Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2019, Vol. 55, No. 4, 2320–2369
https://doi.org/10.1214/18-AIHP950

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Tube estimates for diffusions under a local strong Hörmander condition

Vlad Bally[a], Lucia Caramellino[b] and Paolo Pigato[c]

[a]Université Paris-Est, LAMA (UMR CNRS, UPEMLV, UPEC), MathRisk INRIA, F-77454 Marne-la-Vallée, France. E-mail: bally@univ-mlv.fr
[b]Dipartimento di Matematica, Università di Roma - Tor Vergata and INdAM-GNAMPA, Via della Ricerca Scientifica 1, I-00133 Roma, Italy. E-mail: caramell@mat.uniroma2.it
[c]Weierstrass Institute, Mohrenstr. 39, 10117 Berlin, Germany. E-mail: paolo.pigato@wias-berlin.de

Received 11 October 2016; revised 12 October 2018; accepted 12 November 2018

Abstract. We study lower and upper bounds for the probability that a diffusion process in $\mathbb{R}^n$ remains in a tube around a deterministic skeleton path up to a fixed time. The diffusion coefficients $\sigma_1, \dots, \sigma_d$ may degenerate, but we assume that they satisfy a strong Hörmander condition involving the first order Lie brackets around the skeleton of interest. The tube is written in terms of a norm which accounts for the non-isotropic structure of the problem: in a small time δ, the diffusion process propagates with speed $\sqrt{\delta}$ in the direction of the diffusion vector fields σ_j and with speed δ in the direction of $[\sigma_i, \sigma_j]$. We first prove short-time (non-asymptotic) lower and upper bounds for the density of the diffusion. Then, we prove the tube estimate using a concatenation of these short-time density estimates.

Résumé. On étudie des bornes inférieures et supérieures pour la probabilité qu'un processus de diffusion dans R^n reste dans un petit tube autour d'un squelette déterministe jusqu'à un temps fixé. Les coefficients de diffusion $\sigma_1, \dots, \sigma_d$ peuvent dégénérer, mais on suppose qu'ils satisfont à une condition d'Hörmander forte sur les coefficients et leurs crochets de Lie de premier ordre autour du squelette d'intérêt. Le tube est écrit en termes d'une norme qui prend en compte la structure non isotrope du problème: en temps δ petit, le processus de diffusion se propage avec vitesse $\sqrt{\delta}$ dans la direction des vecteurs de diffusion σ_j et avec vitesse δ dans la direction de $[\sigma_i, \sigma_j]$. On prouve d'abord des bornes inférieures et supérieures en temps court (non asymptotiques) pour la densité de la diffusion. Ensuite, on prouve l'estimée de tube en utilisant une concaténation de ces estimées de densité en temps court.

MSC: Primary 60H07; 60H10; secondary 60H30

Keywords: Tube estimates; Short-time density estimates; Hypoellipticity; Strong Hörmander condition; Malliavin calculus

1. Introduction

We consider a diffusion process in $\mathbb{R}^n$ solution to

$$dX_t = \sum_{j=1}^{d} \sigma_j(t, X_t) \circ dW_t^j + b(t, X_t)\, dt, \qquad X_0 = x_0, \tag{1.1}$$

where $W = (W^1, \dots, W^d)$ is a standard Brownian motion and $\circ dW_t^j$ denotes the Stratonovich integral. We assume suitable regularity properties for $\sigma_j, b : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ and that the coefficients σ_j, b verify the strong Hörmander condition of order one (that is, involving the σ_j's and their first order Lie brackets $[\sigma_i, \sigma_j]$'s), locally around a skeleton path defined by

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

where ϕ is a deterministic control. In such framework, we find exponential lower and upper bounds for the probability that the diffusion X remains in a small tube around the skeleton path $x(\phi)$. We stress that all the regularity and non degeneracy assumptions that we need are local, that is, for (t, x) belonging to a tube around the path $x_t(\phi)$ (see next (2.11) and (2.12)).

The present article was originally written as two separate preprints [4,5].

Several works have considered this subject, starting from Stroock and Varadhan in [30], where such result is used to prove the support theorem for diffusion processes. In their work, the tube is written in terms of the Euclidean norm, but later on different norms have been used to take into account the regularity of the trajectories [10,19] and their geometric structure [27]. This kind of problems is also related to the Onsager-Machlup functional and large or moderate deviation theory, see e.g. [14,20,21].

In this work, we construct the tube using a norm which reflects the non-isotropic structure of the problem: roughly speaking, in a small time interval of length δ, the diffusion process moves with speed $\sqrt{\delta}$ in the direction of the diffusion vector fields σ_j and with speed $\delta = \sqrt{\delta} \times \sqrt{\delta}$ in the direction of $[\sigma_i, \sigma_j]$. In order to capture the local geometric structure around (t, x), we construct a matrix $A_\delta(t, x)$ based on $\sqrt{\delta}\sigma_i(t, x)$ and $[\sqrt{\delta}\sigma_i, \sqrt{\delta}\sigma_j](t, x)$, $i, j = 1, \dots, d$ (see next definition (2.8)), and use it to define a norm $|\cdot|_{A_\delta(t,x)}$.

We also consider, as in [23], the time-homogeneous case $b(t, x) = b(x)$ and $\sigma(t, x) = \sigma(x)$. In this case, we define a semi-distance d via: $d(x, y) < \sqrt{\delta}$ if and only if $|x - y|_{A_\delta(x)} < 1$. We prove the local equivalence of d, the semi-distance associated with the the matrix norm $|\cdot|_{A_\delta}$, with the Carathéodory distance d_c. This gives a rewriting of the tube estimates in terms of the control distance as well.

A key step in proving our tube estimates is the proof of lower and upper bounds for the density of the solution to (1.1) at a small (but not asymptotic) time, say δ, under the strong Hörmander condition of order one.

Let us give a hint on the main idea of our approach. Usually, when studying the short time behavior of the density of a diffusion process, one employs a stochastic Taylor development of order one in order to isolate a principal term, which is a Gaussian random variable of covariance matrix $\sigma\sigma^T(x)$. In the elliptic case, this matrix is non degenerated and then, in short time, the density of X_t behaves as the density of $\sigma\sigma^T(x)\Delta$, where Δ is a standard normal random variable. On the contrary, if we just assume the Hörmander's condition, $\sigma_1(x), \dots, \sigma_n(x)$ do not span the whole space. In this case we have to involve $[\sigma_i, \sigma_j](x)$, $1 \le i < j \le n$ as well. This is a non trivial problem and we deal with it using a decomposition which leads to a Gaussian random variable with a stochastic covariance matrix. In order to prove that this matrix is non degenerated, we use a result due to Donati-Martin and Yor on the variance of the Brownian path [17]. In this context, the Brownian trajectory $t \to W_t$ appears as a control, and the fact that this trajectory has sufficiently large variance gives the non degeneracy of our covariance matrix. The argument presented here seems to be new, when compared to classical density estimates in hypoelliptic setting. In the context of a degenerate diffusion coefficient which fulfills a strong Hörmander condition, the main result in this direction is due to Kusuoka and Stroock. In the celebrated paper [23], they prove a two-sided Gaussian bound for the density in the control (Carathéodory) distance, assuming that the coefficients do not depend on the time variable and that the drift is generated by the vector fields of the diffusive part, which is a quite restrictive hypothesis. Other notable estimates for the heat kernel under strong Hörmander conditions are provided in [11,12]. The subject has also been widely studied with analytical methods - see for example [22,29]. We stress that these are asymptotic results, whereas we prove estimates for a finite, positive and fixed time. Non-isotropic norms similar to the one used here have been used in [16,27] to provide density estimates for SDEs under Hörmander conditions of weak type. We also refer to [15], which considers the existence of the density for SDEs with time dependent coefficients, under very weak regularity assumptions.

In the present paper, we obtain the tube estimate from a concatenation of short-time density estimates, the fact that our density estimates are not asymptotic being crucial. Examples of application of tube estimates can be found in [7], where the authors prove tube estimates for locally elliptic diffusions; these estimates are then applied to find lower bounds for the probability to be in a ball at fixed time and bounds for the distribution function. In [6], tube estimates are applied to local-stochastic volatility models, finding estimates for the tails of the log-price distribution and estimates on the implied volatility.

The paper is organized as follows. In Section 2 we set-up the framework and give the precise statement of our main results (Theorem 2.4 and 2.9). The lower bound for the density is proved in Section 3, the upper bound in Section 4. The proof of the tube estimate is developed in Section 5. In Appendix A we study the local equivalence between the control metric and our matrix norm. Some technical issues are postponed to the other appendices of the paper.

2. Notation and main results

2.1. *Notations*

Let W be a standard Brownian motion in $\mathbb{R}^d$ and let X denote the process in $\mathbb{R}^n$ already introduced in (1.1), that is

$$dX_t = \sum_{j=1}^{d} \sigma_j(t, X_t) \circ dW_t^j + b(t, X_t)\,dt, \qquad X_0 = x_0. \tag{2.1}$$

Here, the vector fields $\sigma_j, b : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ are four time differentiable in $x \in \mathbb{R}^n$ and one time differentiable in $t \in \mathbb{R}^+$, and we suppose that the derivatives with respect to the space variable $x \in \mathbb{R}^n$ are one time differentiable with respect to the time variable t.

Hereafter, for $k \geq 1$, $\alpha = (\alpha_1, \dots, \alpha_k) \in \{1, \dots, n\}^k$ represents a multi-index with length $|\alpha| = k$ and $\partial_x^\alpha = \partial_{x_{\alpha_1}} \cdots \partial_{x_{\alpha_k}}$. We allow the case $k = 0$ by setting $\alpha = \varnothing$ (the void multi-index), $|\alpha| = 0$ and $\partial_x^\alpha = \mathrm{Id}$.

For $f, g : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ we define the directional derivative (w.r.t. the space variable x) $\partial_g f(t,x) = \sum_{i=1}^n g^i(t,x)\partial_{x_i} f(t,x)$, and we recall that the Lie bracket (again w.r.t. the space variable) 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) singular value of M. We recall that singular values are the square roots of the eigenvalues of MM^T, and that, when M is symmetric, singular values coincide with the absolute values of the eigenvalues of M. In particular, when M is a covariance matrix, $\lambda_*(M)$ and $\lambda^*(M)$ coincide with the smallest and the largest eigenvalues of M. We consider the following norm on $\mathbb{R}^n$:

$$|y|_M = \sqrt{\big\langle \big(MM^T\big)^{-1} y, y\big\rangle}. \tag{2.2}$$

We introduce the $n \times d^2$ matrix $A(t,x)$ defined as follows. We set $m = d^2$ and define the function

$$l(i,p) = (p-1)d + i \in \{1, \dots, m\}, \qquad p, i \in \{1, \dots, d\}. \tag{2.3}$$

Notice that $l(i,p)$ is invertible. For $l = 1, \dots, m$, we set the (column) vector field $A_l(t,x)$ in $\mathbb{R}^n$ as follows:

$$\begin{aligned} A_l(t,x) &= [\sigma_i, \sigma_p](t,x) \quad \text{if } l = l(i,p) \text{ with } i \neq p, \\ &= \sigma_i(t,x) \quad \text{if } l = l(i,p) \text{ with } i = p \end{aligned} \tag{2.4}$$

and we set the $n \times m$ matrix $A(t,x)$ to be the one having $A_1(t,x), \dots, A_m(t,x)$ as its columns, that is

$$A(t,x) = \big[A_1(t,x), \dots, A_m(t,x)\big]. \tag{2.5}$$

We denote by $\lambda(t,x)$ the smallest singular value of $A(t,x)$, so

$$\lambda(t,x)^2 = \lambda_*\big(A(t,x)\big)^2 = \inf_{|\xi|=1} \sum_{i=1}^m \big\langle A_i(t,x), \xi\big\rangle^2. \tag{2.6}$$

For fixed $R > 0$ we define the $m \times m$ diagonal scaling matrix $\mathcal{D}_R$ as

$$\begin{aligned} (\mathcal{D}_R)_{l,l} &= R \quad \text{if } l = l(i,p) \text{ with } i \neq p, \\ &= \sqrt{R} \quad \text{if } l = l(i,p) \text{ with } i = p \end{aligned} \tag{2.7}$$

and the scaled directional matrix

$$A_R(t,x) = A(t,x)\mathcal{D}_R. \tag{2.8}$$

Notice that the lth column of the matrix $A_R(t,x)$ is given by $\sqrt{R}\sigma_i(t,x)$ if $l = l(i,p)$ with $i = p$, and if $i \neq p$ then the lth column of $A_R(t,x)$ is $R[\sigma_i, \sigma_p](t,x) = [\sqrt{R}\sigma_i, \sqrt{R}\sigma_p](t,x)$.

This matrix and the associated norm $|\cdot|_{A_R(\cdot,\cdot)}$ in (2.2) are the tools that allow us to account of the different speeds of propagation of the diffusion: $\sqrt{R}$ (diffusive scaling) in the direction of σ and R in the direction of the first order Lie brackets. In particular, straightforward computations easily give that

$$\frac{1}{\sqrt{R}\lambda^*(A(t,x))}|y| \le |y|_{A_R(t,x)} \le \frac{1}{R\lambda_*(A(t,x))}|y|. \tag{2.9}$$

2.2. *Short-time density estimates*

We suppose that the diffusion coefficients fulfill the following requests:

Assumption 2.1. *There exists a constant $\kappa > 0$ such that*, $\forall t \in [0,1], \forall x \in \mathbb{R}^n$:

$$\sum_{j=1}^d \left|\sigma_j(t,x)\right| + \left|b(t,x)\right| + \sum_{j=1}^d \sum_{0\le|\alpha|\le 2} \left|\partial_x^\alpha \partial_t \sigma_j(t,x)\right| \le \kappa\big(1+|x|\big),$$

$$\sum_{j=1}^d \sum_{1\le|\alpha|\le 4} \left|\partial_x^\alpha \sigma_j(t,x)\right| + \sum_{1\le|\alpha|\le 3} \left|\partial_x^\alpha b(t,x)\right| \le \kappa.$$

Remark that Assumption 2.1 ensures the strong existence and uniqueness of the solution to (2.1). We do not assume here ellipticity but a non degeneracy of Hörmander type:

Assumption 2.2. *Let x_0 denote the starting point of the diffusion X solving* (2.1). *We suppose that the smallest eigenvalues of A at the initial condition satisfies*

$$\lambda(0,x_0) > 0.$$

Notice that Assumption 2.2 is actually equivalent to requiring that the first order Hörmander condition holds at the starting point x_0, i.e. the vector fields $\sigma_i(0,x_0)$, $[\sigma_j,\sigma_p](0,x_0)$, as $i,j,p = 1,\dots,d$, span the whole $\mathbb{R}^n$. In particular, it is known that under this assumption the law of X_t is absolutely continuous w.r.t. the Lebesgue measure, see [21,26].

We also consider the following assumption, as a stronger version of Assumption 2.1 (morally we ask for boundedness instead of sublinearity of the coefficients, in the spirit of Kusuoka-Stroock estimates in [23]).

Assumption 2.3. *There exists a constant $\kappa > 0$ such that for every $t \in [0,1]$ and $x \in \mathbb{R}^n$ one has*

$$\sum_{0\le|\alpha|\le 4}\left[\left|\partial_x^\alpha b(t,x)\right| + \sum_{j=1}^d \left|\partial_x^\alpha \sigma_j(t,x)\right| + \left|\partial_x^\alpha \partial_t \sigma_j(t,x)\right|\right] \le \kappa.$$

The first result of this paper is the following:

Theorem 2.4. *Let Assumption* 2.1 *and* 2.2 *hold. Let p_{X_t} denote the density of X_t, $t > 0$, with starting condition $X_0 = x_0$. Then the following statements hold.*

(1) *There exist positive constants r, δ_*, C, depending on $\lambda(0,x_0)$ and κ, such that for every $\delta \le \delta_*$ and for every y such that $|y - x_0 - b(0,x_0)\delta|_{A_\delta(0,x_0)} \le r$,*

$$\frac{1}{C\sqrt{\det A_\delta A_\delta^T(0,x_0)}} \le p_{X_\delta}(y).$$

(2) *For any $p > 1$, there exists a positive constant C, depending on $\lambda(0,x_0)$ and κ, such that for every $\delta \le 1$ and for every $y \in \mathbb{R}^n$,*

$$p_{X_\delta}(y) \le \frac{1}{\sqrt{\det A_\delta A_\delta^T(0,x_0)}} \times \frac{C}{1 + |y - x_0|^p_{A_\delta(0,x_0)}}.$$

(3) *If also Assumption 2.3 holds (boundedness of coefficients) there exists a constant C, depending on $\lambda(0,x_0)$ and κ, such that for every $\delta \le 1$ and for every $y \in \mathbb{R}^n$,*

$$p_{X_\delta}(y) \le \frac{C}{\sqrt{\det A_\delta A_\delta^T(0,x_0)}} \exp\left(-\frac{1}{C}|y - x_0|_{A_\delta(0,x_0)}\right).$$

Remark 2.5. It might appear contradictory that the lower estimate (1) in Theorem 2.4 is centered in $x_0 + \delta b(0,x_0)$, whereas the upper estimates are centered in x_0. In fact, this is important only for the lower bound, the upper bounds (2) and (3) holding either if we write $|y - x_0 - \delta b(0,x_0)|_{A_\delta(0,x_0)}$ or $|y - x_0|_{A_\delta(0,x_0)}$ (see next Remark 4.5). When proving the tube estimate we are mostly interested in the *diagonal* density estimates, meaning the local estimate around the drifted initial condition $x_0 + b(0,x_0)\delta$.

For the application to the tube estimate it is also crucial that our results are not asymptotic, but hold uniformly for δ small enough. Also notice that the upper bounds in (2) and (3) of Theorem 2.4 give the tail estimates, which are exponential if we assume the boundedness of the coefficients, polynomial otherwise.

Remark 2.6. A global two-sided bound for the density of X_t is proved in [23], under the *strong* Hörmander non-degeneracy condition. It is also assumed that the coefficients do not depend on 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). In the present paper, on the contrary, we allow for a general drift and time dependence in the coefficients, but we consider only first order Lie brackets. Moreover, in Assumption 2.1, we also relax the hypothesis of bounded coefficients. Anyway, the two estimates are strictly related, since our matrix norm is locally equivalent to the Carathéodory control metric, as we discuss in Appendix A.

2.3. *Tube estimates*

For a control $\phi \in L^2([0,T],\mathbb{R}^n)$ we consider the skeleton $x(\phi)$ associated to (2.1), that is,

$$dx_t(\phi) = \sum_{j=1}^d \sigma_j\big(t, x_t(\phi)\big)\phi_t^j\,dt + b\big(t, x_t(\phi)\big)\,dt, \qquad x_0(\phi) = x_0. \tag{2.10}$$

In the following, we use a function $R : [0,T] \to (0,1]$ that plays the role of radius function for the tube around $x(\phi)$.

For $(t,x) \in \mathbb{R}^+ \times \mathbb{R}^n$ we denote by $n(t,x)$ a constant such that

$$\forall s \in \big[(t-1)\vee 0, t+1\big], \forall y \in B(x,1) \text{ one has}$$
$$\sum_{|\alpha|=0}^4 \left(\big|\partial_x^\alpha b(s,y)\big| + \big|\partial_t\partial_x^\alpha b(s,y)\big| + \sum_{j=1}^d \big|\partial_x^\alpha \sigma_j(s,y)\big| + \big|\partial_t\partial_x^\alpha \sigma_j(s,y)\big|\right) \le n(t,x). \tag{2.11}$$

We consider now a "regularity property" already introduced in [8], which is needed to control the growth of certain quantities along the skeleton path. For $\mu \geq 1$ and $0 < h \leq 1$ we denote by $L(\mu, h)$ the following class of functions:

$$L(\mu, h) = \{ f : \mathbb{R}^+ \to \mathbb{R}^+ \text{ such that } f(t) \leq \mu f(s) \text{ for } |t-s| \leq h \}.$$

To prove the tube estimate, we make use of the following hypotheses: there exist 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

$$\begin{aligned}
&(H_1) \qquad n\big(t, x_t(\phi)\big) \leq n_t, \quad \forall t \in [0, T], \\
&(H_2) \qquad \lambda\big(t, x_t(\phi)\big) \geq \lambda_t, \quad \forall t \in [0, T], \\
&(H_3) \qquad R_\cdot, |\phi_\cdot|^2, n_\cdot, \lambda_\cdot \in L(\mu, h).
\end{aligned} \tag{2.12}$$

Recall that $\phi \in L^2([0, T], \mathbb{R}^n)$ is the control giving the skeleton path and $R : [0, T] \to (0, 1]$ stands for the radius function. Notice that the functions n_t and λ_t depend on the control ϕ through (2.12).

Remark 2.7. Hypothesis (H_2) implies that for each $t \in (0, T)$, the space $\mathbb{R}^n$ is spanned by the vectors $(\sigma_i(t, x_t(\phi)), [\sigma_j, \sigma_p](t, x_t(\phi)))_{i,j,p=1,\dots,d, j<p}$, meaning that a strong Hörmander condition holds locally along the curve $x_t(\phi)$.

Remark 2.8. Hypotheses (2.11) and (2.12) are local assumptions around the skeleton path $x_t(\phi)$, $t \in [0, T]$, in the sense that inequality (2.11) concerns (s, y) with $s \in [(t-1) \vee 0, t+1]$ and $y \in B(x, 1)$ with $x = x_t(\phi)$. In particular we do not assume any Lipschitz or global growth conditions, so it is not guaranteed that a unique strong solution to equation (2.1) exists. In the following, X is just a process that satisfies (2.1) on the time interval $[0, T]$.

