6.3.43}$$

Comparing with (6.3.41), note that $\frac{2\alpha}{2-\gamma} < \frac{2(\alpha+\gamma-1)}{2-\gamma}$. Thus, assume $2(\alpha+\gamma-1) < q < \frac{2\alpha}{2-\gamma}$ (note that, being by assumption $\alpha+\gamma > 2$, indeed $2(\alpha+\gamma-1) < \frac{2\alpha}{2-\gamma}$). An upper bound for (6.3.42) is given by

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{V_{\tau_1}^{\gamma-1}(h-\tau_1)}{1+V_{\tau_1}^{\gamma-1}(h-\tau_1)}V_{\tau_1}^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right] \\
&\qquad \leq \mathbb{E}\left[\left(\frac{(V_0+L_{\tau_1})^{\gamma-1}h}{1+(V_0+L_{\tau_1})^{\gamma-1}h}(V_0+L_{\tau_1})^{2-\gamma}\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right] \\
&\qquad = \mathbb{E}\left[\left(\frac{(V_0+L_{\tau_1})^{\gamma-1}h}{1+(V_0+L_{\tau_1})^{\gamma-1}h}(V_0+L_{\tau_1})^{2-\gamma}\right)^{q/2}\right]\mathbb{P}(i(h)=1),
\end{aligned} \tag{6.3.44}$$

where we used the facts that V_0 and L_{τ_1} are independent of $\{i(h) = 1\}$. Now, (6.3.44) is estimated exactly as (6.3.35), but with $V_0 + L_{\tau_1}$ in place of V_0. Since the tails of $V_0 + L_{\tau_1}$ are the same as those of L_{τ_1}, i.e. regularly varying with exponent α, while $\mathbb{P}(i(h) = 1) \sim h$, we get

$$\mathbb{E}\left[\left(\frac{(V_0+L_{\tau_1})^{\gamma-1}h}{1+(V_0+L_{\tau_1})^{\gamma-1}h}(V_0+L_{\tau_1})^{2-\gamma}\right)^{q/2}\right] \sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha}{\gamma-1}}\mathbb{P}(i(h)=1) \sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}}.$$

Summing up:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=1\}}\right] &\sim h^{\frac{\gamma-2}{2(\gamma-1)}q+\frac{\alpha+\gamma-1}{\gamma-1}} && \text{for } 2(\alpha+\gamma-1) < q < \frac{2\alpha}{2-\gamma} \\
\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\mathbf{1}_{\{i(h)=0\}}\right] &= +\infty && \text{for } q > \frac{2\alpha}{2-\gamma}.
\end{aligned} \tag{6.3.45}$$

Case $i(h) \geq 2$. This case goes along the same line as for $\gamma > 2$, using the upper bound obtained for $i(h) = 1$. The details are omitted. The proof for $1 < \gamma < 2$ is thus completed.

Part 3: $\gamma = 2$

In this case we have, in the case of no jumps ($i(h) = 0$),

$$V_t = \left(V_0^{-1} + Ct\right)^{-1}$$

and therefore

$$\int_0^h V_t dt = \frac{1}{C}\left[\log(V_0^{-1} + Ch) - \log(V_0^{-1})\right] = \frac{1}{C}\left[\log(1 + ChV_0)\right]. \qquad (6.3.46)$$

An upper bound for $\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right]$ is obtained using (6.3.46) and the inequality, valid for $y, r > 0$,

$$\log(1 + y) \leq \frac{1}{r} y^r,$$

which gives

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \leq \frac{1}{C^{q/2}} C^{rq/2} h^{rq/2} \mathbb{E}\left(V_0^{rq/2}\right).$$

Since $\mathbb{E}\left(V_0^{rq/2}\right) < +\infty$ for $\frac{rq}{2} < \alpha + 1$, letting $\frac{rq}{2} \uparrow \alpha + 1$ we obtain

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \leq C h^{rq/2}, \qquad (6.3.47)$$

for some $C > 0$ and every r such that $\frac{rq}{2} < \alpha + 1$. A corresponding lower bound is obtained using the inequality

$$\log(1 + y) \geq \frac{1}{2} \mathbf{1}_{(1,+\infty)}(y),$$

which gives

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \geq \frac{1}{(2C)^{q/2}} \mathbb{P}(ChV_0 > 1) \sim h^{\alpha+1}, \qquad (6.3.48)$$

where we have used Theorem 6.9 for the last inequality. By (6.3.47) and (6.3.48) we have

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2} \mathbf{1}_{\{i(h)=0\}}\right] \sim h^{\alpha+1}.$$

The cases with $i(h) \geq 1$ are similar to what seen in Parts 1 and 2, and are omitted.

□

Proof of Theorem 6.10. We now complete the proof of Theorem 6.10. We need to extend Proposition 6.12 in two directions: a) generalize from $f(v) = Cv^\gamma$ to any f satisfying Assumption **A4**; b) extend to Lévy subordinator satisfying Assumptions **A2** and **A3**, thus with a possibly infinite characteristic measure ν.
Step 1. We keep all assumption of Proposition 6.12, except that we require $f(v) = Cv^\gamma$ only for $v > \epsilon$, for some $\epsilon > 0$, and f satisfies Assumption **A4**. In other words we do not prescribe the asymptotics of f near $v = 0$. Let V, V' be solutions, respectively, of the equations

$$\begin{aligned} dV_t &= -f(V_t)dt + dL_t \\ dV'_t &= -CV'^\gamma_t + dL_t. \end{aligned}$$