For $K, q > 0$, $\mu \geq 1$, $h \in (0, 1]$, $n : [0, T] \to [1, +\infty)$, $\lambda : [0, T] \to (0, 1]$ and $\phi \in L^2([0, T], \mathbb{R}^n)$, we set the functions

$$\begin{aligned}
H_t &= K\left(\frac{\mu n_t}{\lambda_t}\right)^q, \\
R_t^*(\phi) &= \exp\left(-K\left(\frac{\mu n_t}{\lambda_t}\right)^{q/3}\right)\left(\frac{1}{h} + |\phi|_t^2\right)^{-1}
\end{aligned} \tag{2.13}$$

The main result of this paper is the following:

Theorem 2.9. *Let $\mu \geq 1$, $h \in (0, 1]$, $n : [0, T] \to [1, +\infty)$, $\lambda : [0, T] \to (0, 1]$, $R : [0, T] \to (0, 1]$ and $\phi \in L^2([0, T], \mathbb{R}^n)$ be such that (H_1)–(H_3) in (2.12) hold. Then there exist K, q such that, for H as in (2.13),*

$$\exp\left(-\int_0^T H_t\left(\frac{1}{R_t} + \frac{1}{h} + |\phi_t|^2\right) dt\right) \leq \mathbb{P}\Big(\sup_{t \leq T} \big|X_t - x_t(\phi)\big|_{A_{R_t}(t, x_t(\phi))} \leq 1\Big).$$

Moreover, there exist K, q such that, for H and $R_\cdot^*(\phi)$ as in (2.13), if $R_t \leq R_t^*(\phi)$ for every $t \in [0, T]$, one has

$$\mathbb{P}\Big(\sup_{t \leq T} \big|X_t - x_t(\phi)\big|_{A_{R_t}(t, x_t(\phi))} \leq 1\Big) \leq \exp\left(-\int_0^T \frac{e^{-H_t}}{R_t} + \frac{1}{H_t}\left(\frac{1}{h} + |\phi_t|^2\right) dt\right).$$

Remark 2.10. The estimates in Theorem 2.9 allow for a regime shift, meaning that the dimension of the space generated by the σ_i's and the $[\sigma_i, \sigma_j]$'s may change along the tube, and this is accounted by the variation of A_R along $x_t(\phi)$.

Remark 2.11. The fact that $R \in L(\mu, h)$ implies that $\inf_{t \in [0,T]} R_t > 0$. So, the radius of the tube is small, but cannot go to 0 at any time. Notice also that the upper bound in Theorem 2.9 only holds for small $R_t \leq R_t^*(\phi)$. This depends on the fact that, to obtain an upper bound, we need to be able to use our density estimate on the whole tube. For the lower bound it is enough that the density estimate holds on a smaller tube, contained in the one that we mean to estimate. For this reason, in this case the estimate holds for all $t \to R_t \leq 1$.

Remark 2.12. Let us look at the lower bound. If R_t is large enough (for every $t \in [0,T]$), meaning $R_t \geq (\frac{1}{h}+|\phi_t|^2)^{-1}$, then $\frac{1}{R_t} \leq (\frac{1}{h}+|\phi_t|^2)$ and the significant contribution in the exponential is given by h (regularity parameter) and the energy $\int_0^T |\phi_t|^2\,dt$. Something similar also holds for the lower bound: if R_t is large enough, meaning $H_t e^{-H_t}(\frac{1}{h}+|\phi_t|^2)^{-1} \leq R_t \leq R_t^*(\phi)$, then $\frac{e^{-H_t}}{R_t} \leq \frac{1}{H_t}(\frac{1}{h}+|\phi_t|^2)$ and the significant contribution is given by h and $\int_0^T |\phi_t|^2\,dt$. In what follows we give also explicit examples where the significant contribution comes from the radius R.

Remark 2.13. Suppose $X_t = W_t$ and $x(\phi) = 0$, so that $n_t = 1$, $\lambda_t = 1$, $\mu = 1$ and $\phi_t = 0$. Take $R_t = R$ constant. Then $|X_t - x_t(\phi)|_{A_R(t,x_t(\phi))} = R^{-1/2} W_t$ and we obtain $\exp(-C_1 T/R) \leq \mathbb{P}(\sup_{t\leq T}|W_t| \leq \sqrt{R}) \leq \exp(-C_2 T/R)$ which is consistent with the standard estimate (see [21]).

Let us suppose now, as in [23], $\sigma(t,x) = \sigma(x)$. We recall that we have defined the semi-distance d through $d(x,y) < \sqrt{R}$ if $|x-y|_{A_R(x)} < 1$. We prove in Appendix A the local equivalence of d and the Carathéodory distance d_c. Such equivalence allows us to state Theorem 2.9 in the control metric:

Theorem 2.14. *Suppose that the diffusion coefficients σ_j, $j = 1,\dots,d$, in (2.1) depend on the space variable x only and that the hypotheses of Theorem 2.9 hold. Then, denoting d_c the Carathéodory distance, we have that there exist K, q such that, for H as in (2.13),*

$$\exp\left(-\int_0^T H_t\left(\frac{1}{R_t}+\frac{1}{h}+|\phi_t|^2\right)dt\right) \leq \mathbb{P}\left(\sup_{t\leq T} d_c\big(X_t, x_t(\phi)\big) \leq \sqrt{R_t}\right).$$

Moreover, there exist K, q such that, for H and $R^*_\cdot(\phi)$ as in (2.13), if $R_t \leq R_t^*(\phi)$ for every $t \in [0,T]$, one has

$$\mathbb{P}\left(\sup_{t\leq T} d_c\big(X_t, x_t(\phi)\big) \leq \sqrt{R_t}\right) \leq \exp\left(-\int_0^T \frac{e^{-H_t}}{R_t} + \frac{1}{H_t}\left(\frac{1}{h}+|\phi_t|^2\right)dt\right).$$

We present now two examples of application.

Example 1 (Grushin diffusion). Consider a positive, fixed R and the two dimensional diffusion process

$$X_t^1 = x_1 + W_t^1, \qquad X_t^2 = x_2 + \int_0^t X_s^1\,dW_s^2.$$

Here

$$A_R A_R^T(x) = \begin{pmatrix} R & 0 \\ 0 & R(x_1^2+2R) \end{pmatrix},$$

so the associated norm is $|\xi|^2_{A_R(x)} = \frac{\xi_1^2}{R} + \frac{\xi_2^2}{R(x_1^2+2R)}$. On $\{x_1 = 0\}$, $|\xi|^2_{A_R(x)} = \frac{\xi_1^2}{R} + \frac{\xi_2^2}{2R^2}$ and consequently $\{\xi : |\xi|_{A_R(x)} \leq 1\}$ 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 [7]). 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 $B_R(x) = R\sigma(x)$. We have

$$|\xi|^2_{B_R(x)} = \frac{1}{R}\xi_1^2 + \frac{1}{Rx_1^2}\xi_2^2 \geq \frac{1}{R}\xi_1^2 + \frac{1}{R(x_1^2+2R)}\xi_2^2 = |\xi|^2_{A_R(x)},$$

and the two norms are equivalent for R small. Let us now take $x_t(\phi)=(0,0)$. We have $n_s=1$ and $\lambda_s=1$. Let $I_t(W)=\int_0^t W_s^1\,dW_s^2$. With this notation, $X_t-x_t(\phi)=(W_t^1,I_t(W))$ and using a scaling argument, we obtain

$$\begin{aligned}
e^{-C_1T/R} &\le \mathbb{P}\biggl(\sup_{t\le T/R}\biggl\{\bigl|W_t^1\bigr|^2+\frac{|I_t(W)|^2}{2}\biggr\}\le 1\biggr)\\
&=\mathbb{P}\biggl(\sup_{t\le T}\biggl\{\frac{1}{R}\bigl|W_t^1\bigr|^2+\frac{|I_t(W)|^2}{2R^2}\biggr\}\le 1\biggr)\\
&=\mathbb{P}\Bigl(\sup_{t\le T}|X_t-x_t|^2_{A_R(x_t)}\le 1\Bigr)\le e^{-C_2T/R}.
\end{aligned}$$

Example 2 (Principal invariant diffusion on the Heisenberg group). Consider on $\mathbb{R}^3$ the vector fields $\partial_{x_1}-\frac{x_2}{2}\partial_{x_3}$ and $\partial_{x_2}-\frac{x_1}{2}\partial_{x_3}$. The associated Markov process is the triple given by a Brownian motion on $\mathbb{R}^2$ and its Lévy area, that is

$$X_t^1=x_1+W_t^1,\qquad X_t^2=x_2+W_t^2,\qquad X_t^3=x_3+\frac{1}{2}\int_0^t X_s^1\,dW_s^2-\frac{1}{2}\int_0^t X_s^2\,dW_s^1.$$

We refer e.g. to [1,18,24], where gradient bounds for the heat kernel are obtained, and [9]. Since the diffusion is in dimension $n=3$ and the driving Brownian in dimension $d=2$, ellipticity cannot hold. Direct computations give

$$\sigma_1(x)=\begin{pmatrix}1\\0\\-\frac{x_2}{2}\end{pmatrix},\qquad \sigma_2(x)=\begin{pmatrix}0\\1\\\frac{x_1}{2}\end{pmatrix},\qquad [\sigma_1,\sigma_2](x)=\partial_{\sigma_1}\sigma_2-\partial_{\sigma_2}\sigma_1=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

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 the control $\phi\equiv 0$ and 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 Lévy area), we obtain

$$\begin{aligned}
\mathbb{P}\biggl(\sup_{t\le T/R}\biggl\{\bigl|W_t^1\bigr|^2+\bigl|W_t^2\bigr|^2+\frac{|A_t(W)|^2}{2}\biggr\}\le 1\biggr)&=\mathbb{P}\biggl(\sup_{t\le T}\biggl\{\frac{|W_t^1|^2+|W_t^2|^2}{R}+\frac{|A_t(W)|^2}{2R^2}\biggr\}\le 1\biggr)\\
&=\mathbb{P}\Bigl(\sup_{t\le T}|X_t|^2_{A_R(x_t(\phi))}\le 1\Bigr).
\end{aligned}$$

Appling our estimate we have

$$e^{-C_1T/R}\le\mathbb{P}\biggl(\sup_{t\le T/R}\biggl\{\bigl|W_t^1\bigr|^2+\bigl|W_t^2\bigr|^2+\frac{|A_t(W)|^2}{2}\biggr\}\le 1\biggr)\le e^{-C_2T/R}.$$

3. Lower bound for the density

We study here the lower bound for the density of X_δ.

3.1. *The key-decomposition*

We start with the decomposition of the process that will allow us to prove the lower bound in short (but not asymptotic) time.

We first use a development in stochastic Taylor series of order two of the diffusion process X defined through (2.1). This gives

$$X_t=x_0+Z_t+b(0,x_0)t+R_t, \tag{3.1}$$

where

$$Z_t = \sum_{i=1}^d a_i W_t^i + \sum_{i,j=1}^d a_{i,j} \int_0^t W_s^i \circ dW_s^j$$
$$\text{with } a_i = \sigma_i(0, x_0), a_{i,j} = \partial_{\sigma_i}\sigma_j(0, x_0) \tag{3.2}$$

and

$$\begin{aligned} R_t = & \sum_{j,i=1}^d \int_0^t \int_0^s \big(\partial_{\sigma_i}\sigma_j(u, X_u) - \partial_{\sigma_i}\sigma_j(0, x_0)\big) \circ dW_u^i \circ dW_s^j \\ & + \sum_{i=1}^d \int_0^t \int_0^s \partial_b \sigma_i(u, X_u)\, du \circ dW_s^i + \sum_{i=1}^d \int_0^t \int_0^s \partial_u \sigma_j(u, X_u)\, du \circ dW_s^i \\ & + \sum_{i=1}^d \int_0^t \int_0^s \partial_{\sigma_i} b(u, X_u) \circ dW_u^i\, ds + \int_0^t \int_0^s \partial_b b(u, X_u)\, du\, ds. \end{aligned}$$

Since $R_t = \mathcal{O}(t^{3/2})$, we expect the behavior of X_t and Z_t to be somehow close for small values of t. Our first goal is to give a decomposition for Z_t in (3.2). We start introducing some notation. We fix $\delta > 0$ and set

$$s_k(\delta) = \frac{k}{d}\delta, \quad k = 1, \dots, d.$$

We now consider the following random variables: for $i, k = 1, \dots, d$,

$$\Delta_k^i(\delta, W) = W_{s_k(\delta)}^i - W_{s_{k-1}(\delta)}^i, \qquad \Delta_k^{i,j}(\delta, W) = \int_{s_{k-1}(\delta)}^{s_k(\delta)} \big(W_s^i - W_{s_{k-1}(\delta)}^i\big) \circ dW_s^j. \tag{3.3}$$

Notice that $\Delta_k^{i,j}(\delta, W)$ is the Stratonovich integral, but for $i \neq j$ it coincides with the Itô integral. When no confusion is possible we use the short notation $s_k = s_k(\delta)$, $\Delta_k^i = \Delta_k^i(\delta, W)$, $\Delta_k^{i,j} = \Delta_k^{i,j}(\delta, W)$. We also denote the random vector $\Delta(\delta, W)$ in $\mathbb{R}^m$

$$\begin{aligned} \Delta_l(\delta, W) &= \Delta_p^{i,p}(\delta, W) \quad \text{if } l = l(i, p) \text{ with } i \neq p, \\ &= \Delta_p^p(\delta, W) \quad \text{if } l = l(i, p) \text{ with } i = p. \end{aligned} \tag{3.4}$$

(recall $l(i, p)$ in (2.3)). Moreover, with $\sum_{l>p}^d = \sum_{p=1}^d \sum_{l=p+1}^d$, we define

$$\begin{aligned} V(\delta, W) &= \sum_{p=1}^d \Bigg[\sum_{i \neq p} \Delta_p^i + \sum_{i \neq j, i \neq p, j \neq p} a_{i,j} \Delta_p^{i,j} + \sum_{l=p+1}^d \sum_{i \neq p} \sum_{j \neq l} a_{i,j} \Delta_l^j \Delta_p^i + \frac{1}{2} \sum_{i \neq p} a_{i,i} \big|\Delta_p^i\big|^2 \Bigg]; \\ \varepsilon_p(\delta, W) &= \sum_{l>p}^d \sum_{j \neq l} a_{p,j} \Delta_l^j + \sum_{p>l}^d \sum_{j \neq l} a_{j,p} \Delta_l^j + \sum_{j \neq p} a_{p,j} \Delta_p^j, \quad p = 1, \dots, d; \\ \eta_p(\delta, W) &= \frac{1}{2} a_{p,p} \big|\Delta_p^p\big|^2 + \sum_{l>p}^d a_{p,l} \Delta_l^l \Delta_p^p + \Delta_p^p \varepsilon_p(\delta, W), \quad p = 1, \dots, d. \end{aligned} \tag{3.5}$$

We have the following decomposition:

Lemma 3.1. *Let $\Delta(\delta, W)$ and $A(0, x_0)$ be given in (3.4) and (2.5) respectively. One has*

$$Z_\delta = V(\delta, W) + A(0, x_0)\Delta(\delta, W) + \eta(\delta, W), \tag{3.6}$$

where $V(\delta, W)$ is given in (3.5) and $\eta(\delta, W) = \sum_{p=1}^{d} \eta_p(\delta, W)$, $\eta_p(\delta, W)$ being given in (3.5).

The proof of Lemma 3.1 is quite long, so it is postponed to Appendix B.

Remark 3.2. 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: $(W_s^p - W_{s_{k-1}}^p)_{s_{k-1}\leq s\leq s_k}$, $p = 1, \dots, d$. Let us fix p. For $s \in (s_{p-1}, s_p)$ we have the processes $(W_s^i - W_{s_{p-1}}^i)_{s_{p-1}\leq s\leq s_p}$, $i = 1, \dots, d$. 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 $\Delta_p^{i,p} = \int_{s_{p-1}}^{s_p} (W_s^i - W_{s_{p-1}}^i)\, dW_s^p$, $i \neq p$. 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. Under appropriate hypotheses on the controls, its covariance matrix is non degenerate. This a non trivial matter: notice that $W_s^i - W_{s_{p-1}}^i$ are Brownian trajectories, so our goal is to find a subset of the Wiener space which has strictly positive probability and such that the corresponding paths give a non degenerate covariance matrix (see Q_p in next formula (3.8)). We treat in Appendix C the problem of the choice of the controls. The argument is based on a result in [17] concerning the variance of the Brownian path.

Remark 3.3. As we mentioned in Remark 3.2, the central idea in this paper is to isolate a Gaussian random variable which represents the principal term of the short time behavior. Then, we employ it in order to make our analysis. The same strategy has already been used in [8] and in [27] in the framework of degenerate diffusion processes satisfying a weak Hörmander condition. The construction of the Gaussian principal term was completely different (and much less involved) there.

We now emphasize the scaling in δ in the random vector $\Delta(\delta, W)$. We define $B_t = \delta^{-1/2} W_{t\delta}$ and denote

$$\begin{aligned}
\Theta_l &= \frac{1}{\delta}\Delta_p^{i,p} = \int_{\frac{p-1}{d}}^{\frac{p}{d}} \big(B_s^i - B_{\frac{p-1}{d}}^i\big)\, dB_s^p \quad \text{if } l = l(i,p) \text{ with } i \neq p, \\
&= \frac{1}{\sqrt{\delta}}\Delta_p^p = B_{\frac{p}{d}}^p - B_{\frac{p-1}{d}}^p \quad \text{if } l = l(i,p) \text{ with } i = p,
\end{aligned}$$

$l(i, p)$ being given in (2.3). For $p = 1, \dots, d$ we denote with $\Theta_{(p)}$ the pth block of Θ with length d, that is

$$\Theta_{(p)} = (\Theta_{(p-1)d+1}, \dots, \Theta_{pd}),$$

so that $\Theta = (\Theta_{(1)}, \dots, \Theta_{(d)})$. We will also denote

$$l(p) = l(p, p) = (p-1)d + p \quad \text{and} \quad \Theta_{l(p)} = \frac{1}{\sqrt{\delta}}\Delta_p^p. \tag{3.7}$$

Consider now the σ field

$$\mathcal{G} := \sigma\big(W_s^j - W_{s_{p-1}(\delta)}^j, s_{p-1}(\delta) \leq s \leq s_p(\delta), p = 1, \dots, d, j \neq p\big).$$

Then conditionally to $\mathcal{G}$ the random variables $\Theta_{(p)}$, $p=1,\dots,d$ are independent centered Gaussian d dimensional vectors and the covariance matrix Q_p of $\Theta_{(p)}$ is given by

$$\begin{aligned}
&Q_p^{p,j}=Q_p^{j,p}=\int_{\frac{p-1}{d}}^{\frac{p}{d}}\big(B_s^j-B_{\frac{p-1}{d}}^j\big)\,ds, \quad j\neq p,\\
&Q_p^{i,j}=\int_{\frac{p-1}{d}}^{\frac{p}{d}}\big(B_s^j-B_{\frac{p-1}{d}}^j\big)\big(B_s^i-B_{\frac{p-1}{d}}^i\big)\,ds, \quad j\neq p, i\neq p,\\
&Q_p^{p,p}=\frac{1}{d}.
\end{aligned} \tag{3.8}$$

It is easy to see that $\det Q_p\neq 0$ almost surely. It follows that conditionally to $\mathcal{G}$ the random variable $\Theta=(\Theta_{(1)},\dots,\Theta_{(d)})$ is a centered $m=d^2$ dimensional Gaussian vector. The explicit density of Θ represents the main instrument in our analysis. Its covariance matrix Q is a block-diagonal matrix built with Q_p, $p=1,\dots,d$:

$$Q=\begin{pmatrix} Q_1 & & \\ & \ddots & \\ & & Q_d\end{pmatrix}. \tag{3.9}$$

In particular $\det Q=\prod_{p=1}^d \det Q_p\neq 0$ almost surely, and $\lambda_*(Q)=\min_{p=1,\dots,d}\lambda_*(Q_p)$. We also have $\lambda^*(Q)=\max_{p=1,\dots,d}\lambda^*(Q_p)$. We will need to work on subsets where we have a quantitative control of this quantities, so we will come back soon on these covariance matrices. But let us show now how one can rewrite decomposition (3.6) in terms of the random vector Θ. As a consequence, the scaled matrix $A_\delta=A_\delta(0,x_0)$ in (2.8) will appear.

We denote by $A_\delta^i\in\mathbb{R}^m$, $i=1,\dots,n$ the rows of the matrix A_δ. We also denote $S=\langle A_\delta^1,\dots,A_\delta^n\rangle\subset\mathbb{R}^m$ and $S^\perp$ its orthogonal. Under Assumption 2.2 the columns of A_δ span $\mathbb{R}^n$ so the rows $A_\delta^1,\dots,A_\delta^n$ are linearly independent and $S^\perp$ has dimension $m-n$. We take Γ_δ^i, $i=n+1,\dots,m$ to be an orthonormal basis in $S^\perp$ and we denote $\Gamma_\delta^i=A_\delta^i(0,x_0)$ for $i=1,\dots,n$. We also denote $\underline{\Gamma}_\delta$ the $(m-n)\times m$ matrix with rows Γ_δ^i, $i=n+1,\dots,m$. Finally we denote by Γ_δ the $m\times m$ dimensional matrix with rows Γ_δ^i, $i=1,\dots,m$. Notice that

$$\Gamma_\delta\Gamma_\delta^T=\begin{pmatrix} A_\delta A_\delta^T(0,x_0) & 0\\ 0 & \mathrm{Id}_{m-n}\end{pmatrix}, \tag{3.10}$$

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

$$|y|_{\Gamma_\delta}^2=|y_{(1)}|_{A_\delta(0,x_0)}^2+|y_{(2)}|^2,$$

where we recall that $|y|_{\Gamma_\delta}^2=\langle(\Gamma_\delta\Gamma_\delta^T)^{-1}y,y\rangle$. For $a\in\mathbb{R}^m$ we define the immersion

$$J_a:\mathbb{R}^n\to\mathbb{R}^m, \quad \big(J_a(z)\big)_i=z_i, i=1,\dots,n \quad\text{and}\quad \big(J_a(z)\big)_i=\big\langle\Gamma_\delta^i,a\big\rangle, \quad i=n+1,\dots,m.$$

In particular $J_0(z)=(z,0,\dots,0)$ and

$$|J_0z|_{\Gamma_\delta}=|z|_{A_\delta(0,x_0)}.$$

Finally we denote

$$\begin{aligned}
&V_\omega=V(\delta,W),\\
&\eta_\omega(\Theta)=\sum_{p=1}^d\left(\frac{a_{p,p}}{2}\delta\Theta_{l(p)}^2+\delta^{1/2}\Theta_{l(p)}\varepsilon_p(\delta,W)+\sum_{q>p}^d a_{p,q}\delta\Theta_{l(q)}\Theta_{l(p)}\right),
\end{aligned} \tag{3.11}$$

where $V(\delta, W)$ and $\varepsilon_p(\delta, W)$ are defined in (3.5) and $\Theta_{l(p)}$ is given in (3.7). We notice that $\eta_\omega(\Theta) = \sum_{p=1}^d \eta_p(\delta, W)$, $\eta_p(\delta, W)$ being defined in (3.5). We also remark that both $V(\delta, W)$ and $\varepsilon_p(\delta, W)$ are $\mathcal{G}$-measurable, so (3.11) stresses a dependence on ω which is $\mathcal{G}$-measurable and a dependence on the random vector Θ whose conditional law w.r.t. $\mathcal{G}$ is Gaussian.

Now the decomposition (3.6) 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

$$J_\Theta(Z_\delta) = J_0(V_\omega) + \Gamma_\delta\Theta + J_0\bigl(\eta_\omega(\Theta)\bigr).$$

We now multiply with Γ_δ^{-1}: setting

$$\widetilde{Z} = \Gamma_\delta^{-1} J_\Theta(Z_\delta), \qquad \widetilde{V}_\omega = \Gamma_\delta^{-1} J_0(V_\omega), \qquad \widetilde{\eta}_\omega(\Theta) = \Gamma_\delta^{-1} J_0\bigl(\eta_\omega(\Theta)\bigr) \tag{3.12}$$

and

$$G = \Theta + \widetilde{\eta}_\omega(\Theta),$$

we get

$$\widetilde{Z} = \widetilde{V}_\omega + G. \tag{3.13}$$

Our aim is to obtain lower bounds for the density of X_δ. As we mentioned at the beginning of this section, X_δ is close to Z_δ, which itself is strongly related to $\widetilde{Z}$ (we discuss this in details in Section 3.2). So, we focus now on $\widetilde{Z}$. We will work conditionally on $\mathcal{G}$ and we will use some localization procedures. Since $\widetilde{V}_\omega$ is $\mathcal{G}$-measurable, it represents just a translation which turns out to be small (see (3.26)). Let us now consider $G = \Theta + \widetilde{\eta}_\omega(\Theta)$. Conditionally to $\mathcal{G}$, Θ is a Gaussian random variable, so it is Θ which gives us access to explicit estimates (this is the core of our analysis). Now, G appears as a perturbation of Θ, so we will use the local inverse function theorem in order to transfer estimates from Θ to G. The fact that $\widetilde{\eta}_\omega(\Theta)$ is an explicit function of Θ allows us to use some explicit localization procedures, in order to ensure that $\nabla_\Theta \widetilde{\eta}_\omega(\Theta)$ is small. So, $\nabla_\Theta G = \mathrm{Id} + \nabla_\Theta \widetilde{\eta}_\omega(\Theta)$ is invertible and finally we can obtain quantitative estimates (this is done in Appendix D).

3.2. *Localized density for the principal term*

We study here the density of $\widetilde{Z}$ in (3.13), "around" (that is, localized on) a suitable set of Brownian trajectories (see $\Lambda_{\rho,\varepsilon}$ in next (3.15)), where we have a quantitative control on the "non-degeneracy" (conditionally to $\mathcal{G}$) of the main Gaussian random variable Θ.

We denote

$$q_p(B) = \sum_{j \neq p} \bigl| B^j_{\frac{p}{d}} - B^j_{\frac{p-1}{d}} \bigr| + \sum_{j \neq p, i \neq p, i \neq j} \left| \int_{\frac{p-1}{d}}^{\frac{p}{d}} \bigl(B^j_s - B^j_{\frac{i-1}{d}} \bigr)\, dB^i_s \right|. \tag{3.14}$$

For fixed $\varepsilon, \rho > 0$, we define

$$\begin{aligned} \Lambda_{\rho,\varepsilon,p} &= \Bigl\{ \det Q_p \geq \varepsilon^\rho, \sup_{\frac{p-1}{d} \leq t \leq \frac{p}{d}} \sum_{j \neq p} \bigl| B^j_t - B^j_{\frac{p-1}{d}} \bigr| \leq \varepsilon^{-\rho}, q_p(B) \leq \varepsilon \Bigr\}, \quad p = 1, \dots, d, \\ \Lambda_{\rho,\varepsilon} &= \bigcap_{p=1}^d \Lambda_{\rho,\varepsilon,p}. \end{aligned} \tag{3.15}$$

Notice that $\Lambda_{\rho,\varepsilon,p} \in \mathcal{G}$ for every $p = 1, \dots, d$. We have the following.

Lemma 3.4. *Let $\Lambda_{\rho,\varepsilon}$ be as in (3.15). There exist c and ε_* such that for every $\varepsilon \le \varepsilon_*$ one has*

$$\mathbb{P}(\Lambda_{\rho,\varepsilon}) \ge c \times \varepsilon^{\frac{1}{2}m(d+1)}. \tag{3.16}$$

Proof. We apply here Proposition C.3. Let $p \in \{1,\dots,d\}$ be fixed and consider the Brownian motion $\widehat{B}_t = \sqrt{d}(B_{\frac{p-1+t}{d}} - B_{\frac{p-1}{d}})$. Let $Q(\widehat{B})$ be the matrix in (C.1). Up to a permutation of the components of $\widehat{B}$, we easily get $Q^{p,p}(\widehat{B}) = d \times Q_p^{p,p}$, $Q^{p,j}(\widehat{B}) = d^{3/2} \times Q_p^{p,j}$ for $j \ne p$, $Q^{i,j}(\widehat{B}) = d^2 \times Q_p^{i,j}$ for $i \ne p$ and $j \ne p$. Therefore,

$$\det Q_p = d^{2d-1} \det Q(\widehat{B}) \ge \det Q(\widehat{B}).$$

Let now $q(\widehat{B})$ be the quantity defined in (C.3). With $q_p(B)$ as in (3.14), it easily follows that

$$q_p(B) \le q(\widehat{B}).$$

Moreover, $\sup_{t\le 1} |\widehat{B}_t| = \sqrt{d} \sup_{\frac{p-1}{d} \le s \le \frac{p}{d}} |B_s - B_{\frac{p-1}{d}}| \ge \sup_{\frac{p-1}{d} \le s \le \frac{p}{d}} |B_s - B_{\frac{p-1}{d}}|$. As a consequence, with $\Upsilon_{\rho,\varepsilon}$ the set defined in (C.4), we get

$$\Upsilon_{\rho,\varepsilon}(\widehat{B}) \subset \Lambda_{\rho,\varepsilon,p}$$

and by using (C.5), we may find some constants c and ε_* such that $\mathbb{P}(\Lambda_{\rho,\varepsilon,p}) \ge c\varepsilon^{\frac{1}{2}d(d+1)}$, for $\varepsilon \le \varepsilon_*$. This holds for every p. Since $\Lambda_{\rho,\varepsilon} = \bigcap_{p=1}^d \Lambda_{\rho,\varepsilon,p}$, by using the independence property we get (3.16). □

Let Q be the matrix in (3.9). On the set $\Lambda_{\rho,\varepsilon} \in \mathcal{G}$ we have $\det Q = \prod_{p=1}^d \det Q_p \ge \varepsilon^{d\rho}$. Remark that

$$\frac{\lambda^*(Q)}{\sqrt{m}} \le |Q|_l := \left(\frac{1}{m} \sum_{1\le i,j\le m} Q_{i,j}^2\right)^{1/2} \le \lambda^*(Q). \tag{3.17}$$

For $a > 0$ we introduce the following function,

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

which is a mollified version of $1_{[-a,a]}$. We can now define our *localization variables*.

$$\widetilde{U}_\varepsilon = \big(\psi_{a_1}(1/\det Q)\big)\psi_{a_2}\big(|Q|_l\big)\psi_{a_3}\big(q(B)\big), \quad \text{with } a_1 = \varepsilon^{-d\rho}, a_2 = \varepsilon^{-2\rho}, a_3 = d\varepsilon \tag{3.18}$$

in which we have set

$$q(B) = \sum_{p=1}^d q_p(B).$$

Remark that $\widetilde{U}_\varepsilon$ is measurable w.r.t. $\mathcal{G}$. The following inclusions hold: for every ε small enough,

$$\Lambda_{\rho,\varepsilon} \subset \left\{\det Q \ge \varepsilon^{d\rho}, |Q|_l \le \varepsilon^{-2\rho}, q(B) \le d\varepsilon\right\} \subset \{\widetilde{U}_\varepsilon = 1\} \subset \{\widetilde{U}_\varepsilon \ne 0\}.$$

We can consider $\widetilde{U}_\varepsilon$ as a smooth version of the indicator function of $\Lambda_{\rho,\varepsilon}$. We also define, for fixed $r > 0$,

$$\bar{U}_r = \prod_{i=1}^n \psi_r(\Theta_i). \tag{3.19}$$

In order to state a lower estimate for the (localized) density of $\widetilde{Z}$ in (3.13), we define the following set of constants:

$$\mathcal{C}_0 = \left\{ C > 0 : \text{there are } K, q > 0 \text{ such that } C = \exp\left(K \left(\frac{\kappa}{\lambda(0, x_0)} \right)^q \right) \right\}, \tag{3.20}$$

κ being the constant Assumption 2.1. We set

$$1/\mathcal{C}_0 = \{C > 0 : 1/C \in \mathcal{C}_0\}.$$

Lemma 3.5. *Suppose Assumption 2.1 and 2.2 both hold. Let $U_{\varepsilon,r} = \widetilde{U}_\varepsilon \bar{U}_r$, $\widetilde{U}_\varepsilon$ and $\bar{U}_r$ being defined in (3.18) and (3.19) respectively, with $\rho = \frac{1}{8m}$. Set $d\mathbb{P}_{U_{\varepsilon,r}} = U_{\varepsilon,r}\, d\mathbb{P}$ and let $p_{\widetilde{Z},U_{\varepsilon,r}}$ denote the density of $\widetilde{Z}$ in (3.13) when we endow Ω with the measure $\mathbb{P}_{U_{\varepsilon,r}}$. Then there exist $C \in \mathcal{C}_0$, $\varepsilon, r \in 1/\mathcal{C}_0$ such that for $|z| \le r/2$,*

$$p_{\widetilde{Z},U_{\varepsilon,r}}(z) \ge \frac{1}{C}. \tag{3.21}$$

Proof. *Step* 1. Lower bound for the localized conditional density given $\mathcal{G}$.

Let $p_{\widetilde{Z},\bar{U}_r|\mathcal{G}}$ denote the localized density w.r.t. the localization $\bar{U}_r$ of $\widetilde{Z}$ conditioned to $\mathcal{G}$, i.e.

$$\mathbb{E}\big[f(\widetilde{Z})\bar{U}_r|\mathcal{G}\big] = \int f(z) p_{\widetilde{Z},\bar{U}_r|\mathcal{G}}(z)\, dz, \tag{3.22}$$

for f positive, measurable, with support included in $B(0, r/2)$. We start proving that there exist $C \in \mathcal{C}_0$, $\varepsilon, r \in 1/\mathcal{C}_0$ such that, on $\widetilde{U}_\varepsilon \neq 0$, for $|z| \le r/2$

$$p_{\widetilde{Z},\bar{U}_r|\mathcal{G}}(z) \ge \frac{1}{C}.$$

We recall (3.13): $\widetilde{Z} = \widetilde{V}_\omega + \Theta + \widetilde{\eta}_\omega(\Theta)$, where $\omega \mapsto \widetilde{V}_\omega$ and $\omega \mapsto \widetilde{\eta}_\omega(\cdot)$ are both $\mathcal{G}$-measurable and the conditional law of Θ given $\mathcal{G}$ is Gaussian. This allows us to use the results in Appendix D. In particular, we are interested in working on the set $\{\widetilde{U}_\varepsilon \neq 0\} \in \mathcal{G}$, so one has to keep in mind that $\omega \in \{\widetilde{U}_\varepsilon \neq 0\}$.

On $\widetilde{U}_\varepsilon \neq 0$, by (3.18) and (3.17) one has $\lambda^*(Q) \le 2\sqrt{m}\varepsilon^{-2\rho}$, and

$$\frac{\varepsilon^{d\rho}}{2} \le \det Q \le \lambda_*(Q)\lambda^*(Q)^{m-1} \le \lambda_*(Q)(2\sqrt{m})^{m-1}\varepsilon^{-2\rho(m-1)},$$

and this gives $\lambda_*(Q) \ge \frac{\varepsilon^{3m\rho}}{(2\sqrt{m})^m}$. So, fixing $\rho = 1/(8m)$, for $\varepsilon \le \varepsilon^*$,

$$\frac{1}{16m^2} \frac{\lambda_*(Q)}{\lambda^*(Q)} \ge C_m \varepsilon^{3m\rho + 2\rho} \ge \varepsilon. \tag{3.23}$$

To apply (D.8) to $G = \Theta + \widetilde{\eta}_\omega(\Theta)$ we need to check the hypothesis of Lemma D.3. We are going to use the notation of Appendix D, in particular for $c_*(\widetilde{\eta}_\omega, r)$ in (D.5) and $c_i(\widetilde{\eta}_\omega)$, $i = 2, 3$, in (D.1). Recall that $\widetilde{\eta}_\omega$ is defined in (3.12) through η_ω given in (3.11). Since the third order derivatives of η_ω are null, we have $c_3(\widetilde{\eta}_\omega) = 0$. Also, 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 $|\partial_{i,j}\eta_\omega(\Theta)| \le \delta \sum_{i,j} |a_{i,j}|$. Using (2.9) we obtain

$$\big|\partial_{i,j}\widetilde{\eta}_\omega(\Theta)\big| = \big|J_0\big(\partial_{i,j}\eta_\omega(\Theta)\big)\big|_{\Gamma_\delta} = \big|\partial_{i,j}\eta_\omega(\Theta)\big|_{A_\delta} \le \frac{\sum_{i,j} |a_{i,j}|}{\lambda(0, x_0)} \le C \in \mathcal{C}_0.$$

So, with h_{η_ω} as in (D.2), we get

$$h_{\eta_\omega} = \frac{1}{16m^2(c_2(\widetilde{\eta}_\omega) + \sqrt{c_3(\widetilde{\eta}_\omega)})} \ge \frac{1}{C_1}, \quad \exists C_1 \in \mathcal{C}_0 \tag{3.24}$$

We compute now the first order derivatives. For $j \notin \{l(p) : p = 1, \dots, d\}$ we have $\partial_j \eta_\omega = 0$ and for $j = l(p)$ we have

$$\partial_j \eta_\omega(\Theta) = \delta \sum_{q=p}^{d} a_{p\wedge q, p\vee q} \Theta_{l(q)} + \sqrt{\delta}\varepsilon_j(\delta, W).$$

So, as above, we obtain $|\partial_j \widetilde{\eta}_\omega(\Theta)| \le C(|\Theta| + |\varepsilon_j(\delta, W)|/\sqrt{\delta})$. Remark now that on $\{\bar{U}_r \neq 0\}$ we have $|\Theta| \le Cr$, and on $\{\widetilde{U}_\varepsilon \neq 0\}$ we have $q(B) \le 2d\varepsilon$, so

$$\sum_{j=1}^{d} \bigl|\varepsilon_j(\delta, W)\bigr| \le C\sqrt{\delta} q(B) \le C\sqrt{\delta}\varepsilon.$$

Therefore

$$c_*(\widetilde{\eta}_\omega, 16r) \le C_2(r + \varepsilon), \quad \exists C_2 \in \mathcal{C}_0. \tag{3.25}$$

We also consider the following estimate of $|\widetilde{V}_\omega| = |V_\omega|_{A_\delta}$. First, we rewrite V_ω as follows:

$$V_\omega = \sum_p a_p \mu_p(\delta, W) + \sum_p \psi_p(\delta, W), \quad \text{with}$$

$$\mu_p(\delta, W) = \sum_{i\neq p} \Delta_i^p \quad \text{and} \quad \psi_p(\delta, W) = \sum_{i\neq j, i\neq p, j\neq p} a_{i,j} \Delta_p^{i,j} + \sum_{l=p+1}^{d} \sum_{i\neq p} \sum_{j\neq l} a_{i,j} \Delta_l^j \Delta_p^i + \frac{1}{2} \sum_{i\neq p} a_{i,i} \bigl|\Delta_p^i\bigr|^2.$$

Using again (2.9) we have

$$\left|\sum_{p=1}^{d} a_p \mu_p(\delta, W)\right|_{A_\delta} = \frac{1}{\sqrt{\delta}} \left| A_\delta J_0 \left(\sum_{p=1}^{d} \mu_p(\delta, W) \right) \right|_{A_\delta} \le \sum_{p=1}^{d} \frac{1}{\sqrt{\delta}} \bigl|\mu_p(\delta, W))\bigr| \le Cq(B)$$

and

$$\bigl|\psi(\delta, W)\bigr|_{A_\delta} \le \frac{|\psi(\delta, W)|}{\delta\sqrt{\lambda(0, x_0)}} \le Cq(B).$$

Since $\omega \in \{\widetilde{U}_\varepsilon \neq 0\}$ we get

$$|\widetilde{V}_\omega| \le Cq(B) \le C_3\varepsilon, \quad \exists C_3 \in \mathcal{C}_0. \tag{3.26}$$

We consider (3.26), and fix $\frac{r}{\varepsilon} = 2C_3 \in \mathcal{C}_0$, so $|\widetilde{V}_\omega| \le r/2$. Then we consider (3.25) and we obtain

$$c_*(\widetilde{\eta}_\omega, 4r) \le C_2(2C_3 + 1)\varepsilon \le \varepsilon^{1/2}, \quad \text{for } \varepsilon \le \frac{1}{(4C_2C_3)^2} \in \frac{1}{\mathcal{C}_0}.$$

Moreover, looking at (3.24)

$$r = 2C_3\varepsilon \le \frac{1}{C_1} \quad \text{for } \varepsilon \le \frac{1}{2C_1C_3} \in \frac{1}{\mathcal{C}_0}.$$

So, with

$$\varepsilon = \varepsilon^* \wedge \frac{1}{(4C_2C_3)^2} \wedge \frac{1}{2C_1C_3} \in \frac{1}{\mathcal{C}_0},$$

and $r = 2C_3\varepsilon$ we have

$$|\widetilde{V}_\omega| \le \frac{r}{2}, \qquad c_*(\widetilde{\eta}_\omega, 4r) \le \varepsilon^{1/2}, \quad r \le \frac{1}{C_1}.$$

Now, by using also (3.23) and (3.24), it follows that (D.6) holds, and we can apply Lemma D.3. We obtain

$$\frac{1}{K \det Q^{1/2}} \exp\left(-\frac{K}{\lambda_*(Q)}|z|^2\right) \le p_{G,\bar{U}_r|\mathcal{G}}(z)$$

for $|z| \le r$, where K does not depend on σ, b. Remark that, using $\lambda_*(Q) \ge \frac{\varepsilon^{3m\rho}}{(2\sqrt{m})^m}$, $\rho = \frac{1}{8m}$, $r/\varepsilon = 2C_1$ and $\varepsilon \le 1/(4C_2C_1)^2$,

$$\begin{aligned}\frac{|z|^2}{\lambda_*(Q)} &\le \frac{(2\sqrt{m})^m r^2}{\varepsilon^{3m\rho}} \le (2\sqrt{m})^m \frac{r^2}{\varepsilon} \le (2\sqrt{m})^m \frac{r^2}{\varepsilon^2}\varepsilon \\ &\le (2\sqrt{m})^m (2C_1)^2 \varepsilon \le \bar{K},\end{aligned}$$

where $\bar{K}$ does not depend on σ, b. Therefore $p_{G,\bar{U}_r|\mathcal{G}}(z) \ge \frac{1}{C}$, for $|z| \le r$, for some $C \in \mathcal{C}_0$, on $\widetilde{U}_\varepsilon \neq 0$. Now, by recalling that $|\widetilde{V}_\omega| \le r/2$ and (3.13), we have

$$p_{\widetilde{Z},\bar{U}_r|\mathcal{G}}(z) \ge \frac{1}{C}, \quad \text{for } |z| \le r/2 \text{ on the set } \{\widetilde{U}_\varepsilon \neq 0\}. \tag{3.27}$$

Step 2. We get rid of the conditioning on $\mathcal{G}$, to have non-conditional bound for $p_{\widetilde{Z},U_{\varepsilon,r}}$.