Assume $V_0 = V'_0 = v > 0$. We claim that

$$|V_t - V'_t| \leq 2\epsilon \tag{6.3.49}$$

a.s., for every $t \geq 0$. This follows from the following fact: there is a constant $\delta > 0$ such that as soon as $|V_t - V'_t| \geq 2\epsilon$,

$$\frac{d}{dt}|V_t - V'_t| \leq -\delta. \tag{6.3.50}$$

To see (6.3.50), suppose first $V_t - V'_t \geq 2\epsilon$. In particular $V_t \geq \epsilon$, so

$$\frac{d}{dt}[V_t - V'_t] = -C\left(V_t^\gamma - V'^\gamma_t\right) < -C(2\epsilon)^\gamma,$$

where we have used the fact that, for $c > 0$, the map $(x+c)^\gamma - x^\gamma$ is increasing for $x > 0$. Suppose now $V'_t - V_t \geq 2\epsilon$. If $V_t \geq \epsilon$ then,

$$\frac{d}{dt}[V'_t - V_t] = -C\left(V'^\gamma_t - V_t^\gamma\right) < -C(2\epsilon)^\gamma;$$

If $V_t < \epsilon$, since f is increasing, then

$$\frac{d}{dt}[V'_t - V_t] = -CV'^\gamma_t + f(V_t) \leq -C(2\epsilon)^\gamma + C\epsilon^\gamma < 0.$$

Thus (6.3.50), and so (6.3.49) is proved. In particular, the law of V_t is stochastically smaller than that of $V'_t + 2\epsilon$, which means that for every g increasing and bounded, $E[g(V_t)] \leq E[g(V'_t + 2\epsilon)]$. By the ergodicity results proved in [89], this inequality can be taken to the limit as $t \to +\infty$, so to a stochastic inequality between the stationary distributions of V and V'. This implies that we can realize, on a suitable probability space, two random variables V_0 and V'_0, independent of the Lévy process L, distributed according to the stationary laws of the corresponding processes, and such that $V_0 \leq V'_0 + 2\epsilon$. By repeating the argument above, we see that the inequality

$V_t \le V'_t + 2\epsilon$ is a.s. preserved for all $t > 0$ also for the stationary processes. It follows that

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\right] \le \mathbb{E}\left[\left(\int_0^h [V'_t + 2\epsilon] dt\right)^{q/2}\right] \le 2^{q/2-1}\left\{\mathbb{E}\left[\left(\int_0^h V'_t dt\right)^{q/2}\right] + (2\epsilon h)^{q/2}\right\}. \tag{6.3.51}$$

Since

$$A(q) = \limsup_{h\to 0} \frac{\log \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\right]}{\log h}, \tag{6.3.52}$$

and

$$A'(q) = \lim_{h\to 0} \frac{\log \mathbb{E}\left[\left(\int_0^h V'_t dt\right)^{q/2}\right]}{\log h} \le \frac{q}{2}, \tag{6.3.53}$$

by (6.3.51) we get

$$A(q) \ge A'(q).$$

By exchanging the role of V and V' we get $A(q) = A'(q)$. Moreover, the existence of the limit (6.3.53), which follows from Proposition 6.12, implies that also (6.3.52) is a limit. Since $A'(q)$ is given by Proposition 6.12, the first extension is obtained.

Step 2. In this step we allow the Lévy process L to have infinite characteristic measure ν and *positive drift* $m > 0$, though satisfying Assumptions **A2** and **A3**. On the other hand we make a specific choice for f: $f(v) = Cv^\gamma$ for $v > \epsilon$, while f is linear in $[0, \epsilon)$, with $f(0) = 0$ and $f(\epsilon) = C\epsilon^\gamma$. Moreover we let $\nu_\epsilon := \nu \mathbf{1}_{[\epsilon,+\infty)}$, which is a finite measure. Denote by $L^{(\epsilon)}$ the compound Poisson process with characteristic measure ν^ϵ, and by $V^{(\epsilon)}$ the solution of

$$dV_t^{(\epsilon)} = -f(V_t^{(\epsilon)})dt + dL_t^{(\epsilon)} \tag{6.3.54}$$

The original Lévy process L can be decomposed in the form $L_t = L^{(\epsilon)} + L^{(<\epsilon)}$, where $L^{(<\epsilon)}$ is independent of $L^{(\epsilon)}$, it has characteristic measure $\nu_\epsilon := \nu \mathbf{1}_{[0,\epsilon)}$ and drift $m > 0$. Writing

$$dV_t = -f(V_t)dt + dL_t, \tag{6.3.55}$$

we obtain

$$d(V_t - V_t^{(\epsilon)}) = -[f(V_t) - f(V_t^{(\epsilon)})]dt + dL^{(<\epsilon)}. \tag{6.3.56}$$

This implies, for instance that whenever $V_0^{(\epsilon)} \le V_0$, then $V_t^{(\epsilon)} \le V_t$ for all $t > 0$. Thus, using as above the ergodicity of V and $V^{(\epsilon)}$, V_t dominates stochastically $V_t^{(\epsilon)}$ also in equilibrium. Thus, as before, we can start the processes from $V_0^{(\epsilon)}$ and V_0, each having the corresponding stationary distribution, and such that $V_0^{(\epsilon)} \le V_0$. Thus $V_t^{(\epsilon)} \le V_t$ for all $t > 0$. Note that with this construction we have that the two processes in (6.3.54) and (6.3.55) are separately stationary, by not necessarily the Markov process $(V_t^{(\epsilon)}, V_t)$, whose law will be denoted by $\mu_t^{(2)}$, is stationary. To fix this we observe that, since the family of distribution $(\mu_t^{(2)})_{t\ge 0}$ is tight, by a standard argument its Cesaro means $\frac{1}{t}\int_0^t \mu_s^{(2)} ds$ admit at least a limit point, which is a stationary distribution for $(V_t^{(\epsilon)}, V_t)$. This limiting operation preserves the stochastic order between the laws of the two components. Thus, we can assume to realize $V_0^{(\epsilon)}$ and V_0 in such a way their *joint* distribution is stationary for (6.3.54) and (6.3.55), and $V_0^{(\epsilon)} \leq V_0$.

Now we use the fact that f is superlinearly increasing, to conclude that

$$f(V_t) - f(V_t^{(\epsilon)}) \geq c[V_t - V_t^{(\epsilon)}]$$

for some $c > 0$. It follows that

$$d(V_t - V_t^{(\epsilon)}) \leq -c[V_t - V_t^{(\epsilon)}] + dL_t^{(<\epsilon)},$$

which implies that

$$0 \leq V_t - V_t^{(\epsilon)} \leq e^{-ct}[V_0 - V_0^{(\epsilon)}] + \int_0^t e^{-c(t-s)} dL_s^{(<\epsilon)}. \tag{6.3.57}$$

Since the law of $V_t - V_t^{(\epsilon)}$ does not depend on t, it must be stochastically dominated by the limit of the law of the r.h.s. of (6.3.57), which is just the stationary distribution of the Ornstein-Uhlenbeck process

$$dZ_t = -cZ_t dt + dL_t^{(<\epsilon)}.$$

As observed e.g. in [51], this stationary law is infinitely divisible with characteristic pair $(m, \tilde{\nu})$, with

$$\tilde{\nu}([x, +\infty)) = \int_{[x,+\infty)} u^{-1} \nu_\epsilon(du).$$

Since ν_ϵ, and therefore $\tilde{\nu}$, has bounded support, the stationary law of Z_t has moments of all order (see e.g. [99]). So, also $V_t - V_t^{(\epsilon)}$ has moments of all order. Thus, using the inequality $(x+y)^q \leq 2^{q-1}[x^q + y^q]$ for $x, y \geq 0$, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^h V_t^{(\epsilon)} dt\right)^{q/2}\right] &\leq \mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\right] \\
&\leq 2^{q/2-1}\left\{\mathbb{E}\left[\left(\int_0^h V_t^{(\epsilon)} dt\right)^{q/2}\right] + \mathbb{E}\left[\left(\int_0^h [V_t - V_t^{(\epsilon)}] dt\right)^{q/2}\right]\right\} \\
&\leq 2^{q/2-1}\left\{\mathbb{E}\left[\left(\int_0^h V_t^{(\epsilon)} dt\right)^{q/2}\right] + h^{q/2}\mathbb{E}\left[\left(V_0 - V_0^{(\epsilon)}\right)^{q/2}\right]\right\}.
\end{aligned}$$

Since, by Proposition 6.12 and steps 1, $\mathbb{E}\left[\left(\int_0^h V_t^{(\epsilon)} dt\right)^{q/2}\right] \sim h^{A(q)}$ for every $\epsilon > 0$ and $A(q) \leq q/2$, it follows that

$$\mathbb{E}\left[\left(\int_0^h V_t dt\right)^{q/2}\right] \sim h^{A(q)},$$

thus completing the proof of this step.
Step 3. The extension of Proposition 6.12 to any f which satisfies (**A4**) is now easy, and it will only be sketched. In a first stage, repeating the argument in step 1, one extends from the special f's used for step 2, to the larger class of f in step 1.

The further extension to a general f which satisfies (**A4**) proceeds as follows: for every $\delta > 0$ we can find f_1 and f_2 such that $f_1 \leq f \leq f_2$, and $f_1(v) = C_1 v^{\gamma-\delta}$, $f_2(v) = C_2 v^{\gamma+\delta}$ for $v > \epsilon$. By using coupling arguments similar to those in step 1, one shows that the scaling function $A(q)$ of the process with drift f is bounded above and below by the scaling functions of the processes with drift f_1 and f_2. The continuity of $A(q)$ w.r.t. γ, and the fact that δ is arbitrary, implies that $A(q)$ is given by Proposition 6.12.

□

6.4 Modeling financial data

In this Section we briefly discuss potential applications of the proposed models to Mathematical Finance. The phenomenon that has initially motivated the introduction of the model is that of multiscaling of moments. Multiscaling has been detected, with similar features, in many time series of financial indices, including DJIA, S&P 500, FTSE 100, Nikkei 225, as well as in various currency exchange rates. We refer to [44, 54, 56, 95, 1, 22] for details. In all cases the scaling function $A(q)$ is estimated from daily data. The empirical q-th moment is computed from data, as function of the (discrete) time width h (see (6.1.1)). The log-log plot of this function is approximately linear for small h, and the slope $\hat{A}(q)$ is an estimator for the scaling function $A(q)$. The curve $q \mapsto \hat{A}(q)$ obtained is a smooth, concave function, which may seem to contradict the piecewise linear prediction for $A(q)$ of the model. However, if the scaling function is estimated form "daily" data *simulated* from the model, the same smoothing appears (see Figure 6.2). This shows that the model is consistent with the observed data on multiscaling of moments. The fact that the model we have introduced is an independent time change of a Brownian motion, allows direct applications in option prices. Indeed, consider the risk-neutral measure under which the price $S_t := e^{X_t}$ with, say $S_0 = 1$, evolves according to

$$\frac{dS_t}{S_t} = \sigma_t dB_t,$$

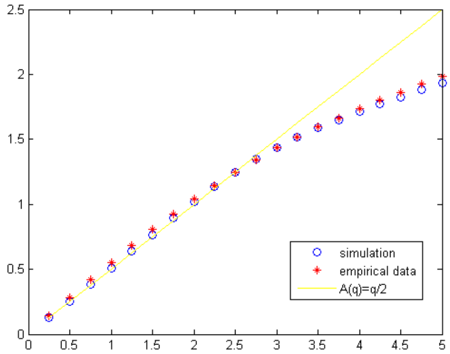
where B is a Brownian motion, and σ_t evolves as in the previous section, and it is independent of B. By Ito's rule

$$dX_t = \sigma_t dB_t - \frac{1}{2}\sigma_t^2 dt. \tag{6.4.1}$$

Note that this martingale measure is not unique, as one could modify the evolution of σ_t, allowing flexibility when the model is calibrated to prices. Equation (6.4.1) implies that

$$S_t = e^{X_t} = e^{W_{I_t} - \frac{1}{2}I_t}, \tag{6.4.2}$$

Figure 6.2: Points indicated with stars refer to the curve $q \to \hat{A}(q)$ computed on the DJIA time series (opening prices 1935-2013). Circles represent data obtained by first simulating the model with estimated values of the parameters, and then by computing $q \to \hat{A}(q)$ from the simulated time series.