Since $\widetilde{U}_\varepsilon$ is $\mathcal{G}$ measurable, for every non-negative and measurable function f with support included in $B(0, r/2)$ we have

$$\mathbb{E}\big(f(\widetilde{Z})U_{\varepsilon,r}\big) = \mathbb{E}\big(\widetilde{U}_\varepsilon \mathbb{E}\big(f(\widetilde{Z})\bar{U}_r|\mathcal{G}\big)\big).$$

By (3.22) and (3.27), we obtain

$$\mathbb{E}\big(f(\widetilde{Z})U_{\varepsilon,r}\big) \ge \frac{1}{C}\mathbb{E}(\widetilde{U}_\varepsilon) \int f(z)\,dz.$$

Since $\Lambda_{\rho,\varepsilon} \subset \{\widetilde{U}_\varepsilon = 1\}$, $\mathbb{E}(\widetilde{U}_\varepsilon) \ge \mathbb{P}(\Lambda_{\rho,\varepsilon})$, so by using (3.16) and $\varepsilon \in 1/\mathcal{C}_0$ we finally get that $\mathbb{E}(\widetilde{U}_\varepsilon) \ge \frac{1}{C}$, so (3.21) is proved. □

3.3. *Lower bound for the transition density*

We study here a lower bound for the density of X_δ, X being the solution to (2.1). Recall decomposition (3.1):

$$X_\delta - x_0 - b(0, x_0)\delta = Z_\delta + R_\delta.$$

Our aim is to "transfer" the lower bound for $\widetilde{Z} = \Gamma_\delta^{-1} J_\Theta(Z_\delta)$ already studied in Lemma 3.5 to a lower bound for X_δ. In order to set up this program, we use results on the distance between probability densities which have been developed in [2]. In particular, we are going to use now Malliavin calculus. Appendix E is devoted to a summary of all the results and notation the present section is based on. In particular, we denote with D the Malliavin derivative with respect to W, the Brownian motion driving the original equation (2.1).

But first of all, we need some properties of the matrix Γ_δ, which can be resumed as follows. We set $\mathrm{SO}(d)$ the set of the $d \times d$ orthogonal matrices and we denote with Id_d the $d \times d$ identity matrix.

Lemma 3.6. *Set* $A_\delta = A_\delta(0, x_0)$ *and* $\bar{\Sigma} = \bar{\Sigma} = \mathrm{Diag}(\lambda_1(A_\delta), \dots, \lambda_n(A_\delta))$, $\lambda_i(A_\delta)$, $i = 1, \dots, n$, *being the singular values of* A_δ (*which are strictly positive because* A_δ *has full rank*). *Let* Γ_δ *be as in* (3.10). *There exist* $\mathcal{U} \in \mathrm{SO}(n)$,

$\underline{\mathcal{U}} \in \mathrm{SO}(m-n)$ *and* $\mathcal{V} \in \mathrm{SO}(m)$ *such that*

$$\Gamma_\delta = \begin{pmatrix} \mathcal{U} & 0 \\ 0^T & \underline{\mathcal{U}} \end{pmatrix} \begin{pmatrix} \bar{\Sigma} & 0 \\ 0^T & \mathrm{Id}_{m-n} \end{pmatrix} \mathcal{V}^T,$$

where 0 *denotes a null* $n \times (m-n)$ *matrix.*

Proof. We recall that

$$\Gamma_\delta = \begin{pmatrix} A_\delta \\ \underline{\Gamma}_\delta \end{pmatrix},$$

where $\underline{\Gamma}_\delta$ is a $(m-n) \times n$ matrix whose rows are vectors of $\mathbb{R}^m$ which are orthonormal and orthogonal with the rows of A_δ. We take a singular value decomposition for A_δ and for $\underline{\Gamma}_\delta$. So, there exist $\mathcal{U} \in \mathrm{SO}(n)$ and $\bar{\mathcal{V}} \in \mathrm{SO}(m)$ such that

$$A_\delta = \mathcal{U}(\bar{\Sigma}\ 0)\bar{\mathcal{V}}^T,$$

0 denoting the $n \times (m-n)$ null matrix. Similarly, there exist $\underline{\mathcal{U}} \in \mathrm{SO}(m-n)$ and $\underline{\mathcal{V}} \in \mathrm{SO}(m)$ such that

$$\underline{\Gamma}_\delta = \underline{\mathcal{U}}\big(0^T\ \mathrm{Id}_{m-n}\big)\underline{\mathcal{V}}^T,$$

the diagonal matrix being Id_{m-n} because the rows of $\underline{\Gamma}_\delta$ are orthonormal. Therefore, we get

$$\Gamma_\delta = \begin{pmatrix} \mathcal{U} & 0 \\ 0^T & \underline{\mathcal{U}} \end{pmatrix} \begin{pmatrix} \bar{\Sigma} & 0 \\ 0^T & \mathrm{Id}_{m-n} \end{pmatrix} \mathcal{V}^T,$$

where $\mathcal{V}$ is a $m \times m$ matrix whose first n columns are given by the first n columns of $\overline{\mathcal{V}}$ and the remaining $m-n$ columns are given by the last $m-n$ columns of $\underline{\mathcal{V}}$. Moreover, since each row of A_δ is orthogonal to any row of $\underline{\Gamma}_\delta$, it immediately follows that all columns of $\mathcal{V}$ are orthogonal. This proves that $\mathcal{V} \in \mathrm{SO}(m)$, and the statement follows. □

Then we have

Lemma 3.7. *Suppose Assumption 2.1 and 2.2 both hold. Let* $U_{\varepsilon,r}$ *denote the localization in Lemma 3.5 and let* $\mathcal{U}$ *and* $\bar{\Sigma}$ *be the matrices in Lemma 3.6. Set*

$$\alpha = \mathcal{U}\bar{\Sigma} \quad \textit{and} \quad \widehat{X}_\delta = \alpha^{-1}\big(X_\delta - x_0 - b(0,x_0)\delta\big).$$

Then there exist $C \in \mathcal{C}_0$, $\delta_*, r \in 1/\mathcal{C}_0$ *such that for* $\delta \le \delta_*$, $|z| \le r/2$,

$$p_{\widehat{X}_\delta, U_{\varepsilon,r}}(z) \ge \frac{1}{C},$$

$p_{\widehat{X}_\delta, U_{\varepsilon,r}}$ *denoting the density of* $\widehat{X}_\delta$ *with respect to the measure* $\mathbb{P}_{U_{\varepsilon,r}}$.

Proof. We set $\widehat{Z}_\delta = \alpha^{-1} Z_\delta$ and we use Proposition E.1, with the localization $U = U_{\varepsilon,r}$, applied to $F = \widehat{X}_\delta$ and $G = \widehat{Z}_\delta$. Recall that the requests in (1) of Proposition E.1 involve several quantities: the lowest singular value (that in this case coincides with the lowest eigenvalue) $\lambda_*(\gamma_{\widehat{X}_\delta})$ and $\lambda_*(\gamma_{\widehat{Z}_\delta})$ of the Malliavin covariance matrix of $\widehat{X}_\delta$ and $\widehat{Z}_\delta$ respectively, as well as $m_{U_{\varepsilon,r}}(p)$ in (E.1), the Sobolev-Malliavin norms $\|\widehat{X}_\delta\|_{2,p,U_{\varepsilon,r}}$, $\|\widehat{Z}_\delta\|_{2,p,U_{\varepsilon,r}}$, and $\|\widehat{X}_\delta - \widehat{Z}_\delta\|_{2,p,U_{\varepsilon,r}} = \|\alpha^{-1} R_\delta\|_{2,p,U_{\varepsilon,r}}$. First of all, by using Assumption 2.1, one easily gets that there exists $C \in \mathcal{C}_0$ such that

$$\big\|\alpha^{-1} R_\delta\big\|_{2,p} \le C\delta^{-1}\delta^{3/2} = C\sqrt{\delta} \quad \text{and} \quad \|\widehat{X}_\delta\|_{2,p} + \|\widehat{Z}_\delta\|_{2,p} \le C.$$

We now check that $m_{U_{\varepsilon,r}}(p) < \infty$ for every p. Standard computations and (C.2) give, for every p,

$$\|1/\det Q\|_{2,p} + \||Q|_l\|_{2,p} + \big\|q(B)\big\|_{2,p} + \|\Theta\|_{2,p} \le C,$$

so we can apply (E.3) and conclude

$$m_{U_{\varepsilon,r}}(p) \le C \in \mathcal{C}_0.$$

We now study the lower eigenvalue of the Malliavin covariance matrix of $\widehat{Z}_\delta$. From the definition of $\widehat{Z}_\delta$, we have

$$\widetilde{Z} = \mathcal{V}\begin{pmatrix} \alpha^{-1} Z_\delta \\ \underline{\mathcal{U}}^T \underline{\Gamma}_\delta \Theta \end{pmatrix} = \mathcal{V}\begin{pmatrix} \widehat{Z}_\delta \\ \underline{\mathcal{U}}^T \underline{\Gamma}_\delta \Theta \end{pmatrix}, \tag{3.28}$$

(see the proof of Lemma 3.6 for the definition of $\underline{\Gamma}_\delta$). As an immediate consequence, one has $\lambda_*(\gamma_{\widehat{Z}_\delta}) \ge \lambda_*(\gamma_{\widetilde{Z}})$, and it suffices to study the lower eigenvalue of the Malliavin covariance matrix of $\widetilde{Z}$. By using (3.13), we have

$$\begin{aligned}
\langle \gamma_{\widetilde{Z}} \xi, \xi\rangle &= \sum_{i=1}^d \int_0^\delta \langle D_s^i \widetilde{Z}, \xi\rangle^2 = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i \widetilde{Z}, \xi\rangle^2 = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i \big(\Theta + \widetilde{\eta}(\Theta)\big), \xi\rangle^2 \\
&\ge \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \left(\frac{1}{2} \langle D_s^i \Theta, \xi\rangle^2 - \langle D_s^i \eta(\Theta), \xi\rangle^2 \right) ds \\
&= S_1 + S_2.
\end{aligned}$$

We write

$$\begin{aligned}
S_1 &= \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \frac{1}{2} \langle D_s^i \Theta, \xi\rangle^2 \ge \frac{\lambda_*(Q)}{2}, \\
S_2 &= \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle \nabla\eta(\Theta) D_s^i \Theta, \xi\rangle^2 ds = \sum_{i=1}^d \int_{s_{i-1}(\delta)}^{s_i(\delta)} \langle D_s^i \Theta, \nabla\eta(\Theta)^T \xi\rangle^2 ds \le \lambda^*(Q) \big|\nabla\eta(\Theta)\big|^2 |\xi|^2,
\end{aligned}$$

so that

$$\lambda_*(\gamma_{\widehat{Z}_\delta}) \ge \lambda_*(\gamma_{\widetilde{Z}}) \ge \lambda_*(Q) \left(\frac{1}{2} - \frac{\lambda^*(Q)}{\lambda_*(Q)} \big|\nabla\eta(\Theta)\big|^2 \right).$$

On $\{\widetilde{U}_\varepsilon \ne 0\}$, we have already proved in Lemma 3.5 that $c_*(\eta, \Theta) \le \frac{\sqrt{\lambda_*(Q)/\lambda^*(Q)}}{2m}$. Since $|\nabla\eta(\Theta)| \le m c_*(\eta, \Theta)$, we obtain

$$\big|\nabla\eta(\Theta)\big| \le \frac{1}{2}\sqrt{\frac{\lambda_*(Q)}{\lambda^*(Q)}},$$

and therefore $\lambda_*(\gamma_{\widehat{Z}_\delta}) \ge 4\lambda_*(\gamma_{\widetilde{Z}}) \ge \lambda_*(Q) \ge \varepsilon^{3m\rho}$. Writing now $\mathbb{E}_{U_{\varepsilon,r}}$ for the integral wrt $d\mathbb{P}_{U_{\varepsilon,r}} = U_{\varepsilon,r}\, d\mathbb{P}$, we have that $\mathbb{E}_{U_{\varepsilon,r}}(\lambda_*(\widehat{Z}_\delta)^{-p}) < \infty$ for every p.

Let us study the lowest eigenvalue of $\gamma_{\widehat{X}_\delta}$. We use here some results from next Section 4, namely Lemma 4.2. There, we actually prove the desired bound for the Malliavin covariance matrix of $\alpha^{-1}(X_\delta - x_0)$. Here we are considering $\widehat{X}_\delta = \alpha^{-1}(X_\delta - x_0 - b(0, x_0)\delta)$, but their Malliavin covariance matrix is the clearly the same. Then, Lemma 4.2 gives that $\mathbb{E}(\lambda_*(\gamma_{\widehat{X}_\delta})^{-p}) < \infty$ for every p.

So, we have proved that all the requests in Proposition E.1 hold. Then, we can apply (E.4) and we get

$$p_{\widehat{X}_\delta, U_{\varepsilon,r}}(z) \ge p_{\widehat{Z}_\delta, U_{\varepsilon,r}}(z) - C'\sqrt{\delta}$$

with $C' \in \mathcal{C}_0$. Now, from (3.28) and (3.21), with a simple change of variables, we get

$$p_{\widehat{Z}_\delta, U_{\varepsilon,r}}(z) \ge \frac{1}{C}, \quad \text{for } |z| \le \frac{r}{2}.$$

We can assert the existence of $\delta_* \in 1/\mathcal{C}_0$ and $C \in \mathcal{C}_0$ such that for all $\delta \leq \delta_*$

$$p_{\widehat{X}_\delta, U_{\varepsilon,r}}(z) \geq \frac{1}{2C},$$

and the statement follows. $\square$

We are now ready for the proof of the lower bound:

Theorem 3.8. *Let Assumption 2.1 and 2.2 hold. Let p_{X_t} denote the density of X_t, $t > 0$. Then there exist positive constants r, δ_*, C such that for every $\delta \leq \delta_*$ and for every y such that $|y - x_0 - b(0, x_0)\delta|_{A_\delta(0,x_0)} \leq r$,*

$$p_{X_\delta}(y) \geq \frac{1}{C\sqrt{\det A_\delta A_\delta^T(0, x_0)}},$$

Here, $C \in \mathcal{C}_0$ and $r, \delta_* \in 1/\mathcal{C}_0$.

Proof. We take the same δ_*, r as in Lemma 3.7 and let $\widehat{X}_\delta$ denotes the r.v. handled in Lemma 3.7. By construction, we have $X_\delta = x_0 + b(0, x_0) + \alpha\widehat{X}_\delta$, so by applying Lemma 3.7 we get

$$\begin{aligned}
\mathbb{E}\big(f(X_\delta)\big) &\geq \mathbb{E}_{U_{\varepsilon,r}}\big(f(X_\delta)\big) = \mathbb{E}_{U_{\varepsilon,r}}\big(f\big(x_0 + b(0, x_0)\delta + \alpha\widehat{X}_\delta\big)\big) \\
&= \int f\big(x_0 + b(0, x_0)\delta + \alpha z\big) p_{\widehat{X}_\delta, U_{\varepsilon,r}}(z)\, dz \\
&\geq \frac{1}{C}\int_{\{|z| \leq r/2\}} f\big(x_0 + b(0, x_0)\delta + \alpha z\big)\, dz \\
&\geq \frac{1}{C|\det\alpha|}\int_{|y|_\alpha \leq r/2} f\big(x_0 + b(0, x_0)\delta + y\big)\, dy.
\end{aligned}$$

From (2.8) we obtain

$$\sqrt{\det A_\delta A_\delta^T(0, x_0)} = \big|\det(\alpha)\big| \tag{3.29}$$

and the statement follows. $\square$

Remark 3.9. We observe that if the diffusion coefficients are bounded, that is Assumption 2.3 holds, then the class $\mathcal{C}_0$ in (3.20) of the constants can be replaced by

$$\mathcal{L}_0 = \left\{C > 0 : \text{ there are } K, q > 0 \text{ such that } C = K\left(\frac{\kappa}{\lambda(0, x_0)}\right)^q\right\}$$

and, as before, $1/\mathcal{L}_0 = \{C > 0 : 1/C \in \mathcal{L}_0\}$. This follows from the fact that in the estimates for $\|\widehat{X}_\delta - \widehat{Z}_\delta\|_{2,p}$ and $\|\widehat{X}_\delta\|_{2,p}$ one does not need anymore to use the Gronwall's Lemma but it suffices to use the boundedness of the coefficients and the Burkholder inequality. A time dependent version of this class is defined in (5.3), to be used in the concatenation along the tube.

4. Upper bound for the density

We study here the upper bound for the density of X_δ. As for the lower bound, we scale X_δ. We recall the results and the notation in Lemma 3.6 and we define the change of variable

$$T_\alpha : \mathbb{R}^n \to \mathbb{R}^n, \quad T_\alpha(y) = \alpha^{-1}y, \quad \text{where } \alpha = \mathcal{U}\bar{\Sigma} \tag{4.1}$$

and its adjoint $T_\alpha^*(v) = \alpha^{-1,T} v$. Note that α is a $n \times n$ matrix. We write $A_{\delta,j}$, for $j = 1, \dots, m$, for the columns of A_δ (which can be $\sqrt{\delta}\sigma_i$ or $\delta[\sigma_i, \sigma_p]$). The following properties hold:

Lemma 4.1. *Let T_α be defined in* (4.1). *Then one has*:

$$|y|_{A_\delta} = |T_\alpha y| = |y|_\alpha, \quad \forall y \in \mathbb{R}^n, \quad \textit{and} \quad \det \alpha = \sqrt{\det A_\delta A_\delta^T} \tag{4.2}$$

$$\forall v \in \mathbb{R}^n \quad \textit{with } |v| = 1, \exists j = 1, \dots, m : \left|T_\alpha^* v \cdot A_{\delta,j}\right| \geq \frac{1}{m} \tag{4.3}$$

$$\forall j = 1, \dots, d, \quad \sqrt{\delta}|T_\alpha \sigma_j| \leq 1 \tag{4.4}$$

Proof. (4.2) follows easily from $\alpha = \mathcal{U}\bar{\Sigma}$ and the definition (2.2) of $|\cdot|_M$. Now, $(T_\alpha^* v)^T A_\delta = v^T \alpha^{-1} A_\delta = [v^T 0]\mathcal{V}^T$. So $|(T_\alpha^* v)^T A_\delta| = |[v^T 0]\mathcal{V}^T| = 1$. Recall that $A_{\delta,j}$ are the columns of A_δ, therefore $\exists j = 1, \dots, m : |(T_\alpha^* v)^T A_{\delta,j}| \geq \frac{1}{m}$, which is equivalent to (4.3). Moreover, $T_\alpha A_\delta = [\mathrm{Id}_n 0]\mathcal{V}^T$. This easily implies that $\forall i = 1, \dots, m$, $|T_\alpha A_{\delta,i}| \leq 1$. For $A_{\delta,i} = \sigma_j(0, x_0)\sqrt{\delta}$ we have (4.4). □

We define now

$$F = \alpha^{-1}(X_\delta - x_0) = T_\alpha(X_\delta - x_0). \tag{4.5}$$

As for the lower bound, we first estimate the density of F, using the results in Appendix E (specifically, (E.5) in Proposition E.1), and then recover the estimates for the density of X_δ via a change of variable. Estimate (E.5) involves the inverse moments of the smallest singular value of the Malliavin covariance matrix of F. The boundedness of the inverse moments of the Malliavin covariance matrix of the hypoelliptic diffusion X_δ is a classic result in Malliavin calculus (see for example [26] [Section 2.3.3]). In this paper, to obtain the desired bound for the density wrt the matrix norm defined via (2.2) and (2.8), we need an analogous bound which keeps track of the different time scales of propagation in the direction of the diffusion vector fields and of their first order Lie brackets. For this reason, we have defined in (4.5) the rescaled diffusion F, whose Malliavin covariance matrix is non-degenerate *uniformly in time*. More precisely, in the following lemma, which is a refinement of the classic non-degeneracy result mentioned above, we upper bound the L^p norm of the inverse of the Malliavin covariance matrix of F by a constant in $\mathcal{C}_0$, $\mathcal{C}_0$ being defined in (3.20). A key fact here is that such constant does not depend on δ.

Lemma 4.2. *Let α, T_α and $F = T_\alpha(X_\delta - x_0)$ be defined as in* (4.1) *and* (4.5). *γ_F denotes the Malliavin covariance matrix of F. Then for any $p > 1$ there exists $C \in \mathcal{C}_0$ such that, for $\delta \leq 1$, $\mathbb{E}|\lambda_*(\gamma_F)|^{-p} \leq C$.*

Proof. The proof is a modification of [26] [Theorem 2.3.3], where at each stage it has to be checked that considering the rescaled diffusion F in (4.5) instead of X_δ allows for bounds not depending on δ. All the details of the proof can be found in the preprint [4], which was replaced by the present article. □

The above lemma allow us to prove the following upper bound for the density.