where W is a Brownian motion, and $I_t := \int_0^t \sigma_s^2 ds$ is independent of W. Thus, under the risk-neutral measure, the price S_t is a time-changed geometric Brownian motion, with independent time-change process. For the special $\gamma > 2$ and the characteristic measure of the Lévy subordinator L_t concentrated on $+\infty$, treated in chapter 5, the representation (6.4.2) is the basis of the computation of sharp asymptotics for the *implied volatility* surface in the regime of small maturity or large strike (see [31, 32]). Moreover, it is shown that the parameters of the model can be tuned to reproduce quite realistic *smile*-shapes for the implied volatility. We remark that the model introduced here, if adopted in a pricing context, would suffer from symmetric smiles. Asymmetry could be introduced adding a Brownian term of Heston type in the equation for the volatility:

$$dV_t = -CV_t^\gamma dt + dL_t + \bar{C}\sqrt{V_t} d\tilde{W}_t,$$

where $\tilde{W}$ is a Brownian Motion correlated with W_t.

6.5 Decay of autocorrelation

In this section we consider models of the form 6.3.1, but we the specific assumptions of Proposition 6.12. In particular, $f(v) = Cv^\gamma$, with $C > 0$, L_t is a Lévy subordinator with zero drift and finite characteristic measure ν (a compound Poisson process). Recall also that $u \mapsto \nu((u, +\infty))$ is regularly varying at infinity with exponent $-\alpha < 0$. More explicitly,

$$\begin{aligned} dX_t &= \sqrt{V_t} dB_t \\ dV_t &= -CV_t^\gamma dt + dL_t \end{aligned}$$

We prove a result in the spirit of theorem 5.5: autocorrelation of squared returns decays exponentially in time at infinity. We define

$$\rho_2(t) := \lim_{h\to 0} corr(|X_h - X_0|^2, |X_{t+h} - X_t|^2).$$

This quantity and volatility autocorrelation $\rho(t) := \lim_{h\to 0} corr(|X_h - X_0|, |X_{t+h} - X_t|)$, that we have considered in theorem 5.5, have a similar meaning, and both have been investigated in the literature (see [45] for details). Define, for $\rho \in (0, 1)$,

$$\alpha_\rho = \rho(1/\rho + \ln\rho - 1), \qquad \beta_\rho = \rho \ln \mathbb{E}\left[(1 + (\gamma - 1)C\,\theta\,\Delta L^{\gamma-1})^{\frac{\gamma}{1-\gamma}}\right],$$

where $\theta \sim \exp(\lambda)$ and $\Delta L \sim \nu$. Remark $\alpha_\rho, \beta_\rho > 0$, $\forall \rho \in (0, 1)$. We set

$$r := \sup_{\rho\in(0,1)} \min\{\alpha_\rho, \beta_\rho, 1\}$$

We have the following estimates for the decay at ∞ of the autocorrelation of squared returns.

Theorem 6.13. *If* $\alpha + \gamma > 3$,

$$-\lambda \le \liminf_{t\to\infty} \frac{\ln(\rho_2(t))}{t} \le \limsup_{t\to\infty} \frac{\ln(\rho_2(t))}{t} \le -\lambda r < 0.$$

Proof. We introduce the notation

$$f_t : v \to (v^{1-\gamma} + (\gamma - 1)Ct)^{\frac{1}{1-\gamma}}.$$

We have

$$f_t'(v) = (1 + (\gamma - 1)Ctv^{\gamma-1})^{\frac{\gamma}{1-\gamma}},$$

which is a decreasing function. We recall that $(\tau_k)_{k\ge 1}$ is the set of positive jump times, and $\tau_0 = 0$. Given $h > 0$, $i(h)$ is the random number of jump times in the interval $(0, h]$. We also denote $\theta_k = \tau_k - \tau_{k-1}$, $k \ge 1$.

$$\begin{aligned}\rho_2(t) &= \lim_{h\to 0} \frac{Cov(|X_h - X_0|^2, |X_{t+h} - X_t|^2)}{\sqrt{Var(|X_h - X_0|^2)Var(|X_{t+h} - X_t|^2)}}\\ &= \lim_{h\to 0} \frac{\mathbb{E}(|X_h - X_0|^2|X_{t+h} - X_t|^2) - \mathbb{E}|X_h - X_0|^2\mathbb{E}|X_{t+h} - X_t|^2}{\mathbb{E}|X_h - X_0|^4 - (\mathbb{E}|X_h - X_0|^2)^2}.\end{aligned}$$

Now, $\mathbb{E}V_0^q < \infty$ for $q < \alpha + \gamma - 1$, and $2 < \alpha + \gamma - 1$. Recall $X_{t+h} - X_t = \int_t^{t+h} \sqrt{V_s}dW_s$. Using uniform integrability analogously to what has been done in the proof of proposition 6.1 we can prove

$$\begin{aligned}&\lim_{h\downarrow 0} \frac{\mathbb{E}(|X_h - X_0|^2|X_{t+h} - X_t|^2) - \mathbb{E}|X_h - X_0|^2\mathbb{E}|X_{t+h} - X_t|^2}{h^2}\\ &\quad= \frac{2}{\pi}\mathbb{E}\,|V_0V_t| - \frac{2}{\pi}\mathbb{E}\,|V_0|\,\mathbb{E}\,|V_t| = \frac{2}{\pi}Cov(V_0, V_t).\end{aligned}$$