Theorem 4.3. *Let Assumption 2.1 and 2.2 hold. Let p_{X_t} denote the density of X_t, $t > 0$. Then, for any $p > 1$, there exists a positive constant $C \in \mathcal{C}_0$ such that for every $\delta \leq 1$ and for every $y \in \mathbb{R}^n$*

$$p_{X_\delta}(y) \leq \frac{1}{\sqrt{\det A_\delta A_\delta^T(0, x_0)}} \frac{C}{1 + |y - x_0|^p_{A_\delta(0,x_0)}}.$$

Proof. Set $F = T_\alpha(X_\delta - x_0)$. We apply estimate (E.5): there exist constants p and a depending only on the dimension n, such that

$$p_F(z) \leq C \max\left\{1, \mathbb{E}\left|\lambda_*(\gamma_F)\right|^{-p} \|F\|_{2,p}\right\} \mathbb{P}\left(|F - z| < 2\right)^a.$$

We first show that $\|F\|_{2,p} \leq C \in \mathcal{C}_0$, as a consequence of Assumption 2.1. We prove just that $\|F\|_p \leq C$ for every p, for the Malliavin derivatives the proof is heavier but analogous. We write

$$F = T_\alpha\left(\sum_{j=1}^{d}\int_0^\delta \sigma_j(t, X_t) \circ dW_t^j + \int_0^\delta b(t, X_t)\,dt\right) = T_\alpha\left(\sum_{j=1}^{d}\sigma_j(0, x_0)W_\delta^j + B_\delta\right),$$

where

$$B_\delta = \sum_{j=1}^{d}\int_0^\delta \bigl(\sigma_j(t, X_t) - \sigma_j(0, x_0)\bigr) \circ dW_t^j + \int_0^\delta b(t, X_t)\,dt.$$

Therefore

$$|F| \leq \sum_{j=1}^{d}\bigl|T_\alpha\sigma_j(0, x_0)W_\delta^j\bigr| + |T_\alpha B_\delta|. \tag{4.6}$$

(4.4) implies $|T_\alpha\sigma_j(0, x_0)W_\delta^j| \leq CW_\delta^j/\sqrt{\delta}$, for $j = 1, \dots, d$. Moreover $|T_\alpha B_\delta| \leq |B_\delta|_{A_\delta} \leq C|B_\delta|/\delta$. If assumption 2.1 holds we conclude that $\mathbb{E}|F|^p \leq C \in \mathcal{C}_0$.

As in [3], Remark 2.4, it is easy to reduce the estimate of $\mathbb{P}(|F - z| < 2)$ to the tail estimate of F, and then to use Markov inequality to relate the estimate of the tails to the moments of F:

$$\mathbb{P}\bigl(|F - z| < 2\bigr) \leq \mathbb{P}\bigl(|F| > |z|/2\bigr) \leq C\frac{1 \vee \mathbb{E}|F|^p}{1 + |z|^p}, \quad \forall z \in \mathbb{R}^n. \tag{4.7}$$

Since, from Assumption 2.1, all the moments of F are bounded by constants in $\mathcal{C}_0$, we have that for any exponent $p > 1$ this term decays faster than $|z|^{-p}$ for $|z| \to \infty$.

In Lemma 4.2 we have already proved that $\mathbb{E}|\lambda_*(\gamma_F)|^{-q} \leq C \in \mathcal{C}_0$, for $\delta \leq 1$. We conclude that $p_F(z) \leq \frac{C}{1+|z|^p}$. The upper bound for the density of X_δ comes from the simple change of variable $y = x_0 + \alpha z$. For a positive and bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$, we write

$$\mathbb{E}f(X_\delta) = \mathbb{E}f(x_0 + \alpha F) = \int f(x_0 + \alpha z)p_F(z)\,dz$$

and we apply our density estimate, so that

$$\mathbb{E}f(X_\delta) \leq \int \frac{Cf(x_0 + \alpha z)}{1 + |z|^p}\,dz \leq \frac{C}{|\det\alpha|}\int \frac{f(y)}{1 + |x_0 - y|^p_{A_\delta(0,x_0)}}\,dy,$$

in which we have used (4.2). We use again (3.29) and the statement follows. □

Remark 4.4. If Assumption 2.3 holds then the upper estimate in Theorem 4.3 is of exponential type: there exists a constant $C \in \mathcal{C}_0$ such that for every $\delta \leq 1$ and for every $y \in \mathbb{R}^n$

$$p_{X_\delta}(y) \leq \frac{C}{\sqrt{\det A_\delta A_\delta^T(0, x_0)}}\exp\left(-\frac{1}{C}|y - x_0|_{A_\delta(0,x_0)}\right).$$

The proof is identical to the previous one except for the last part. In fact, looking at (4.6), in this case the boundedness of the coefficients allows one to apply the exponential martingale inequality, so instead of (4.7) we obtain the exponential bound $\mathbb{P}(|F| > |y|/2) \leq C\exp(-|y|/C)$. This gives the proof of (3) in Theorem 2.4.

Remark 4.5. In Theorem 3.8 the lower bound is centered at $x_0 + \delta b(x_0)$ but for the upper estimate in Theorem 4.3, one can choose to center at x_0 or at $x_0 + \delta b(x_0)$. In fact, in this case we notice that

$$\left|\delta b(x_0)\right|_{A_\delta(0,x_0)} \le \frac{C'}{\delta}\left|\delta b(x_0)\right| \le C'',$$

so

$$\begin{aligned}\frac{C_1}{1+|x_0-y|_{A_\delta(0,x_0)}} &\le \frac{C_2}{1+|x_0+\delta b(x_0)-y|_{A_\delta(0,x_0)}}\\ &\le \frac{C_3}{1+|x_0-y|_{A_\delta(0,x_0)}},\end{aligned}$$

and the estimate of Theorem 4.3 can be equivalently written as

$$p_{X_\delta}(y) \le \frac{1}{\sqrt{\det A_\delta A_\delta^T(0,x_0)}}\frac{C}{1+|y-x_0-\delta b(x_0)|^p_{A_\delta(0,x_0)}}.$$

5. Tube estimates

The proof of Theorem 2.9 is inspired by the approach in [7]. A similar procedure is also used in [27] in a weak Hörmander framework. Such a proof strongly uses the estimates for the density developed in Sections 3 and 4 and it is crucial that these estimates hold in a time interval of a fixed small length. The proof consists in a "concatenation" of such estimates in order to recover the whole time interval $[0,T]$. Since the "concatenation" works around the skeleton path $x(\phi)$, it suffices that the properties of all objects hold only locally around $x(\phi)$, as required in (2.12). In order to set-up this program, we need the precise behavior of the norm $|\cdot|_{A_R}$. So, we first present the desired properties for $|\cdot|_{A_R}$ (Section 5.1) and then we proceed with the proof of Theorem 2.9 (Section 5.2).

5.1. *Matrix norms*

Recall the definitions (2.5) and (2.8) for $A(t,x)$ and $A_R(t,x)$ respectively. We work with the norm $|y|^2_{A_R(t,x)} = \langle (A_R A_R^T(t,x))^{-1}y, y\rangle$, $y\in\mathbb{R}^n$.

Lemma 5.1. *Let $x\in\mathbb{R}^n$, $t\ge 0$, $R>0$ and recall that $\lambda^*(A(t,x))$ and $\lambda_*(A(t,x))$ denote the largest and lowest singular value of $A(t,x)$.*

(i) *For every $y\in\mathbb{R}^n$ and $0<R\le R'\le 1$*

$$\sqrt{\frac{R}{R'}}|y|_{A_R(t,x)} \ge |y|_{A_{R'}(t,x)} \ge \frac{R}{R'}|y|_{A_R(t,x)} \tag{5.1}$$

(ii) *For every $z\in\mathbb{R}^m$ and $R>0$*

$$\left|A_R(t,x)z\right|_{A_R(t,x)} \le |z|. \tag{5.2}$$

(iii) *For every $\varphi\in L^2([0,T];\mathbb{R}^m)$,*

$$\left|\int_0^r \varphi_s\,ds\right|^2_{A_R(t,x)} \le r\int_0^r |\varphi_s|^2_{A_R(t,x)}\,ds, \quad r\in[0,T].$$

Proof. For fixed $x\in\mathbb{R}^n$ and $t\ge 0$, during the proof we omit in $A(t,x)$ and $A_R(t,x)$ the dependence on (t,x), so we simply write A and A_R

(i) For $0 < R \le R' \le 1$, 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 (5.1).

(ii) For $z \in \mathbb{R}^m$, we write $z = A_R^T y + w$ with $y \in \mathbb{R}^n$ and $w \in (\mathrm{Im} A_R^T)^\perp = \mathrm{Ker}\, A_R$. Then $A_R z = A_R A_R^T y$ so that

$$\begin{aligned}
|A_R z|_{A_R}^2 &= \left|A_R A_R^T y\right|_{A_R}^2 = \left\langle \left(A_R A_R^T\right)^{-1} A_R A_R^T y, A_R A_R^T y\right\rangle \\
&= \left\langle y, A_R A_R^T y\right\rangle = \left\langle A_R^T y, A_R^T y\right\rangle = \left|A_R^T y\right|^2 \le |z|^2
\end{aligned}$$

and (5.2) holds.

(iii) For $\varphi \in L^2([0,T];\mathbb{R}^m)$ and $r \in [0,T]$,

$$\begin{aligned}
\left|\int_0^r \varphi_s\, ds\right|_{A_R}^2 &= \left\langle \left(A_R A_R^T\right)^{-1} \int_0^r \varphi_s\, ds, \int_0^r \varphi_s\, ds\right\rangle = \int_0^r \int_0^r \left\langle \left(A_R A_R^T\right)^{-1} \varphi_s, \varphi_u\right\rangle ds\, du \\
&= \frac{1}{2} \int_0^r \int_0^r \left\langle \left(A_R A_R^T\right)^{-1} (\varphi_s - \varphi_u), \varphi_s - \varphi_u\right\rangle ds\, du \\
&\quad - \int_0^r \int_0^r \left(\left\langle \left(A_R A_R^T\right)^{-1} \varphi_s, \varphi_s\right\rangle - \left\langle \left(A_R A_R^T\right)^{-1} \varphi_u, \varphi_u\right\rangle\right) ds\, du \\
&= \frac{1}{2} \int_0^r \int_0^r \left(|\varphi_s - \varphi_u|_{A_R}^2 - 2|\varphi_s|_{A_R}^2\right) ds\, du \\
&\le \int_0^r \int_0^r |\varphi_u|_{A_R}^2\, ds\, du = r \int_0^r |\varphi_u|_{A_R}^2\, du.
\end{aligned}$$

□

Next Lemma 5.2 is strictly connected to Remark 2.10, where we stressed that our result allows for a regime switch along the tube. In fact, here we fix $R > 0$, two points (t,x) and (s,y) and we get an equivalence between the norms $|\cdot|_{A_R(t,x)}$ and $|\cdot|_{A_R(s,y)}$ without supposing that in these two points the Hörmander condition holds "under the same regime". To compensate this lack of uniformity, we suppose that the distance between (t,x) and (s,y) is bounded by $\sqrt{R}$, and we will need to take this fact into account. In the concatenation procedure of next Section 5.2, the size of the intervals, to which we apply our density estimates, will have to depend on the radius of the tube.

We set

$$O = \left\{(t,x) \in [0,T] \times \mathbb{R} : \lambda(t,x) > 0\right\}$$

which is open, and under (2.11), we set (recall that λ is defined in (2.6) and n in (2.11))

$$\mathcal{L} = \left\{C : O \to \mathbb{R}_+ : \text{ there are } K, q > 0 \text{ such that } C(t,x) = K\left(\frac{n(t,x)}{\lambda(t,x)}\right)^q\right\}. \tag{5.3}$$

We also define

$$1/\mathcal{L} = \{c : O \to \mathbb{R}_+ \text{ such that } 1/c \in \mathcal{L}\}.$$

Notice that this is the time dependent version of the class of constants in Remark 3.9.

Lemma 5.2. *Assume* (2.11) *and let* $\mathcal{L}$ *as in* (5.3). *There exists* $C^* \in \mathcal{L}$ *such that for every* $(t,x), (s,y) \in O$ *and* $R \in (0,1]$ *satisfying*

$$|x - y| + |t - s| \le \sqrt{R}/C^*(t,x), \tag{5.4}$$

then for every $z \in \mathbb{R}^n$ one has

$$\frac{1}{4}|z|^2_{A_R(t,x)} \le |z|^2_{A_R(s,y)} \le 4|z|^2_{A_R(t,x)}. \tag{5.5}$$

Proof. (5.5) is equivalent to

$$4\big(A_R A_R^T\big)(t,x) \ge \big(A_R A_R^T\big)(s,y) \ge \frac{1}{4}\big(A_R A_R^T\big)(t,x),$$

so we prove the above inequalities. Let $A_{R,k}$, $k = 1, \dots, m$, denote the columns of A_R. We use $(a+b)^2 \ge \frac{1}{2}a^2 - b^2$:

$$\begin{aligned}
\big\langle A_R A_R^T(s,y)z, z\big\rangle &= \sum_{k=1}^m \big\langle A_{R,k}(s,y), z\big\rangle^2 \\
&= \sum_{k=1}^m \big(\big\langle A_{R,k}(t,x), z\big\rangle + \big\langle A_{R,k}(s,y) - A_{R,k}(t,x), z\big\rangle\big)^2 \\
&\ge \frac{1}{2}\sum_{k=1}^m \big\langle A_{R,k}(t,x), z\big\rangle^2 - \sum_{k=1}^m \big\langle A_{R,k}(s,y) - A_{R,k}(t,x), z\big\rangle^2.
\end{aligned}$$

We use (2.11): for every (s,y) such that $|t-s| \le 1$ and $|x-y| \le 1$, we have

$$\big\langle A_R A_R^T(s,y)z, z\big\rangle \ge \frac{1}{2}\sum_{k=1}^m \big\langle A_{R,k}(t,x), z\big\rangle^2 - C_1 n(t,x)^\alpha R\big(|x-y|^2 + |t-s|^2\big)|z|^2,$$

in which $C_1 > 0$ and $\alpha \ge 1$ denote universal constants. Notice that

$$\sum_{k=1}^m \big\langle A_{R,k}(t,x), z\big\rangle^2 = \big\langle A_R A_R^T(t,x)z, z\big\rangle \ge \lambda_*^2\big(A_R(t,x)\big)|z|^2 \ge R^2\lambda_*^2\big(A(t,x)\big)|z|^2.$$

We choose the constants (K,q) characterizing $C^*(t,x)$ such that $K \ge 2\sqrt{C_1} \vee 1$ and $q \ge \alpha$. So, under (5.4) we obtain

$$C_1 n(t,x)^\alpha R\big(|x-y|^2 + |t-s|^2\big)|z|^2 \le \frac{1}{4}\sum_{k=1}^m \big\langle A_{R,k}(t,x), z\big\rangle^2$$

and

$$\big\langle\big(A_R A_R^T\big)(s,y)z, z\big\rangle \ge \frac{1}{4}\sum_{k=1}^m \big\langle A_{R,k}(t,x), z\big\rangle^2 = \frac{1}{4}\big\langle\big(A_R A_R^T\big)(t,x)z, z\big\rangle.$$

The converse inequality follows from analogous computations and inequality $(a+b)^2 \le 2a^2 + 2b^2$. □

We prove that moving along the skeleton associated to a control $\phi \in L^2([0,T], \mathbb{R}^d)$ for a small time δ, the trajectory remains close to the initial point in the A_δ-norm. To this purpose, we assume the conditions (H_1) and (H_2) in (2.12). Notice that these give $(t, x_t(\phi)) \in O$ for every t. Moreover, in such a case the set $\mathcal{L}$ can be replaced by the following class of functions:

$$\mathcal{A} = \left\{C : [0,T] \to \mathbb{R}_+ : \text{ there are } K, q > 0 \text{ such that } C_t = K\left(\frac{n_t}{\lambda_t}\right)^q\right\}, \tag{5.6}$$

n_t and λ_t being defined in (2.12). We also set

$$1/\mathcal{A} = \big\{c : [0,T] \to (0,1] : 1/c_t \in \mathcal{A}\big\}.$$

Lemma 5.3. *Let $x(\phi)$ be the skeleton path (2.10) associated to $\phi \in L^2([0,T],\mathbb{R}^d)$. Assume (H_1) and (H_2) in (2.12). Then there exists $\delta^*, \varepsilon^* \in 1/\mathcal{A}$ such that for every $t \in [0,T]$, $\delta_t \le \delta_t^*$, $\varepsilon_t(\delta_t) \le \varepsilon_t^*$, $s \in [0,\delta_t]$ with $t+s \le T$ and for every $z \in \mathbb{R}^n$ one has*

$$\frac{1}{4}|z|^2_{A_{\delta_t}(t,x_t(\phi))} \le |z|^2_{A_{\delta_t}(t+s,x_{t+s}(\phi))} \le 4|z|^2_{A_{\delta_t}(t,x_t(\phi))}. \tag{5.7}$$

Moreover, there exists $\bar{C} \in \mathcal{A}$ such that

$$\sup_{0\le s\le\delta_t} \left|x_{t+s}(\phi) - \left(x_t(\phi) + b\left(t,x_t(\phi)\right)s\right)\right|_{A_{\delta_t}(t,x_t(\phi))} \le \bar{C}_t\left(\varepsilon_t(\delta_t) \vee \sqrt{\delta_t}\right), \tag{5.8}$$

where

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

Proof.
Set $s_t = \inf\{s>0 : |x_{t+s}(\phi) - x_t(\phi)| \ge 1\}$. From (2.11) and (H_1) in (2.12), we have

$$1 = \left|x_{t+s_t}(\phi) - x_t(\phi)\right| \le n_t\left(s_t + \sqrt{s_t}\varepsilon_t(s_t)\right).$$

We take $\underline{C} \in \mathcal{A}$ such that $n_t(\sqrt{s_t} + \varepsilon_t(s_t)) \le \underline{C}_t^{1/2}$, so that $s_t \ge 1/\underline{C}_t$. Take now $\delta^* \in 1/\mathcal{A}$ such that $\delta^* \le 1/\underline{C}$. Then if $s \le \delta_t \le \delta_t^*$, one has $s \le s_t$ and again from (2.11) and (H_1) in (2.12) we have

$$\left|x_{t+s}(\phi) - x_t(\phi)\right| + |s| \le \sqrt{\delta_t}\left(n_t\left(\sqrt{\delta_t^*} + \varepsilon_t(\delta_t)\right) + \sqrt{\delta_t^*}\right).$$

By continuity, for every $\varepsilon^* \in 1/\mathcal{A}$ and for every t there exists $\hat{\delta}_t$ such that $\varepsilon_t(\hat{\delta}_t) \le \varepsilon_t^*$. So, there actually exists $\delta_t \le \delta_t^*$ for which $\varepsilon_t(\delta_t) \le \varepsilon_t^*$. For such a δ_t, we have

$$\left|x_{t+s}(\phi) - x_t(\phi)\right| + |s| \le \sqrt{\delta_t}\left(n_t\left(\sqrt{\delta_t^*} + \varepsilon_t^*\right) + \sqrt{\delta_t^*}\right).$$

We now choose $\delta^*, \varepsilon^* \in 1/\mathcal{A}$ in order that the last factor in the above right hand side is smaller than $1/C^*(t,x_t(\phi))$, where $C^*(t,x)$ is the function in $\mathcal{L}$ for which Lemma 5.2 holds. Then (5.4) is satisfied with $R = \delta_t$, $x = x_t(\phi)$, $y = x_{t+s}(\phi)$ and s replaced by $t+s$. Hence (5.7) follows by applying (5.5).

We prove now (5.8). For the sake of simplicity, we let x_t denote the skeleton path $x_t(\phi)$. We write

$$\begin{aligned} J_{t,s} &:= x_{t+s} - x_t - b(t,x_t)s \\ &= \int_t^{t+s} \left(\dot{x}_u - b(u,x_u)\right)du + \int_t^{t+s}\left(b(u,x_u) - b(t,x_t)\right)du \\ &= \int_t^{t+s} \sigma(u,x_u)\phi_u\,du + \int_t^{t+s}\left(b(u,x_u) - b(t,x_t)\right)ds, \end{aligned}$$

so that

$$|J_{t,s}|^2_{A_{\delta_t}(t,x_t)} \le 2s\int_t^{t+s}\left|\sigma(u,x_u)\phi_u\right|^2_{A_{\delta_t}(t,x_t)}\,dt + 2s\int_t^{t+s}\left|b(u,x_u) - b(t,x_t)\right|^2_{A_{\delta_t}(t,x_t)}\,du.$$

In the above right hand side, we apply (5.7) to the norm in the first term and we use (2.9) in the second one. We obtain:

$$\begin{aligned} |J_{t,s}|^2_{A_{\delta_t}(t,x_t)} &\le 2s\int_t^{t+s} 4\left|\sigma(u,x_u)\phi_u\right|^2_{A_{\delta_t}(u,x_u)}\,du + 2s\int_t^{t+s}\frac{1}{\delta_t^2\lambda_t^2}\left|b(u,x_u) - b(t,x_t)\right|^2 du \\ &\le 8s\int_t^{t+s}\left|\sigma(u,x_u)\phi_u\right|^2_{A_{\delta_t}(u,x_u)}\,du + 2\delta_t\int_t^{t+\delta_t}\frac{1}{\delta_t^2\lambda_t^2}\times n_t^2\left(|u-t| + |x_u - x_t|\right)^2 du. \end{aligned}$$

We have already proved that, for $u \in [t, t+s]$, $|u-t| + |x_u - x_t| \le \sqrt{\delta_t}/C_t^*$, with $C^* \in \mathcal{A}$, so

$$|J_{t,s}|^2_{A_{\delta_t}(t,x_t)} \le 8s \int_t^{t+s} \left|\sigma(u, x_u)\phi_u\right|^2_{A_\delta(u,x_u)} du + \bar{C}_t \delta_t,$$

with $\bar{C} \in \mathcal{A}$. It remains to study the first term in the above right hand side. For $i = 1, \dots, m$, we set $\psi^{(j-1)d+j} = \frac{1}{\sqrt{\delta_t}}\phi^j$ for $j = 1, \dots, d$, $\psi^i = 0$ otherwise. Then, recalling (2.8), we can write $\sigma(u, x_u)\phi_u = A_{\delta_t}(u, x_u(\phi))\psi_u$, so that, by (5.2),

$$\left|\sigma(u, x_u)\phi_u\right|^2_{A_{\delta_t}(u,x_u)} = \left|A_{\delta_t}(u, x_u)\psi_u\right|^2_{A_{\delta_t}(u,x_u)} \le |\psi_u|^2 = \frac{1}{\delta_t}|\phi_u|^2.$$

Hence, for $s \le \delta_t$, we finally have $|J_{t,s}|^2_{A_{\delta_t}(t,x_t)} \le 8\varepsilon_t(\delta_t)^2 + \bar{C}_t\delta_t$, and the statement follows. □

Remark 5.4. Let us finally discuss an inequality which will be used in next Appendix A. Fix $x \in \mathbb{R}^n$ and let $x(\phi)$ be the skeleton path (2.10) associated to $\phi \in L^2([0,T], \mathbb{R}^d)$ with starting condition $x_0(\phi) = x$. Assume simply (2.11) and recall $\mathcal{L}$ defined in (5.3). Then looking at the proof of Lemma (5.3), we have the following result: if $(0, x) \in O$, there exists $\overline{\delta}, \overline{\varepsilon} \in 1/\mathcal{L}$ and $\overline{C} \in \mathcal{L}$ such that if $\delta \le \overline{\delta}(0, x)$, $\varepsilon_0(\delta) \le \overline{\varepsilon}(0, x)$ and $s \in [0, \delta]$ then

$$\sup_{0\le s\le\delta} \left|x_s(\phi) - \big(x + b(0, x)s\big)\right|_{A_\delta(0,x)} \le \overline{C}(0, x)\big(\varepsilon_0(\delta) \vee \sqrt{\delta}\big). \tag{5.9}$$

5.2. *Proof of main theorem on tube estimates*

This section is organized as follows: the lower bound in Theorem 2.9 is proved in next Theorem 5.8, whereas the upper bound in Theorem 2.9 is studied in next Theorem 5.9. Since the proofs are long and technical we begin by giving the principal elements.

Remark 5.5 (Localization). Our aim is to estimate the probability of the set

$$\Lambda_T = \Big\{ \sup_{0\le t\le T} \left|X_t - x_t(\phi)\right|_{A_{R_t}(t,x_t(\phi))} \le 1 \Big\}$$

and our hypotheses (H_i), $i = 1, 2, 3$ are "local hypotheses", along a tube around the curve $x_t(\phi)$. Our first goal is to stress that, in order to deal with our problem, there is no loss of generality in taking "global" bounds on the coefficients. This, thanks to a localization procedure, that we explain here for the lower bound, which is the most complicated case. For the upper bound case, a similar localization hinges on the fact that $R_t \le R_t^*(\phi)$. Let us now focus on the lower bound case: since $n_t \ge 1$

$$\Lambda_T \supset \Lambda_T^* = \Big\{ \sup_{0\le t\le T} n_t^2 \left|X_t - x_t(\phi)\right|_{A_{R_t}(t,x_t(\phi))} \le 1 \Big\}$$

and we will estimate the probability of Λ_T^*. On this set we have, by (2.9),

$$\left|X_t - x_t(\phi)\right| \le \left|X_t - x_t(\phi)\right|_{A_{R_t}(t,x_t(\phi))} \sqrt{R_t}\lambda^*\big(t, x_t(\phi)\big) \le \frac{\lambda^*(t, x_t(\phi))}{n_t^2} \le 1,$$

so that $\Lambda_T^* \subset \{\sup_{0\le t\le T} |X_t - x_t(\phi)| \le 1\}$. Let $\tau = \inf\{t > 0 : |X_t - x_t(\phi)| \ge 1\}$. We have that $X_{t\wedge\tau}$ coincides with $\overline{X}_{t\wedge\tau}$, where $\overline{X}_t$ is the solution of the same SDE but with globally bounded coefficients $\overline{\sigma}$ and $\overline{b}$. More precisely we may assume the following hypothesis. Let $t \in (0, T)$. For every $y \in \mathbb{R}^n$ and $s \in (t, t+h)$ (recall that h is given in (H_3))

$$\sum_{0\le|\alpha|\le 4} \left|\partial_x^\alpha \overline{\sigma}(s, y)\right| + \left|\partial_s \partial_x^\alpha \overline{\sigma}(s, y)\right| + \left|\partial_x^\alpha \overline{b}(s, y)\right| + \left|\partial_s \partial_x^\alpha \overline{b}(s, y)\right| \le n_t. \tag{5.10}$$

The construction of the coefficients $\overline{\sigma}$ and $\overline{b}$ is done in a standard way: one takes $\overline{\sigma}(s,y)=\sigma(s,y)$ for $y\in B(x_s(\phi),1)$. So for such (s,y) one has the inequality $|\overline{\sigma}(s,y)|=|\sigma(s,y)|\le n(s,x_s(\phi))\le n_t$. Then one takes the extension for every $y\in\mathbb{R}^n$ which keeps this restriction. Similar restrictions are obtained for the derivatives. So, in the following we will assume that (5.10) holds true for our coefficients.

Remark 5.6 (Short time estimates). We assume the localization as in Remark 5.5. In particular, the coefficients b,σ_i, $i=1,\dots,d$, verify (5.10). So, Assumption 2.3 holds with $\kappa=n_t$. Let us fix $t\ge 0$ and $\delta>0$. We want to estimate the density $p(t,t+\delta,x,y)$ of the solution X at time $t+\delta$ with the starting condition $X_t=x$. We define the (open) set

$$O=\{(t,x):\lambda(t,x)>0\}.$$

We set

$$\mathcal{L}=\left\{C:O\to\mathbb{R}_+:\text{ there are } K,q>0 \text{ such that } C(t,x)=K\left(\frac{n_t}{\lambda(t,x)}\right)^q\right\}.$$

We are slightly abusing the notation here (compare with (5.3)) but this is not a problem because of (5.10). Upper and lower bounds proved in Theorem 3.8 and 4.3 can now be written for a general starting condition (t,x) in place of $(0,x_0)$: there exist $C\in\mathcal{L}$, $r^*,\delta^*\in 1/\mathcal{L}$ such that for $(t,x)\in O$, $\delta\le\delta^*(t,x)$ and for every y such that $|y-x-b(t,x)\delta|_{A_\delta(t,x)}\le r^*(t,x)$ one has

$$\frac{1}{C(t,x)\sqrt{\det A_\delta A_\delta^T(t,x)}}\le p(t,t+\delta,x,y)\le\frac{e^{C(t,x)}}{\sqrt{\det A_\delta A_\delta^T(t,x)}},\tag{5.11}$$

where $p(t,s,x,\cdot)$ denotes the density of the solution X at time s of the equation in (2.1) but with the starting condition $X_t=x$.

Remark 5.7 (Concatenation). The plan for the proof is the following. Consider first the lower bound for the tube (see next Theorem 5.8). For $\phi\in L^2[0,T]$, let $x(\phi)$ be the skeleton associated to (2.1) given in (2.10). We set a discretization $0=t_0<t_1<\cdots<t_N=T$ of the time interval $[0,T]$. Then, as k varies, we consider the events

$$D_k=\left\{\sup_{t_k\le t\le t_{k+1}} n_t^2\left|X_t-x_t(\phi)\right|_{A_{R_t}(t,x_t(\phi))}\le 1\right\}\quad\text{and}\quad\Gamma_k=\left\{y:\left|y-x_{t_k}(\phi)\right|_{A_{R_{t_k}}(t_k,x_{t_k}(\phi))}\le r_k\right\},$$

where $r_k<1$ is a radius that will be suitably defined in the sequel. We denote $\mathbb{P}_k$ the conditional probability

$$\mathbb{P}_k(\cdot)=\mathbb{P}(\cdot|W_t,t\le t_k;X_{t_k}\in\Gamma_k).$$

We will bound from below $\mathbb{P}(\sup_{t\le T}|X_t-x_t(\phi)|_{A_{R_t}(t,x_t(\phi))}\le 1)$ by computing the product of the probabilities $\mathbb{P}_k(D_k\cap\{X_{t_{k+1}}\in\Gamma_{k+1}\})$, and this computation uses the lower estimate of the densities given in (5.11). This estimate is applicable because at each step we condition to $X_{t_k}\in\Gamma_k$, and on this small set we control the eigenvalue $\lambda(t_k,X_k)$, and as a consequence we control the constant in the density estimate (5.11).

We recall the set $\mathcal{A}$ in (5.6) and $1/\mathcal{A}$ defined as usual.

Theorem 5.8. *Let $\mu\ge 1$, $h\in(0,1]$, $n:[0,T]\to[1,+\infty)$, $\lambda:[0,T]\to(0,1]$, $\phi\in L^2([0,T],\mathbb{R}^n)$ and $R:[0,T]\to(0,1]$ be such that (H_1)–(H_3) in (2.12) hold. Then there exist $\bar{K},\bar{q}>0$ such that*

$$\exp\left(-\int_0^T\bar{K}\left(\frac{\mu n_t}{\lambda_t}\right)^{\bar{q}}\left(\frac{1}{h}+\frac{1}{R_t}+|\phi_t|^2\,dt\right)\right)\le\mathbb{P}\left(\sup_{t\le T}\left|X_t-x_t(\phi)\right|_{A_{R_t}(t,x_t(\phi))}\le 1\right).\tag{5.12}$$

Proof. *Step* 1. We first set-up some quantities which will be used in the rest of the proof.

We recall (H_3): $R_\cdot, |\phi_\cdot|^2, n_\cdot, \lambda_\cdot \in L(\mu, h)$, where $f \in L(\mu, h)$ if and only if $f(t) \le \mu f(s)$ for $|t-s| \le h$. We set, for $q_1, K_1 > 1$ to be fixed in the sequel,

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

Then straightforward computations give that $f_R \in L(\mu^{2q_1+1}, h)$. We define

$$\delta(t) = \inf\left\{\delta > 0 : \int_t^{t+\delta} f_R(s)\,ds \ge \frac{1}{\mu^{2q_1+1}}\right\}. \tag{5.13}$$

We have

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

so $\delta(t) \le h$. We now prove that $\delta(\cdot) \in L(\mu^{4q_1+1}, h)$. In fact, if $0 < t - t' \le h$,

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

so $\delta(t') \le \mu^{4q_1+2}\delta(t)$. Since the converse holds as well, we get $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$. We now prove a further property for $\delta(\cdot)$: we have

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

so

$$\delta(t) \le \frac{1}{f_R(t)} \le \frac{R_t}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1} \le \frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1} \in 1/\mathcal{A} \tag{5.14}$$

(recall that $R_t, \lambda_t \le 1$ and $n_t \ge 1$ for every t). We also set the energy over the time interval $[t, t+\delta(t)]$:

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

Since $n, \lambda \in L(\mu, h)$ and $\delta(t) \le h$, for $s \in (t, t+\delta(t))$ we have

$$f_R(s) \ge K_1\left(\frac{\mu n_s}{\lambda_s}\right)^{q_1}|\phi_s|^2 \ge \frac{K_1}{\mu^{2q_1}}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}|\phi_s|^2.$$

Hence

$$\frac{1}{\mu^{2q_1+1}} = \int_t^{t+\delta(t)} f_R(s)\,ds \ge \frac{K_1}{\mu^{2q_1}}\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\int_t^{t+\delta(t)} |\phi_s|^2\,ds,$$

which gives that

$$\varepsilon_t\big(\delta(t)\big)^2 \le \frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1} \in 1/\mathcal{A}. \tag{5.15}$$

Step 2. We set now some notation and properties that will be used in the "concatenation", which is developed in the following steps.

We define the time grid as

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

and introduce the following notation on the grid:

$$\delta_k = \delta(t_k), \qquad \varepsilon_k = \varepsilon_{t_k}(\delta_k), \qquad n_k = n_{t_k}, \lambda_k = \lambda_{t_k}, \qquad X_k = X_{t_k}, \qquad x_k = x_{t_k}(\phi), \qquad R_k = R_{t_k}.$$

Recall that $\delta(t) < h$ for every t, so we have

$$R_k/\mu \le R_t \le \mu R_k, \quad \text{for } t_k \le t \le t_{k+1}.$$

We also define

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

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

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

Let $r^* \in 1/\mathcal{A}$ be the radius-function in Remark 5.6, associated to the points $(t, x_t(\phi))$ as $t \in [0, T]$. We set $r^*_k = r^*_{t_k}$.

Let us see some properties.

For all $t_k \le t \le t_{k+1}$, we have $R_t \ge R_k/\mu \ge \delta_k/\mu$ and, by using (5.1), we obtain

$$|\xi|_{A_{R_t}(t,x_t)} \le \sqrt{\frac{\delta_k}{R_k}} |\xi|_{A_{\delta_k/\mu}(t,x_t)} \le |\xi|_{A_{\delta_k/\mu}(t,x_t)},$$

last inequality holding because $\delta_k \le R_k$. Since $\delta_k/\mu \le \delta_k$, we apply again (5.1) to the norm in the right hand side above and we get

$$|\xi|_{A_{R_t}(t,x_t)} \le \mu |\xi|_{A_{\delta_k}(t,x_t)}. \tag{5.16}$$

Taking $\xi = x_t - \hat{x}_k(t)$, we have

$$\left|x_t - \hat{x}_k(t)\right|_{A_{R_t}(t,x_t)} \le \mu \left|x_t - \hat{x}_k(t)\right|_{A_{\delta_k}(t,x_t)}.$$

By (5.14) and (5.15), we can choose q_1, K_1 large enough such that $\delta(t) \le \delta^*(t)$, $\varepsilon_t(\delta(t)) \le \varepsilon^*(t)$ where $\delta^* \in 1/\mathcal{A}$ and $\varepsilon^* \in 1/\mathcal{A}$ are the functions in Lemma 5.3. So, we apply (5.7) to the norm in the above right hand side and we obtain

$$\left|x_t - \hat{x}_k(t)\right|_{A_{R_t}(t,x_t)} \le \mu \times 4 \left|x_t - \hat{x}_k(t)\right|_{A_{\delta_k}(t_k,x_k)}.$$

We use now (5.8): for some $\bar{C} \in \mathcal{A}$, we get

$$\left|x_t - \hat{x}_k(t)\right|_{A_{\delta_k}(t_k,x_k)} \le \bar{C}_k (\varepsilon_k \vee \sqrt{\delta_k}),$$

where $\bar{C}_k = \bar{C}_{t_k}$, and, as a consequence of the estimate above, we have also

$$\left|x_t - \hat{x}_k(t)\right|_{A_{R_t}(t,x_t)} \le 4\mu \bar{C}_k (\varepsilon_k \vee \sqrt{\delta_k}),$$

for all $t \in [t_k, t_{k+1}]$ and for all k. By recalling that $x_{t_{k+1}} - \hat{x}_k(t_{k+1}) = x_{k+1} - \hat{x}_k$, and possibly choosing K_1, q_1 larger, we can resume by asserting that $\delta_k \le \delta^*_{t_k}$ in Remark 5.6 with initial condition (t_k, x_k), and

$$|x_{k+1} - \hat{x}_k|_{A_{\delta_k}(t_k,x_k)} \le r^*_k/4 \quad \text{for all } k, \tag{5.17}$$

$$\left|\hat{x}_k(t) - x_t\right|_{A_{R_t}(t,x_t)} \le \frac{1}{4n_t^2} \quad \text{for all } t \in [t_k, t_{k+1}] \text{ and for all } k. \tag{5.18}$$

We have already noticed that, under our settings, (5.7) holds, so that

$$\frac{1}{2}|\xi|_{A_{\delta_k}(t_k,x_k)} \le |\xi|_{A_{\delta_k}(t_{k+1},x_{k+1})} \le 2|\xi|_{A_{\delta_k}(t_k,x_k)}.$$

Since $\delta(\cdot) \in L(\mu^{4q_1+2}, h)$, one has $\delta_k/\delta_{k+1} \le \mu^{4q_1+2}$ and $\delta_{k+1}/\delta_k \le \mu^{4q_1+2}$. So, using (5.1) to the right hand side of the above inequality we easily get

$$\frac{1}{2\mu^{2q_1+1}}|\xi|_{A_{\delta_k}(t_k,x_k)} \le |\xi|_{A_{\delta_{k+1}}(t_{k+1},x_{k+1})} \le 2\mu^{2q_1+1}|\xi|_{A_{\delta_k}(t_k,x_k)} \quad \text{for all } k. \tag{5.19}$$

Step 3. We are ready to set-up the concatenation for the lower bound.

We set, for K_2 and q_2 to be fixed in the sequel,

$$r_k = \frac{1}{K_2\mu^{2q_1+2q_2+1}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}. \tag{5.20}$$

Moreover, since $\lambda, n \in L(\mu, h)$ and $\delta_k \le h$, one easily gets $r_{k+1}/r_k \le \mu^{2q_2}$ for every k.

We define

$$\Gamma_k = \{y : |y - x_k|_{A_{\delta_k}(t_k,x_k)} \le r_k\} \quad \text{and} \quad \mathbb{P}_k(\cdot) = \mathbb{P}(\cdot|W_t, t \le t_k; X_k \in \Gamma_k),$$

that is, $\mathbb{P}_k$ is the conditional probability with respect to the knowledge of the Brownian motion up to time t_k and the fact that $X_k \in \Gamma_k$. The aim of this step is to prove that

$$\mathbb{P}_k(X_{k+1} \in \Gamma_{k+1}) \ge 2\mu^{-4nq_1}\exp\bigl(-K_3(\log\mu + \log n_k - \log\lambda_k)\bigr) \quad \text{for all } k \tag{5.21}$$

for some constant K_3 depending on K_1, K_2, q_1 and q_2.

We denote $\rho_k(X_k, y)$ the density of X_{k+1} with respect to this probability. We prove that

$$\Gamma_{k+1} \subset \{y : |y - \hat{X}_k|_{A_{\delta_k}(t_k,X_k)} \le r_k^*\}. \tag{5.22}$$

If (5.22) holds, as we will see, then we can apply the lower bound in Remark 5.6 to $\rho_k(X_k, y)$. We use here the estimate given in (5.11): there exists $\underline{C} \in \mathcal{A}$ such that

$$\rho_k(X_k, y) \ge \frac{1}{\underline{C}_k\sqrt{\det A_{\delta_k}A_{\delta_k}^T(t_k, X_k)}} \quad \text{for all } y \in \Gamma_{k+1}, \tag{5.23}$$

where $\underline{C}_k = \underline{C}_{t_k}$. Here, a direct application of (5.11) would give a constant which depends on $\lambda(t_k, X_k)$, but using the fact that $X_k \in \Gamma_k$ and so it is close to x_k, we can show that the same is true with $\lambda(t_k, x_k)$, and so the constant can be taken in $\mathcal{A}$. Let us show that (5.22) holds. We estimate

$$|y - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le |y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} + |x_{k+1} - \hat{x}_k|_{A_{\delta_k}(t_k,x_k)} + |\hat{x}_k - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)}$$

and by using (5.17) we obtain

$$|y - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le |y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} + \frac{r_k^*}{4} + |\hat{x}_k - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)}. \tag{5.24}$$

Using (5.19), the fact that $r_{k+1}/r_k \le \mu^{2q_2}$ and recalling that $|y - x_{k+1}|_{A_{\delta_{k+1}}(t_{k+1},x_{k+1})} \le r_{k+1}$, we obtain

$$\begin{aligned} |y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} &\le 2\mu^{2q_1+1}|y - x_{k+1}|_{A_{\delta_{k+1}}(t_{k+1},x_{k+1})} \le 2\mu^{2q_1+1}r_{k+1} \\ &\le 2\mu^{2q_1+2q_2+1}r_k \le \frac{2}{K_2}\left(\frac{\lambda_k}{n_k}\right)^{q_2}. \end{aligned}$$

(2.11) also gives $|\hat{x}_k - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le C_k|x_k - X_k|_{A_{\delta_k}(t_k,x_k)}$, where $C_k = C_{t_k}$ and C is a suitable function in $\mathcal{A}$, and the conditioning with respect to Γ_k gives $|\hat{x}_k - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le C_k r_k$. Similarly, $|\hat{x}_k(t) - \hat{X}_k(t)|_{A_{R_t}(t,x_t)} \le C_k|x_k - X_k|_{A_{R_t}(t,x_t)}$ and by using firstly (5.16) and secondly (5.7), we get

$$\left|\hat{x}_k(t) - \hat{X}_k(t)\right|_{A_{R_t}(t,x_t)} \le C_k \times \mu|x_k - X_k|_{A_{\delta_k}(t,x_t)} \le C_k\mu \times 2|x_k - X_k|_{A_{\delta_k}(t_k,x_k)} \le 2\mu C_k r_k,$$

for every $t \in [t_k, t_{k+1}]$. Recalling (5.20), K_2 and q_2 (possibly large) such that $|y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} \le r_k^*/8$, $|\hat{x}_k - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le r_k^*/8$, and

$$\left|\hat{X}_k(t) - \hat{x}_k(t)\right|_{A_{R_t}(t,x_t)} \le \frac{1}{4n_t^2}, \quad \text{for all } t \in [t_k, t_{k+1}] \text{ and for all } k. \tag{5.25}$$

From (5.24), this implies $|y - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le r_k^*/2$. On the event Γ_k, we also have, from (2.9), $|x_k - X_k| \le |x_k - X_k|_{A_{\delta_k}(t_k,x_k)}\lambda^*(A(t_k,x_k))\sqrt{\delta_k} \le n_{t_k}^{\alpha}\sqrt{\delta_k}r_k$, for some universal constant $\alpha > 0$. So, we can fix K_2 and q_2 in order that Lemma 5.2 holds with $R = \delta_k$, $x = x_k$, $y = X_k$, $t = t_k$ and $s = 0$. Then, we get

$$\frac{1}{2}|\xi|_{A_{\delta_k}(t_k,x_k)} \le |\xi|_{A_{\delta_k}(t_k,X_k)} \le 2|\xi|_{A_{\delta_k}(t_k,x_k)}.$$

These inequalities give two consequences. First, we have

$$|y - \hat{X}_k|_{A_{\delta_k}(t_k,X_k)} \le 2|y - \hat{X}_k|_{A_{\delta_k}(t_k,x_k)} \le r_k^*,$$

so that (5.22) actually holds and then (5.23) holds as well. As a second consequence, we have that

$$\begin{aligned}\left\{y : |y - x_{k+1}|_{A_{\delta_k}(t_k,X_k)} \le \frac{r_{k+1}}{4\mu^{2q_1+1}}\right\} &\subset \left\{y : |y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} \le \frac{r_{k+1}}{2\mu^{2q_1+1}}\right\} \\ &\subset \left\{y : |y - x_{k+1}|_{A_{\delta_{k+1}}(t_{k+1},x_{k+1})} \le r_{k+1}\right\} = \Gamma_{k+1},\end{aligned}$$

in which we have used (5.19). Since $r_{k+1}/(4\mu^{2q_1+1}) \ge r_k/(4\mu^{2q_1+2q_2+1})$, we obtain

$$\Gamma_{k+1} \supset \left\{y : |y - x_{k+1}|_{A_{\delta_k}(t_k,X_k)} \le \frac{r_k}{4\mu^{2q_1+2q_2+1}}\right\}.$$

By recalling that $r_k/(4\mu^{2q_1+2q_2+1}) = \frac{1}{4K_2\mu^{4q_1+4q_2+2}}(\frac{\lambda_k}{n_k})^{q_2}$, we can write, with Leb_n denoting the Lebesgue measure in $\mathbb{R}^n$,

$$\mathrm{Leb}_n(\Gamma_{k+1}) \ge \sqrt{\det\left(A_{\delta_k}A_{\delta_k}^T(t_k, X_k)\right)}\left(\frac{1}{4K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^n.$$

So, from (5.23),

$$\mathbb{P}_k(X_{k+1} \in \Gamma_{k+1}) \ge \frac{1}{\underline{C}_k}\left(\frac{1}{4K_2\mu^{4q_1+4q_2+2}}\left(\frac{\lambda_k}{n_k}\right)^{q_2}\right)^n,$$

where $\underline{C}_k$ is the constant in (5.23). This implies (5.21), for some constant K_3 depending on K_2 and q_2.