The same procedure applied to the variance in the denominator leads to

$$\rho_2(t) = corr(V_0, V_t) = \frac{Cov(V_0, V_t)}{Var(V_0)}.$$

So, to prove the statement we just have to prove

$$-\lambda \leq \liminf_{t\to\infty} \frac{\ln(Cov(V_0, V_t))}{t} \leq \limsup_{t\to\infty} \frac{\ln(Cov(V_0, V_t))}{t} \leq -\lambda r < 0.$$

Now,

$$Cov(V_0, V_t) = Cov(V_0, V_t 1_{\{i(h)=0\}}) + Cov(V_0, V_t 1_{\{i(h)\geq 1\}})$$

On $\{i(h) = 0\}$, $V_t = f_t(V_0)$. The first summand is

$$\begin{aligned} Cov(V_0, V_t 1_{\{i(t)=0\}}) &= \mathbb{E}[V_0 f_t(V_0) 1_{\{i(t)=0\}}] - \mathbb{E}V_0 \mathbb{E}[f_t(V_0) 1_{\{i(t)=0\}}] \\ &= \mathbb{E}[V_0 f_t(V_0)]\mathbb{P}(i(t) = 0) - \mathbb{E}V_0 \mathbb{E}[f_t(V_0)]\mathbb{P}(i(t) = 0) \\ &= Cov(V_0, f_t(V_0))e^{-\lambda t}. \end{aligned}$$

This term easily gives

$$-\lambda \leq \liminf_{t\to\infty} \frac{\ln(Cov(V_0, V_t 1_{\{i(h)=0\}}))}{t}.$$

Now we consider $Cov(V_0, V_t 1_{\{i(t)\geq 1\}})$. We can express $V_t 1_{\{i(t)\geq 1\}}$ as $g(V_0, Y)$ where g is increasing in V_0 and Y and V_0 are independent r.v.s. Therefore

$$\begin{aligned} Cov(V_0, g(V_0, Y)) &= \mathbb{E}[V_0 g(V_0, Y)] - \mathbb{E}V_0 \mathbb{E}g(V_0, Y) \\ &= \mathbb{E}[\mathbb{E}[V_0 g(V_0, Y)|Y]] - \mathbb{E}[V_0]\mathbb{E}[\mathbb{E}[g(V_0, Y)|Y]] \\ &= \mathbb{E}\big[\mathbb{E}[V_0 g(V_0, Y)|Y] - \mathbb{E}[V_0|Y]\mathbb{E}[g(V_0, Y)|Y]\big]. \end{aligned}$$

$\mathbb{E}[V_0 g(V_0, Y)|Y] - \mathbb{E}[V_0|Y]\mathbb{E}[g(V_0, Y)|Y]$ is a.s. positive, because of the properties of covariance and because g is increasing in V_0. So also its expectation is positive, $Cov(V_0, V_t 1_{\{i(t)\geq 1\}}) \geq 0$ and the lower bound is proved. Now we prove an upper bound for the same term. On $\{i(t) \geq 1\}$

$$V_t = f_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}} + f_{\theta_{i(t)}}(\Delta L_{\tau_{i(t)-1}} + f_{\theta_{i(t)-1}}(\ldots f_{\theta_1}(V_0) + \Delta L_{\tau_1} \ldots)))$$

It is easy to see that

$$\begin{aligned} \frac{dV_t}{dV_0} &= f'_{t-\tau_{i(t)}}(...) \times ... \times f'_{\theta_1}(...) \\ &\leq f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) \times ... \times f'_{\theta_2}(\Delta L_{\tau_1}) \times f'_{\theta_1}(V_0). \end{aligned}$$

Defining

$$\hat{V}_t := f_{\theta_1}(V_0) \times f'_{\theta_2}(\Delta L_{\tau_1}) \times ... \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}),$$

$V_t - \hat{V}_t$ is decreasing in V_0, and therefore $Cov(V_0, V_t 1_{\{i(t)\geq 1\}}) \leq Cov(V_0, \hat{V}_t 1_{\{i(t)\geq 1\}})$, Conditioning w.r.t. τ_1, and using independence

$$\begin{aligned}
&Cov(V_0, \hat{V}_t 1_{\{i(t)\geq 1\}}) \\
&\quad = \mathbb{E}[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0) \times f'_{\theta_2}(\Delta L_{\tau_1}) \times ... \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)\geq 1\}}] \\
&\quad = \mathbb{E}\big[\mathbb{E}[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0) \times f'_{\theta_2}(\Delta L_{\tau_1}) \times ... \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)\geq 1\}} | \tau_1]\big] \\
&\quad = \mathbb{E}\left[\mathbb{E}[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0) 1_{\{i(t)\geq 1\}} | \tau_1] \times \mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) | \tau_1\right]\right],
\end{aligned}$$

where we agree that the product over k is 1 if $i(t) = 1$. Now we decompose:

$$\begin{aligned}
&\mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) | \tau_1\right] \\
&\quad = \mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)\geq \rho\lambda t\}} | \tau_1\right] \\
&\quad + \mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)< \rho\lambda t\}} | \tau_1\right]
\end{aligned}$$

Being $f'_t \leq 1$, from a standard result on tails of Poisson processes, the second summand admits the following upper bound:

$$\begin{aligned}
&\mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)< \rho\lambda t\}} | \tau_1\right] \\
&\quad \leq \mathbb{P}\left(i(t) < \rho\lambda t | \tau_1\right) \\
&\quad \leq \mathbb{P}\left(Po(\lambda t) < \rho\lambda t - \tau_1 + 1 | \tau_1\right) \\
&\quad \leq \mathbb{P}\left(Po(\lambda t) < \rho\lambda t + 1\right) \sim_{t\sim\infty} \mathbb{P}\left(Po(\lambda t) < \rho\lambda t\right) \\
&\quad \leq e^{-\lambda t}\left(\frac{e}{\rho}\right)^{\rho\lambda t} \\
&\quad \leq \exp(-\alpha_\rho \lambda t).
\end{aligned}$$