Step 4. We give here the proof of the lower bound (5.12).

We set

$$D_k = \left\{\sup_{t_k \le t \le t_{k+1}} n_t^2|X_t - x_t|_{A_{R_t}(t,x_t)} \le 1\right\} \quad \text{and} \quad E_k = \left\{\sup_{t_k \le t \le t_{k+1}} n_t^2\left|X_t - \hat{X}_k(t)\right|_{A_{R_t}(t,x_t)} \le \frac{1}{2}\right\}.$$

For $t \in [t_k, t_{k+1}]$, by using (5.18) and (5.25) we have

$$
\begin{aligned}
|X_t - x_t|_{A_{R_t}(t,x_t)} &\le \big|X_t - \hat{X}_k(t)\big|_{A_{R_t}(t,x_t)} + \big|\hat{X}_k(t) - \hat{x}_k(t)\big|_{A_{R_t}(t,x_t)} + \big|\hat{x}_k(t) - x_t\big|_{A_{R_t}(t,x_t)} \\
&\le \big|X_t - \hat{X}_k(t)\big|_{A_{R_t}(t,x_t)} + \frac{1}{2n_t^2},
\end{aligned}
$$

so that $E_k \subset D_k$. Moreover, by passing from Stratonovich to Itô integrals and by using (2.9), we have

$$
\begin{aligned}
\big|X_t - \hat{X}_k(t)\big|_{A_{R_t}(t,x_t)} &\le \big|\sigma(t_k, X_{t_k})(W_t - W_{t_k})\big|_{A_{R_t}(t,x_t)} \\
&\quad + \left|\int_{t_k}^t \big(\sigma(s,X_s) - \sigma(t_k,X_k)\big)\,dW_s\right|_{A_{R_t}(t,x_t)} \\
&\quad + \left|\int_{t_k}^t \big(b(s,X_s) - b(t_k,X_k)\big)\,ds\right|_{A_{R_t}(t,x_t)} \\
&\quad + \sum_{l=1}^d \left|\int_{t_k}^t \nabla\sigma_l(s,X_s)\big(\sigma_l(s,X_s) - \sigma_l(t_k,X_k)\big)\,ds\right|_{A_{R_t}(t,x_t)} \\
&\le \left|\frac{\sqrt{\mu}}{\sqrt{R_k}}\sigma(t_k,X_{t_k})(W_t - W_{t_k})\right|_{A(t,x_t)} + \left|\frac{\mu}{R_k}\int_{t_k}^t \big(\sigma(s,X_s) - \sigma(t_k,X_k)\big)\,dW_s\right| \\
&\quad + \left|\frac{\mu}{R_k}\int_{t_k}^t \big(b(s,X_s) - b(t_k,X_k)\big)\,ds\right| \\
&\quad + \sum_{l=1}^d \left|\frac{\mu}{R_k}\int_{t_k}^t \frac{\nabla\sigma_l(s,X_s)}{2}\big(\sigma_l(s,X_s) - \sigma_l(t_k,X_k)\big)\,ds\right|.
\end{aligned}
$$

We use now the exponential martingale inequality (see also Remark 5.5) and we find that

$$
\mathbb{P}_k\big(E_k^c\big) \le \exp\left(-\frac{1}{K_4}\left(\frac{\lambda_k}{\mu n_k}\right)^{q_4}\frac{R_k}{\delta_k}\right)
$$

for some constants K_4, q_4. From (5.14), $R_k/\delta_k \ge K_1(\mu n_k/\lambda_k)^{q_1}$, so by choosing K_1 and q_1 possibly larger and by recalling (5.21), we can conclude that

$$
\mathbb{P}_k\big(E_k^c\big) \le \mu^{-4nq_1}\exp\big(-K_3(\log\mu + \log n_k - \log\lambda_k)\big) \le \frac{1}{2}\mathbb{P}_k(X_{k+1} \in \Gamma_{k+1}).
$$

Hence,

$$
\begin{aligned}
\mathbb{P}_k\big(\{X_{k+1} \in \Gamma_{k+1}\} \cap D_k\big) &\ge \mathbb{P}_k\big(\{X_{k+1} \in \Gamma_{k+1}\} \cap E_k\big) \ge \mathbb{P}_k(X_{k+1} \in \Gamma_{k+1}) - \mathbb{P}_k\big(E_k^c\big) \\
&\ge \frac{1}{2}\mathbb{P}_k(X_{k+1} \in \Gamma_{k+1}) \ge \exp\big(-K_5(\log\mu + \log n_k - \log\lambda_k)\big),
\end{aligned}
\tag{5.26}
$$

for some constant K_5. Let now $N(T) = \max\{k : t_k \le T\}$. From definition (5.13),

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

From (5.26),

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\le T}|X_t-x_t|_{A_{R_t}(t,x_t)}\le 1\Big) &\ge \mathbb{P}\left(\bigcap_{k=1}^{N(T)}\{X_{k+1}\in\Gamma_{k+1}\}\cap D_k\right)\\ &\ge \prod_{k=1}^{N(T)}\exp\big(-K_5(\log\mu+\log n_k-\log\lambda_k)\big)\\ &=\exp\left(-K_5\sum_{k=1}^{N(T)}(\log\mu+\log n_k-\log\lambda_k)\right).\end{aligned}$$

Since

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

the lower bound (5.12) follows. □

We can now address the problem of the upper bound. In this case we need to consider a sufficiently small radius $R_t\le R_t^*(\phi)$, such that the density estimates hold on the whole tube (cf. also with Remark 2.11). This allows us to use the same ideas that we explained in Remarks 5.5, 5.6, 5.7. In particular, Remark 5.5 holds in similar fashion, with a simpler justification, consequence of assumption $R_t\le R_t^*(\phi)$. Also, it is not necessary in this case to introduce the analogues of the sets D_k's, but only the analogues of the sets Γ_k's, that we call Δ_k's.

Similarly to (2.13), for $\mu\ge 1$, $h\in(0,1]$ and $K_*,q_*>0$, we denote

$$\bar R_t^*(\phi)=\exp\left(-K_*\left(\frac{\mu n_t}{\lambda_t}\right)^{q_*}\right)\left(\frac{1}{h}+|\phi|_t^2\right)^{-1} \tag{5.27}$$

Theorem 5.9. *Let $\mu\ge 1$, $h\in(0,1]$, $n:[0,T]\to[1,+\infty)$, $\lambda:[0,T]\to(0,1]$, $\phi\in L^2([0,T],\mathbb{R}^n)$ and $R:[0,T]\to(0,1]$ be such that (H_1)–(H_3) in (2.12) hold. Then there exist $K_*,q_*,\bar K,\bar q>0$ such that for $\bar R^*(\phi)$ as in (5.27), if $R_t\le\bar R_t^*(\phi)$ one has*

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\le T}\big|X_t-x_t(\phi)\big|_{A_{R_t}(t,x_t(\phi))}\le 1\Big)\\ &\quad\le\exp\left(-\int_0^T\bar K\left(\frac{\mu n_t}{\lambda_t}\right)^{\bar q}\left[\frac{\exp(-K_*(\frac{\mu n_t}{\lambda_t})^{3q_*})}{R_t}+\frac{1}{h}+|\phi_t|^2\right]dt\right).\end{aligned} \tag{5.28}$$

Proof. We refer here to notation and arguments already introduced and developed in the proof of Theorem 5.8. So, when we recall here Step 1, 2 and 3, we intend to refer to the same steps developed in the proof of Theorem 5.8.

We define, with the same K_1,q_1 as in Step 1,

$$g_R(t)=K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\left(\frac{\exp(-K_*(\frac{\mu n_t}{\lambda_t})^{q_*}\mu^{2q_*})}{R_t}+\frac{1}{h}+|\phi_t|^2\right).$$

We work here with $\delta(t)$ as in the proof od Theorem 5.8 but defined from g_R:

$$\delta(t)=\inf\left\{\delta>0:\int_t^{t+\delta}g_R(s)\,ds\ge\frac{1}{\mu^{2q_1+1}}\right\}.$$

We set, as before,

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

As in Step 1, we can check estimates similar to (5.14) and (5.15): we have indeed,

$$\delta(t)\le\frac{h}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}\le\frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}\quad\text{and}\quad\varepsilon_t\big(\delta(t)\big)^2\le\frac{1}{K_1}\left(\frac{\lambda_t}{\mu n_t}\right)^{q_1}.$$

In particular, $\delta(t)\le h$. With these definitions we set a time grid $\{t_k:k=0,\dots,N(T)\}$ and all the associated quantities as in Step 2. As we did for the lower bound, since we estimate the probability of remaining in the tube for any $t\in[t_k,t_{k+1}]$, we can suppose that the bound in Assumption 2.3 holds on $\mathbb{R}^+\times\mathbb{R}^n$ (recall Remark 5.5). The short time density estimate (5.11) holds again. Recall now that $R_\cdot\in L(\mu,h)$, and this gives the analogous to (5.19):

$$\frac{1}{2\sqrt{\mu}}|\xi|_{A_{R_k}(t_k,x_k)}\le|\xi|_{A_{R_{k+1}}(t_{k+1},x_{k+1})}\le 2\sqrt{\mu}|\xi|_{A_{R_k}(t_k,x_k)}. \tag{5.29}$$

We define

$$\Delta_k=\big\{y:|y-x_k|_{A_{R_k}(t_k,x_k)}\le 1\big\}\quad\text{and}\quad\tilde{\mathbb{P}}_k(\cdot)=\mathbb{P}(\cdot\,|W_t,t\le t_k;X_k\in\Delta_k),$$

so $\tilde{\mathbb{P}}_k$ is the conditional probability given the Brownian path up to time t_k and the fact that $X_k\in\Delta_k$.

Now, since $\delta(t)\le h$ and $R,\lambda,n,|\phi|^2\in L(\mu,h)$, recalling $R_t\le\bar{R}_t^*(\phi)$ in (5.27), we have

$$\begin{aligned}\int_t^{t+\delta(t)}K_1\left(\frac{\mu n_s}{\lambda_s}\right)^{q_1}\left(\frac{1}{h}+|\phi|_s^2\right)ds&\le\mu^{2q_1+1}K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\left(\frac{1}{h}+|\phi|_t^2\right)\delta(t)\\&\le\mu^{2q_1+1}K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\exp\left(-K_*\left(\frac{\mu n_t}{\lambda_t}\right)^{q_*}\right)\frac{\delta(t)}{R_t}\end{aligned}$$

and

$$\begin{aligned}&\int_t^{t+\delta(t)}K_1\left(\frac{\mu n_s}{\lambda_s}\right)^{q_1}\frac{\exp(-K_*(\frac{\mu n_s}{\lambda_s})^{q_*}\mu^{2q_*})}{R_s}\,ds\\&\quad\le\mu^{2q_1+1}K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\exp\left(-K_*\left(\frac{\mu n_t}{\lambda_t}\right)^{q_*}\right)\frac{\delta(t)}{R_t}.\end{aligned}$$

We obtain

$$1=\mu^{2q_1+1}\int_t^{t+\delta(t)}g_R(s)\,ds\le 2\mu^{4q_1+2}K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\exp\left(-K_*\left(\frac{\mu n_t}{\lambda_t}\right)^{q_*}\right)\frac{\delta(t)}{R_t}$$

so

$$\frac{R_t}{\delta(t)}\le 2\mu^{4q_1+2}K_1\left(\frac{\mu n_t}{\lambda_t}\right)^{q_1}\exp\left(-K_*\left(\frac{\mu n_t}{\lambda_t}\right)^{q_*}\right). \tag{5.30}$$

As we did in Step 3, if q_*,K_* are large enough, R_k is small enough and the upper bound for the density holds on Δ_{k+1}. By using inequality

$$\sqrt{\frac{\delta_k}{R_k}}\frac{1}{2\sqrt{\mu}}|\xi|_{A_{\delta_k}(t_k,x_k)}\le\frac{1}{2\sqrt{\mu}}|\xi|_{A_{R_k}(t_k,x_k)}\le|\xi|_{A_{R_{k+1}}(t_{k+1},x_{k+1})},$$

which can be proved using (5.29) and (5.1), one gets

$$\mathrm{Leb}_n\big(y : |y - x_{k+1}|_{A_{R_{k+1}}(t_{k+1},x_{k+1})} \le 1\big) \le \left(\frac{R_k}{\delta_k}\right)^{n/2} 2^n \mu^{n/2}\,\mathrm{Leb}_n\big(y : |y - x_{k+1}|_{A_{\delta_k}(t_k,x_k)} \le 1\big)$$
$$= \left(\frac{R_k}{\delta_k}\right)^{n/2} 2^n \mu^{n/2}\sqrt{\det\big(A_{\delta_k}A^T_{\delta_k}(t_k,x_k)\big)}.$$

Now, using the upper estimate for the density (5.11), we obtain

$$\tilde{\mathbb{P}}_k(X_{k+1} \in \Delta_{k+1}) \le e^{\overline{C}_k}\left(\frac{R_k}{\delta_k}\right)^{n/2},$$

where $\overline{C}_k = \overline{C}_{t_k}$, $\overline{C} \in \mathcal{A}$ (see the constant in the upper bound in (5.11); also here, a direct application of (5.11) would give a constant which depends on $\lambda(t_k, X_k)$, but using the fact that $X_k \in \Delta_k$ and so it is close to x_k, we can show that the constant can be taken in $\mathcal{A}$).

Recall (5.30), for $t = t_k$

$$\frac{R_k}{\delta_k} \le 2\mu^{4q_1+2}K_1\left(\frac{\mu n_k}{\lambda_k}\right)^{q_1}\exp\left(-K_*\left(\frac{\mu n_k}{\lambda_k}\right)^{q_*}\right)$$

so we chose now K_*, q_* large enough to have

$$\tilde{\mathbb{P}}_k(X_{k+1} \in \Delta_{k+1}) \le \exp(-K_2)$$

for a constant $K_2 > 0$. From the definition of $N(T)$

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

So, we have

$$\mathbb{P}\Big(\sup_{t\le T}\big|X_t - x_t(\phi)\big|_{A_{R_t}(t,x_t(\phi))} \le 1\Big) \le \mathbb{E}\left(\prod_{k=1}^{N(T)}\tilde{\mathbb{P}}_k(\Delta_{k+1})\right)$$
$$\le \prod_{k=1}^{N(T)}\exp(-K_2) = \exp\big(-K_2N(T)\big) \le \exp\left(-K_2\int_0^T g_R(t)\right)$$

and (5.28) holds as a consequence. □

Appendix A: On the equivalence between matrix norm and control distance

We establish here the local equivalence between the norm $|\cdot|_{A_R(t,x)}$ and the control (Carathéodory) distance d_c. We use in a crucial way the alternative characterization of d_c given in [25]. These results hold in the homogeneous case, so we consider now the vector fields $\sigma_j(t,x) = \sigma_j(x)$, and the associated norm $|\cdot|_{A_R(t,x)} = |\cdot|_{A_R(x)}$ (see (2.2) and (2.8)). We assume in this section the following bound on σ: there exists $\kappa : \mathbb{R}^n \to [1,+\infty)$ such that

$$\sup_{|y-x|\le 1}\sum_{0\le|\alpha|\le 4}\sum_{j=1}^{d}\big|\partial_x^\alpha\sigma_j(y)\big| \le \kappa(x), \quad \forall x \in \mathbb{R}^n. \tag{A.1}$$

So, (A.1) agrees with (2.11) in the homogeneous case and when $b = 0$.

We now introduce a quasi-distance d which is naturally associated to the family of norms $|y|_{A_R(x)}$. We define the open set

$$O = \{x \in \mathbb{R}^n : \lambda_*(A(x)) > 0\} = \{x : \det(AA^T(x)) \neq 0\}.$$

For $x, y \in O$, we define $d(x, y)$ by

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

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 $|y - x| \leq \sqrt{R}$. In the hypoelliptic case considered here, R appears as a radius, but we have ellipsoids instead of balls. It is straightforward to see that d is a quasi-distance on O, meaning that d verifies the following three properties (see [25]):

(i) for every $x \in O$ and $r > 0$, the set $\{y \in O : d(x, y) < r\}$ is open;
(ii) $d(x, y) = 0$ if and only if $x = y$;
(iii) for every compact set $K \Subset O$ there exists $C > 0$ such that for every $x, y, z \in K$ one has $d(x, y) \leq C(d(x, z) + d(z, y))$.

We recall the definition of equivalence of quasi-distances. Two quasi-distances $d_1 : D \times D \to \mathbb{R}^+$ and $d_2 : D \times D \to \mathbb{R}^+$ are equivalent if for every compact set $K \subset D$ there exists a constant C such that for every $x, y \in K$

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

d_1 and d_2 are locally equivalent if for every $\xi \in D$ there exists an open neighborhood V of ξ such that d_1 and d_2 are equivalent on V.

We introduce now the control metric. Without loss of generality, we assume $T = 1$,

For a control $\psi \in L^p([0, 1], \mathbb{R}^d)$, $p \in [1, \infty]$, let $u(\psi)$ satisfy the following controlled equation:

$$du_t(\psi) = \sum_{j=1}^{d} \sigma_j(u_t(\psi))\psi_t^j \, dt. \tag{A.2}$$

Notice that the equation for $u(\psi)$ is actually the skeleton equation (2.10) when the drift b is null.

For $x, y \in O$ and $\delta \in (0, 1]$, $p \in [1, \infty]$ we denote $C^p_{\sigma,\delta}(x, y)$ the set of controls $\psi \in L^p([0, \delta]; \mathbb{R}^d)$ such that the corresponding solution $u(\psi)$ to (A.2) satisfies $u_0(\psi) = x$ and $u_\delta(\psi) = y$. For $\psi \in C^2_{\sigma,\delta}(x, y)$, we set the associated energy

$$\varepsilon_\psi(\delta) = \left(\int_0^\delta |\psi_s|^2 \, ds\right)^{1/2}.$$

We also write $\|\psi\|_2 = \varepsilon_\psi(1)$ and define the control (Carathéodory) distance between $x, y \in \mathbb{R}^n$ as

$$d_c(x, y) = \inf_{\psi \in C^2_{\sigma,1}(x,y)} \|\psi\|_2.$$

Notice that for any fixed $\delta \in (0, 1]$,

$$d_c(x, y) = \sqrt{\delta} \inf_{\psi \in C^2_{\sigma,\delta}(x,y)} \varepsilon_\psi(\delta). \tag{A.3}$$

Indeed, for each $x, y \in \mathbb{R}^n$ and $\psi \in C^2_{\sigma,1}(x, y)$, take $\phi_t = \delta^{-1}\psi(t\delta^{-1})$ and $\xi_t = u_{t/\delta}(\psi)$. Then, $d\xi_t = \sum_{j=1}^d \sigma_j(\xi_t)\phi_t^j \, dt$ and of course $\xi_0 = x$, $\xi_\delta = y$. Moreover, $\|\psi\|_2 = \sqrt{\delta}\varepsilon_\phi(\delta)$.