For the first summand,

$$\begin{aligned}
&\mathbb{E}\left[\prod_{k=2}^{i(t)} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) \times f'_{t-\tau_{i(t)}}(\Delta L_{\tau_{i(t)}}) 1_{\{i(t)\geq \rho\lambda t\}} | \tau_1\right] \\
&\quad \leq \mathbb{E}\left[\prod_{k=2}^{[\rho\lambda t]+1} f'_{\theta_k}(\Delta L_{\tau_{k-1}}) | \tau_1\right] \leq \mathbb{E}\left[f'_\theta(\Delta L)\right]^{[\rho\lambda t]} \\
&\quad = \mathbb{E}\left[(1 + (\gamma - 1) C\, \theta\, \Delta L^{\gamma-1})^{\frac{\gamma}{1-\gamma}}\right]^{[\rho\lambda t]}
\end{aligned}$$

where $\theta \sim \exp(\lambda)$ and $\Delta L \sim \nu$. This part decays exponentially as well,

$$\sim \exp\left(-\beta_\rho \lambda t\right).$$

Remark that the upper bounds we have found are deterministic and we can take them out of the expectation. Now,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0) 1_{\{i(t)\geq 1\}} | \tau_1]\right] \\
&= \mathbb{E}\left[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0) 1_{\{i(t)\geq 1\}}\right] \\
&\to \mathbb{E}\left[(V_0 - \mathbb{E}V_0) f_{\theta_1}(V_0)\right]
\end{aligned}$$

for $t \to \infty$, for bounded convergence. Therefore this part does not give any exponential contribution. □

Remark 6.14. On the leverage effect: We have considered the multiscaling phenomenon for a stochastic volatility model

$$dX_t = \sigma_t dW_t$$

where the volatility process σ is independent of W. We wonder if it is possible to add a term that explains the *leverage effect* of financial markets, but does not change the multiscaling phenomenon. A first attempt could be to take a Lévy process J correlated to σ_t, and define

$$dX_t = \sigma_t dW_t - dJ_t.$$

In the specific example of OU with superlinear drift this might be something stricly connected to the process of jumps of the volatility, e.g. $J = const\, L$, or J with the same jump times of L but with jumps of size one (the standard Poisson process associated to L). This would give

$$X_h = \int_0^h \sigma_s dW_s - (J_h - J_0),$$

and when computing the moments we have to deal with

$$\mathbb{E}|J_h - J_0|^q.$$

This depends on the precise law of the process J, but in general scales as h, independently of q, and therefore it changes the multiscaling behavior introducing a term with a larger scaling, for $q > 2$, and producing a scaling exponent $A(q) = \frac{q}{2} \wedge 1$. Some more informations on this issue can be found in [39]

Bibliography

[1] A. Andreoli, F. Caravenna, P. Dai Pra, and G. Posta. Scaling and multiscaling in financial series: a simple model. *Advances in Applied Probability*, 44(4):1018–1051, 2012.

[2] D.G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, 73:890–896, 1967.

[3] D. Bakry, F. Baudoin, M. Bonnefont, and D. Chafai. On gradient bounds for the heat kernel on the heisenberg group. *Journal of Functional Analysis*, 255(8):1905 – 1938, 2008.

[4] F. Baldovin and A. Stella. Scaling and effciency determine the irreversible evolution of a market. *PNAS*, 104(50):19741–19744, 2007.

[5] V. Bally and L. Caramellino. Riesz transform and integration by parts formulas for random variables. *Stochastic Processes and their Applications*, 121(6):1332 – 1355, 2011.

[6] V. Bally and L. Caramellino. Tubes estimates for diffusion processes under a local Hörmander condition of order one. *ArXiv e-prints*, February 2012.

[7] V. Bally and L. Caramellino. Positivity and lower bounds for the density of Wiener functionals. *Potential Analysis*, 39:141–168, April 2013.

[8] V. Bally and L. Caramellino. On the distances between probability density functions. *Electron. J. Probab.*, 19:no. 110, 1–33, 2014.

[9] V. Bally and S. De Marco. Some estimates in extended stochastic volatility models of heston type. *Risk and Decision Analysis*, 2(4):195–206, 2011.

[10] V. Bally, B. Fernandez, and A. Meda. Estimates for the probability that a itô process remains near a path. *Stochastic Processes and their Applications*, 121:2087–2113, 2011.

[11] V. Bally and A. Kohatsu-Higa. Lower bounds for densities of Asian type stochastic differential equations. *J. Funct. Anal.*, 258(9):3134–3164, 2010.

[12] O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society Series B*, 63(2):167–241, 2001.

[13] 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.

[14] F. Baudoin. *An introduction to the Geometry of stochastic flows*. Imperial College Press, 2004.

[15] F. Baudoin and M. Hairer. A version of hörmander's theorem for the fractional brownian motion. *Prob. Theory and Rel. Fields*, 139 (3-4):373–395, 2007.

[16] G. Ben Arous, M. Gradinaru, and M. Ledoux. Hölder norms and the support theorem for diffusions. *Annales de l'IHP - Probabilités et Statistiques*, 30:415–436, 1994.

[17] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. I. *Probab. Theory Related Fields*, 90(2):175–202, 1991.

[18] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. II. *Probab. Theory Related Fields*, 90(3):377–402, 1991.

[19] J. Bertoin. *Lévy processes*, volume 121. Cambridge university press, 1998.

[20] K. Bichteler, J-B. Gravereaux, and J. Jacod. *Malliavin Calculus for Processes with Jumps*, volume Stochastic mongraphs Volume 2. Gordon and Breach Science Publishers, 1987.

[21] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[22] M. Bonino, M. Camelia, and P. Pigato. A multivariate model for financial indexes and an algorithm for detection of jumps in the volatility. *arXiv preprint, arXiv:1404.7632*, 2014.

[23] Bonino, M. Portfolio allocation and monitoring under volatility shocks. Master's thesis, Università degli Studi di Padova, 2011.

[24] U. Boscain and S. Polidoro. Gaussian estimates for hypoelliptic operators via optimal control. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 18(4):333–342, 2007.

[25] C. T. Brownlees and R. F. Engle. Volatility, correlation and tails for systemic risk measurement. *Available at SSRN 1611229*, 2012.

[26] L.E. Calvet, A. J. Fisher, and B. B. Mandelbrot. A multifractal model of asset returns. *Cowles Foundation Discussion Papers*, pages 1164–1166, 1997.

[27] L.E. Calvet and A.J. Fisher. Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58, 2001.

[28] L.E. Calvet and A.J. Fisher. *Multifractal Volatility.* Academic Press, 2008.

[29] Camelia, M. Variazione quadratica di processi stocastici: un'applicazione alla finanza. *Undergraduate thesis.*, page Università degli Studi di Padova, 2012.

[30] M. Capitaine. On the onsager-machlup functional for elliptic diffusion processes. In *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 313–328. Springer Berlin Heidelberg, 2000.

[31] F. Caravenna and J. Corbetta. General smile asymptotics with bounded maturity. *arXiv preprint, arXiv:1411.1624*, 2014.

[32] F. Caravenna and J. Corbetta. The asymptotic smile of a multiscaling stochastic volatility model. *arXiv preprint, arXiv:1501.03387*, 2015.

[33] T. Cass and P. Friz. Densities for rough differential equations under hörmander's condition. *Annals of Mathematics*, 171(3):2115–2141, 2010.

[34] P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions. *Prob. Theory and Rel. Fields*, 123:453–483, 2002.

[35] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41:135–155, 1973.

[36] F. Collet and P. Dai Pra. The role of disorder in the dynamics of critical fluctuations of mean field models. *Electron. J. Probab.*, 17, 2012.

[37] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[38] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall, Boca Raton, 2004.

[39] J. Corbetta. *General smile asymptotics and a multiscaling stochastic volatility model.* PhD thesis, Università degli Studi di Milano-Bicocca, 2015.

[40] J. M. Coron. *Control and nonlinearity*, volume 136 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2007.

[41] P. Dai Pra and P. Pigato. Multi-scaling of moments in stochastic volatility models. *preprint*, 2014.

[42] F. Delarue and S. Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *J. Funct. Anal.*, 259(6):1577–1630, 2010.

[43] T. Di Matteo. Multi-scaling in finance. *Quantitative Finance*, 7(1), 2007.

[44] T. Di Matteo, T. Aste, and M. M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *J. Banking Finance*, 29:827–851, 2005.

[45] Z. Ding, C. W. J. Granger, and R. F. Engle. A long-memory property of stock market returns and a new model. *Journal of Empirical Finance*, pages 83–106, 1993.

[46] C. Donati-Martin and M. Yor. Fubini's theorem for double wiener integrals and the variance of the brownian path. *Ann. Inst. Henri Poincaré*, 27:181–200, 1991.

[47] B. K. Driver and T. Melcher. Hypoelliptic heat kernel inequalities on the heisenberg group. *Journal of Functional Analysis*, 221(2):340 – 365, 2005.

[48] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pages 176–223, 2002.

[49] R. Eon and M. Gradinaru. Gaussian asymptotics for a non-linear langevin type equation driven by an a-stable l´evy noise. *hal-01107765*, 2015.

[50] V. Fasen, C. Klüppelberg, and A. Lindner. Extremal behavior of stochastic volatility models. In *Stochastic finance*, pages 107–155. Springer, 2006.

[51] V. Fasen, C. Klüppelberg, and A. Lindner. Extremal behavior of stochastic volatility models. *In Stochastic finance*, pages 107–155, 2006.

[52] P. Foschi, S. Pagliarani, and A. Pascucci. Black-scholes formulae for asian options in local volatility models. *J. Comput. Appl. Math.*, 237:442–459, 2013.

[53] P. Friz, T. Lyons, and D. Stroock. Lévy's area under conditioning. *Annales de l'IHP - Probabilités et Statistiques*, 42:89–101, 2006.

[54] S. Galluccio, G. Caldarelli, M. Marsili, and Y. C. Zhang. Scaling in currency exchange. *Physica A*, 245:423–36, 1997.

[55] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *ArXiv e-prints*, October 2014.

[56] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381:767–70, 1996.

[57] S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62, September 1996.

[58] A. Guillin. Averaging principle of sde with small diffusion: Moderate deviations. *Ann. Probab.*, 31(1):413–443, 01 2003.

[59] M. Hairer and J. C. Mattingly. Slow energy dissipation in anharmonic oscillator chains. *Comm. Pure Appl. Math.*, 62(8):999–1032, 2009.

[60] Z. He and J. M. Maheu. Real time detection of structural breaks in garch models. *Computational Statistics & Data Analysis*, 54(11):2628–2640, 2010.

[61] R. Höpfner, E. Löcherbach, and M. Thieullen. Ergodicity for a stochastic hodgkin-huxley model driven by ornstein-uhlenbeck type input. *ArXiv e-prints*, 2014.

[62] R. Höpfner, E. Löcherbach, and M. Thieullen. Strongly degenerate time inhomogeneous sdes: densities and support properties. application to a hodgkin-huxley system with periodic input. *ArXiv e-prints*, 2014.

[63] L. Huang and S. Menozzi. A Parametrix Approach for some Degenerate Stable Driven SDEs. *ArXiv e-prints*, February 2014.

[64] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1989.

[65] C. Inclán and G. C. Tiao. Use of cumulative sums of squares for retrospective detection in the changes fo variance. *J. Amer. Statist. Assoc.*, 89:913–923, 1994.

[66] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, 2002.

[67] C. Klüppelberg, A. Lindner, and R. Maller. Continuous time volatility modelling: Cogarch versus ornstein–uhlenbeck models. In *From stochastic calculus to mathematical finance*, pages 393–419. Springer, 2006.

[68] C. Kluppelberg, A. Lindner, and R. A. Maller. Continuous time volatility modelling: Cogarch versus ornstein-uhlenbeck models. In From Stochastic Calculus to Mathematical Finance, Yu. Kabanov, R. Lipster and J. Stoyanov (Eds.). Springer, 2007.

[69] A. Kohatsu-Higa and M. Yamazato. On moments and tail behaviors of storage processes. *Journal of Applied Probability*, 40(4):1069–1086, 2003.

[70] S. Kusuoka and D. Stroock. Application of the malliavin calculus, part i. *Stochastic analysis (Katata/Kyoto, 1982), North-Holland*, pages 271–306, 1984.

[71] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. II. *Journal of the Faculty of Science of the Univ. of Tokyo Sect IA Math*, pages 1–76, 1985.

[72] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 34(2):391–442, 1987.

[73] H.-Q. Li. Estimations asymptotiques du noyau de la chaleur sur les groupes de heisenberg. *Comptes Rendus Mathematique*, 344(8):497 – 502, 2007.

[74] P. Malliavin. Stochastic calculus of variation and hypoelliptic operators. *Proceedings of the International Symposium on SDEs, Kyoto 1976*, pages 195–263, 1978.

[75] P. Malliavin and A. Thalmaier. *Stochastic Calculus of Variations in Mathematical Finance*. Springer-Verlag, Berlin, 2006.

[76] MATLAB. *version 7.8.0 (R2009a)*. The MathWorks Inc., Natick, Massachusetts, 2009.

[77] J. C. Mattingly, T. Suidan, and E. Vanden-Eijnden. Simple systems with anomalous dissipation and energy cascade. *Comm. Math. Phys.*, 276(1):189–220, 2007.

[78] A. Nagel, E. M. Stein, and S. Wainger. Balls and metrics defined by vector fields. I. Basic properties. *Acta Math.*, 155(1-2):103–147, 1985.

[79] D. Nualart. *Malliavin Calculus and Related Topics*. Springer, Berlin, 2006.

[80] A. Pascucci and S. Polidoro. Harnack inequalities and Gaussian estimates for a class of hypoelliptic operators. *Trans. Amer. Math. Soc.*, 358(11):4873–4893 (electronic), 2006.

[81] P. Pigato. Tube estimates for diffusion processes under a weak hormander condition. *Preprint*, 2014.

[82] Pigato, P. A multivariate model for financial indexes subject to volatility shocks. Master's thesis, Università degli Studi di Padova, 2011.

[83] B. Podobnik, D. F. Fu, H. E. Stanley, and P. Ch. Ivanov. Power-law autocorrelated stochastic processes with long-range cross-correlations. *The European Physical Journal B - Condensed Matter and Complex Systems*, 56(1):47–52, 2007.

[84] B. Podobnik, D. Wang, D. Horvatic, I. Grosse, and H. E. Stanley. Time-lag cross-correlations in collective phenomena. *EPL (Europhysics Letters)*, 90(6):68001, 2010.

[85] S. Polidoro. A global lower bound for the fundamental solution of Kolmogorov-Fokker-Planck equations. *Arch. Rational Mech. Anal.*, 137(4):321–340, 1997.

[86] P. E. Protter. *Stochastic Integration and Differential Equations, Version 2.1. Vol. 21*. Springer, 2004.

[87] G. J. Ross. Modelling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and its Applications*, 392(2):350 – 360, 2013.

[88] W. Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

[89] G. Samorodnitsky and M. Grigoriu. Tails of solutions of certain nonlinear stochastic differential equations driven by heavy tailed levy motions. *Stochastic processes and their applications*, 105.1:69–97, 2003.

[90] M. Sanz Solé. *Malliavin calculus : with applications to stochastic partial differential equations.* Fundamental sciences. EPFL press [Paris], Lausanne (Suisse), 2005.

[91] K. Sato. *Lévy processes and infinitely divisible distributions.* Cambridge university press, 1999.

[92] I. Shigekawa. *Stochastic analysis*, volume 224 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI, 2004.

[93] A. Stella and F. Baldovin. Role of scaling in the statistical modeling of finance. *Pramana*, 71:341–352, 2008.

[94] D. Stroock and S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proc. Sixth Berkeley Symp. Math. Statist.*, volume Prob. III, pages 333–359, Berkeley, 1972. Univ. California Press.

[95] J.C. Vassilicos, A. Demos, and F. Tata. No evidence of chaos but some evidence of multifractals in the foreign exchange and the stock market. In *Applications of Fractals and Chaos, eds. A.J. Crilly, R.A. Earnshaw, H. Jones*, pages 249–65. Springer, 1993.

[96] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[97] D. Wang, B. Podobnik, D. Horvatić, and H. E. Stanley. Quantifying and modeling long-range cross-correlations in multiple time series with applications to world stock indices. *Phys. Rev. E*, 83(046121), 2011.

[98] D. Williams. To begin at the beginning. In *Stochastic Integrals*, volume 851 of *Lecture Notes in Mathematics*, pages 1–55. Springer, 1981.

[99] S.J. Wolfe. On moments of infinitely divisible distribution functions. *The Annals of Mathematical Statistics*, 42(6):2036–2043, 1971.