Lastly, we consider the analogous distance defined via the sup-norm $\|\cdot\|_\infty$:

$$d_\infty(x, y) = \inf_{\psi \in C^\infty_{\sigma,1}(x,y)} \|\psi\|_\infty.$$

Under (A.1), we define

$$\mathcal{L}=\left\{C:O\to\mathbb{R}_+:C=K\left(\frac{\kappa(x)}{\lambda(x)}\right)^q,\exists K,q>0\right\}.$$

Notice that $\mathcal{L}$ is actually the set in (5.3) in the homogeneous case.

Theorem A.1. *Suppose that* (A.1) *hold. Then, the semi-distance d is locally equivalent to d_c on O. As a consequence, for every compact set $K\subset O$ there exist r_K and C_K such that for every $x,y\le r_K$ one has $d_c(x,y)\le C_K d(x,y)$.*

Proof. *Step* 1. We first prove that there exists $\bar C\in\mathcal{L}$ such that if $d_c(x,y)\le 1/\bar C^2(x)$ then $d(x,y)\le 2\bar C(x)d_c(x,y)$.

Assume that $d_c(x,y)\le 1/\bar C^2(x)$, with $\bar C\in\mathcal{L}$ to be chosen later. We set $\delta(x)=\bar C^2(x)d_c(x,y)^2$. Notice that $\delta(x)\le 1/\bar C^2(x)$. (A.3) with $\delta=\delta(x)$ gives

$$d_c(x,y)=\sqrt{\delta(x)}\inf_{\phi\in C^2_{\sigma,\delta(x)}(x,y)}\varepsilon_\phi\big(\delta(x)\big)=\bar C(x)d_c(x,y)\inf_{\phi\in C^2_{\sigma,\delta(x)}(x,y)}\varepsilon_\phi\big(\delta(x)\big)$$

and thus,

$$\inf_{\phi\in C^2_{\sigma,\delta(x)}(x,y)}\varepsilon_\phi\big(\delta(x)\big)=\frac{1}{\bar C(x)}<\frac{2}{\bar C(x)}.$$

Hence, there exists $\phi_*\in C^2_{\sigma,\delta(x)}(x,y)$ such that

$$\varepsilon_{\phi_*}\big(\delta(x)\big)<\frac{2}{\bar C(x)}.$$

For every fixed x, we apply Remark 5.4 to ϕ_* (recall that here $b\equiv 0$): there exists $\bar\delta,\bar\varepsilon\in 1/\mathcal{L}$ and $\bar C\in\mathcal{L}$ such that (with the slightly different notation of the present section)

$$\big|u_\delta(\phi_*)-x\big|_{A_\delta(x)}\le\overline{C}(x)\big(\varepsilon_{\phi_*}(\delta)\vee\sqrt{\delta}\big),$$

for every δ such that $\delta\le\bar\delta(x)$ and $\varepsilon_{\phi_*}(\delta)\le\bar\varepsilon(x)$. We have just proved that $\delta(x)\le 1/\bar C^2(x)$ and $\varepsilon_{\phi_*}(\delta(x))\le 2/\bar C(x)$. So, possibly taking $\bar C$ larger, we can actually use $\delta=\delta(x)$. Since $u_{\delta(x)}(\phi_*)=y$, the above inequality gives

$$|y-x|_{A_\delta(x)}\le\overline{C}(x)\big(\varepsilon_{\phi_*}\big(\delta(x)\big)\vee\sqrt{\delta(x)}\big)\le 2.$$

By (2.9), we obtain $|y-x|_{A_{4\delta}(x)}\le 1$, that is $d(x,y)\le\sqrt{4\delta(x)}=2\bar C(x)d_c(x,y)$, and the statement follows.

Step 2. We prove now the converse inequality. We use a result from [25], for which we need to recall the definition of the quasi-distance d_* (denoted by ρ_2 in [25]). The definition we give here is slightly different but clearly equivalent. For $\theta\in\mathbb{R}^m$, consider the equation

$$dv_t(\theta)=A\big(v_t(\theta)\big)\theta\,dt. \tag{A.4}$$

We denote

$$\bar C_A(x,y)=\big\{\theta\in\mathbb{R}^m:\ \text{the solution } v(\theta) \text{ to (A.4) satisfies } v_0(\theta)=x \text{ and } v_1(\theta)=y\big\}.$$

Notice that $\theta\in\bar C_A(x,y)$ is a constant vector, and not a time depending control as in (A.2). Moreover, recalling the definitions (2.4)–(2.5) for A, (A.4) involves also the vector fields $[\sigma_i,\sigma_j]$, differently from (A.2). In both equations the drift term b does not appear.

Let $\mathcal{D}_R$ be the diagonal matrix in (2.7) and recall that $A_R(x)=A(x)\mathcal{D}_R$. We define

$$d_*(x,y)=\inf\big\{R>0:\ \text{there exists } \theta\in\bar C_A(x,y) \text{ such that } \big|\mathcal{D}_R^{-1}\theta\big|<1\big\}.$$

As a consequence of Theorem 2 and Theorem 4 from [25], 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 from above by d.

Let us be more precise: for $x \in O$, we look for $C \in \mathcal{L}$ and $R \in 1/\mathcal{L}$ such that the following holds: if $0 < R \le R(x)$ and $d(x,y) \le \sqrt{R}$, then there exists a control $\theta \in \bar{C}_A(x,y)$ such that $|\mathcal{D}_R^{-1}\theta| < C(x)$. 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 $C(x)$ and $R(x)$ depending on x.

Recall that $d(x,y) \le \sqrt{R}$ means $|x-y|_{A_R(x)} \le 1$, and this also implies $|x-y| \le \lambda^*(A(x))\sqrt{R}$, by (2.9). Let $v(\theta)$ denote the solution to (A.4) with $v_0(\theta) = x$. We look for θ such that $v_1(\theta) = y$. We define

$$\Phi(\theta) = \int_0^1 A\big(v_s(\theta)\big)\theta\, ds = A(x)\theta + r(\theta)$$

with $r(\theta) = \int_0^1 (A(v_s(\theta)) - A(x))\theta\, ds$. 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: $AA(x)^+ = \mathrm{Id}$; $|x-y|_{A(x)} = |A(x)^+(x-y)|$. Write $\theta = A(x)^+\gamma$, $\gamma \in \mathbb{R}^n$. This implies $A(x)\theta = \gamma$, and so we are looking for $\gamma \in \mathbb{R}^n$ such that

$$\gamma + r\big(A(x)^+\gamma\big) = y - x.$$

One has $r(0) = 0$, $\nabla r(0) = 0$ and, as a consequence, $|r(\theta)| \le C(x)|\theta|^2$, for some $C \in \mathcal{L}$ – from now on, $C \in \mathcal{L}$ will denote a function that may vary from line to line.

From the local inversion theorem (in a quantitative form), there exists $l \in \mathcal{L}$ such that $\gamma \mapsto \gamma + r(A(x)^+\gamma)$ is a diffeomorrphism from $B(0, l_x)$ to $B(0, l_x/2)$. Remark that $|x-y| \le \lambda^*(A(x))\sqrt{R}$. So, taking R_x such that $\lambda^*(A(x))\sqrt{R} = l_x/2$, then for every $R < R_x$ and $|y-x| < \lambda^*(A(x))\sqrt{R}$ then there exists a unique γ such that $\gamma + r(A(x)^+\gamma) = y - x$ and moreover, $|\gamma| \le 2|x-y|$. Now, using (2.9)

$$\big|r\big(A(x)^+\gamma\big)\big|_{A_R(x)} \le \frac{\lambda^*(A(x))|r(A(x)^+\gamma)|}{R} \le C_x\frac{|A(x)^+\gamma|^2}{R} \le C_x\frac{|x-y|^2}{R} \le C_x|x-y|^2_{A_R(x)}.$$

Since $\gamma = x - y - r(A(x)^+\gamma)$,

$$|\gamma|_{A_R(x)} \le |x-y|_{A_R(x)} + C_x|x-y|^2_{A_R(x)} \le C_x,$$

(using $|x-y|_{A_R(x)} \le 1$). We have $|\mathcal{D}_R^{-1}\theta| = |\mathcal{D}_R^{-1}A(x)^+\gamma|$. Since $A_R^+A_R(x) = A_R^+(x)A(x)\mathcal{D}_R$ is an orthogonal projection and $AA^+(x)$ is the identity,

$$\big|\mathcal{D}_R^{-1}\theta\big| \le \big|\mathcal{D}_R^{-1}A(x)^+\gamma\big| \le A\big|A_R^+(x)A(x)\mathcal{D}_R\mathcal{L}_R^{-1}A(x)^+\gamma\big| = \big|A_R^+(x)\gamma\big| = |\gamma|_{A_R(x)}.$$

So $|\mathcal{L}_R^{-1}\theta| \le C_x$, and this implies $d_*(x,y) \le C_x\sqrt{R}$. □

The proof of Theorem 2.14 is now an immediate consequence of Theorem 2.9 and Theorem A.1. The only apparent problem is that in Theorem A.1 the global estimate (A.1) is required, whereas in Theorem 2.14 the local estimate (H_1) in (2.12) holds. But this is not really a problem, since it can be handled as already done for Theorem 2.14 (see Remark 5.5).

Appendix B: Proof of decomposition Lemma

We prove the decomposition (3.6) in Lemma 3.1. We recall Z_t in (3.2):

$$Z_t = \sum_{i=1}^d a_i W_t^i + \sum_{i,j=1}^d a_{i,j}\int_0^t W_s^i \circ dW_s^j$$

with $a_i=\sigma_i(0,x_0)$, $a_{i,j}=\partial_{\sigma_i}\sigma_j(0,x_0)$. Setting $s_l=\frac{l}{d}\delta$, $l=1,\dots,d$, we have

$$Z_\delta=\sum_{l=1}^d Z(s_l)-Z(s_{l-1})=\sum_{l=1}^d\left(\sum_{i=1}^d a_i\Delta_l^i+\sum_{i,j=1}^d a_{i,j}\int_{s_{l-1}}^{s_l}W_s^i\circ dW_s^j\right).$$

Recalling the quantities Δ_l^j and $\Delta_l^{i,j}$ in (3.3), we write

$$\int_{s_{l-1}}^{s_l}W_s^i\circ dW_s^j=W_{s_{l-1}}^i\Delta_l^j+\Delta_l^{i,j}=\left(\sum_{p=1}^{l-1}\Delta_p^i\right)\Delta_l^j+\Delta_l^{i,j}.$$

Then

$$Z_\delta=\sum_{l=1}^d\sum_{i=1}^d a_i\Delta_l^i+\sum_{l=1}^d\sum_{i,j=1}^d a_{i,j}\left(\sum_{p=1}^{l-1}\Delta_p^i\right)\Delta_l^j+\sum_{l=1}^d\sum_{i,j=1}^d a_{i,j}\Delta_l^{i,j}=:S_1+S_2+S_3.$$

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

$$\begin{aligned}
S_3&=\sum_{l=1}^d\sum_{i=1}^d a_{i,i}\Delta_l^{i,i}+\sum_{l=1}^d\sum_{i\neq j}a_{i,j}\Delta_l^{i,j}\\
&=\sum_{l=1}^d\sum_{i=1}^d a_{i,i}\Delta_l^{i,i}+\sum_{l=1}^d\sum_{i\neq l}a_{i,l}\Delta_l^{i,l}+\sum_{l=1}^d\sum_{j\neq l}a_{l,j}\Delta_l^{l,j}+\sum_{l=1}^d\sum_{i\neq j,i\neq l j\neq l}a_{i,j}\Delta_l^{i,j}\\
&=\frac{1}{2}\sum_{l=1}^d\sum_{i=1}^d a_{i,i}\left|\Delta_l^i\right|^2+\sum_{l=1}^d\sum_{i\neq l}a_{i,l}\Delta_l^{i,l}\\
&\quad+\sum_{l=1}^d\sum_{j\neq l}a_{l,j}\left(\Delta_l^j\Delta_l^l-\Delta_l^{j,l}\right)+\sum_{l=1}^d\sum_{i\neq j,i\neq l,j\neq l}a_{i,j}\Delta_l^{i,j}\\
&=\frac{1}{2}\sum_{i=1}^d a_{i,i}\left|\Delta_i^i\right|^2+\frac{1}{2}\sum_{l=1}^d\sum_{i\neq l}a_{i,i}\left|\Delta_l^i\right|^2+\sum_{l=1}^d\sum_{i\neq l}(a_{i,l}-a_{l,i})\Delta_l^{i,l}\\
&\quad+\sum_{l=1}^d\left(\sum_{j\neq l}a_{l,j}\Delta_l^j\right)\Delta_l^l+\sum_{l=1}^d\sum_{i\neq j,i\neq l,\neq j\neq}a_{i,j}\Delta_l^{i,j}.
\end{aligned}$$

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

$$S_2' = \sum_{l>p}^{d} a_{p,l} \Delta_p^p \Delta_l^l, \qquad S_2'' = \sum_{l>p}^{d} \sum_{j \neq l} a_{p,j} \Delta_p^p \Delta_l^j,$$

$$S_2''' = \sum_{l>p}^{d} \sum_{i \neq p}^{d} a_{i,l} \Delta_p^i \Delta_l^l, \qquad S_2^{iv} = \sum_{l>p}^{d} \sum_{i \neq p, j \neq l} a_{i,j} \Delta_p^i \Delta_l^j.$$

We have

$$S_2'' = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=p+1}^{d} \sum_{j \neq l} a_{p,j} \Delta_l^j \right)$$

and

$$S_2''' = \sum_{l=1}^{d} \Delta_l^l \left(\sum_{p=1}^{l-1} \sum_{i \neq p} a_{i,l} \Delta_p^i \right) = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=1}^{p-1} \sum_{j \neq l} a_{j,p} \Delta_l^j \right)$$

so that

$$S_2'' + S_2''' = \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l=p+1}^{d} \sum_{j \neq l} a_{p,j} \Delta_l^j + \sum_{l=1}^{p-1} \sum_{j \neq l} a_{j,p} \Delta_l^j \right).$$

Finally

$$\begin{aligned}
Z_\delta &= \sum_{l=1}^{d} a_l \Delta_l^l + \sum_{l=1}^{d} \sum_{i \neq l} a_i \Delta_l^i \\
&\quad + \sum_{l>p}^{d} a_{p,l} \Delta_p^p \Delta_l^l + \sum_{p=1}^{d} \Delta_p^p \left(\sum_{l>p}^{d} \sum_{j \neq l} a_{p,j} \Delta_l^j + \sum_{p>l}^{d} \sum_{j \neq l} a_{j,p} \Delta_l^j \right) \\
&\quad + \sum_{l>p}^{d} \sum_{i \neq p, j \neq l} a_{i,j} \Delta_p^i \Delta_l^j + \frac{1}{2} \sum_{i=1}^{d} a_{i,i} \left| \Delta_i^i \right|^2 + \frac{1}{2} \sum_{l=1}^{d} \sum_{i \neq l} a_{i,i} \left| \Delta_l^i \right|^2 \\
&\quad + \sum_{l=1}^{d} \sum_{i \neq l} (a_{i,l} - a_{l,i}) \Delta_l^{i,l} + \sum_{l=1}^{d} \left(\sum_{j \neq l} a_{l,j} \Delta_l^j \right) \Delta_l^l + \sum_{l=1}^{d} \sum_{i \neq j, i \neq l, j \neq l} a_{i,j} \Delta_l^{i,j}.
\end{aligned}$$

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

$$\varepsilon_p = \sum_{l>p}^{d} \sum_{j \neq l} a_{p,j} \Delta_l^j + \sum_{p>l}^{d} \sum_{j \neq l} a_{j,p} \Delta_l^j + \sum_{j \neq p} a_{p,j} \Delta_p^j.$$

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 vector $a_{i,p}-a_{p,i}$ corresponds to the bracket $[\sigma_i,\sigma_p](0,x)$. Notice that for $l=l(i,p)$ when $i\neq p$, then $[\sigma_i,\sigma_p](0,x)=A_l(0,x)$, $A_l(0,x)$ being the lth column of $A(0,x)$. The other terms are

$$\sum_{l=1}^{d}\sum_{i\neq l}a_i\Delta_l^i+\sum_{l>p}^{d}\sum_{i\neq p,j\neq l}a_{i,j}\Delta_p^i\Delta_l^j+\frac{1}{2}\sum_{i=1}^{d}a_{i,i}\left|\Delta_i^i\right|^2+\frac{1}{2}\sum_{l=1}^{d}\sum_{i\neq l}a_{i,i}\left|\Delta_l^i\right|^2$$
$$+\sum_{l=1}^{d}\sum_{i\neq j,i\neq l,j\neq l}a_{i,j}\Delta_l^{i,j}+\sum_{l=p+1}^{d}a_{p,l}\Delta_p^p\Delta_l^l.$$

We put everything together and (3.6) is proved.

Appendix C: Support property

The aim of this section is the proof of the inequality in (C.5), which has been strongly used in Lemma 3.4.

Let $B=(B^1,\dots,B^{d-1})$ be a standard Brownian motion. We consider the analogous of the covariance matrix $Q_i(B)$ considered in Section 3.1: we define a symmetric square matrix of dimension $d\times d$ by

$$\begin{aligned}&Q^{d,d}=1,\qquad Q^{d,j}=Q^{j,d}=\int_0^1 B_s^j\,ds,\quad j=1,\dots,d-1,\\&Q^{j,p}=Q^{p,j}=\int_0^1 B_s^jB_s^p\,ds,\quad j,p=1,\dots,d-1\end{aligned}\tag{C.1}$$

and we denote by $\lambda_*(Q)$ (respectively by $\lambda^*(Q)$) the lowest (respectively largest) eigenvalue of Q.

For a measurable function $g:[0,1]\to R^{d-1}$ we denote

$$\alpha_g(\xi)=\xi_d+\int_0^1\langle g_s,\xi_*\rangle\,ds,\qquad \beta_g(\xi)=\int_0^1\langle g_s,\xi_*\rangle^2\,ds-\left(\int_0^1\langle g_s,\xi_*\rangle\,ds\right)^2\quad\text{with}$$
$$\xi=(\xi_1,\dots,\xi_d)\in\mathbb{R}^d\quad\text{and}\quad\xi_*=(\xi_1,\dots,\xi_{d-1}).$$

We need the following two preliminary lemmas.

Lemma C.1. *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

$$\lambda_*(Q)=\inf_{|\xi|=1}\left(\alpha_B^2(\xi)+\beta_B(\xi)\right)\quad\text{and}\quad\lambda^*(Q)\le\sup_{|\xi|=1}\left(\alpha_B^2(\xi)+\beta_B(\xi)\right)\le\left(1+\sup_{t\le1}|B_t|\right)^2.$$

Taking $\xi_*=0$ and $\xi_d=1$ we obtain $\langle Q\xi,\xi\rangle=1$ so that $\lambda_*(Q)\le1\le\lambda^*(Q)$.

Proof. By direct computation

$$\begin{aligned}\langle Q\xi,\xi\rangle&=\xi_d^2+2\xi_d\int_0^1\langle B_s,\xi_*\rangle\,ds+\left(\int_0^1\langle B_s,\xi_*\rangle\,ds\right))^2\\&\quad+\int_0^1\langle B_s,\xi_*\rangle^2\,ds-\left(\int_0^1\langle B_s,\xi_*\rangle\,ds\right)^2\\&=\left(\xi_d+\int_0^1\langle B_s,\xi_*\rangle\,ds\right)^2+\int_0^1\langle B_s,\xi_*\rangle^2\,ds-\left(\int_0^1\langle B_s,\xi_*\rangle\,ds\right)^2.\end{aligned}$$

The remaining statements follow straightforwardly. □

Proposition C.2. *For each $p \geq 1$ one has*

$$\mathbb{E}\big(|\det Q|^{-p}\big) \leq C_{p,d} < \infty, \tag{C.2}$$

where $C_{p,d}$ is a constant depending on p, d only.

Proof. By Lemma 7.29, pg 92 in [13], for every $p \in (0,\infty)$ one has

$$\frac{1}{|\det Q|^p} \leq \frac{1}{\Gamma(p)} \int_{R^d} |\xi|^{d(2p-1)} e^{-\langle Q\xi,\xi\rangle}\, d\xi.$$

Let $\theta(\xi_*) := \int_0^1 \langle B_s, \xi_*\rangle\, ds$. Using the previous lemma

$$\begin{aligned}
\int_{R^d} |\xi|^{d(2p-1)} e^{-\langle Q\xi,\xi\rangle}\, d\xi &= \int_{R^d} \big(\xi_d^2 + |\xi_*|^2\big)^{d(2p-1)/2} e^{-(\xi_d+\theta(\xi_*))^2-\beta_B(\xi_*)}\, d\xi \\
&\leq C \int_{R^{d-1}} \big(\big(1+\theta^2(\xi_*)\big)^{d(2p-1)/2} + |\xi_*|^{d(2p-1)}\big) e^{-\beta_B(\xi_*)}\, d\xi_* \\
&\leq C \int_{R^{d-1}} \sup_{t\leq 1} 1 \vee |B_t|^{d(2p-1)} \big(1+|\xi_*|^{d(2p-1)+1}\big) e^{-\beta_B(\xi_*)}\, d\xi_*.
\end{aligned}$$

We integrate and we use Schwartz inequality in order to obtain

$$\mathbb{E}\bigg(\frac{1}{|\det Q|^p}\bigg) \leq C + C\int_{\{|\xi_*|\geq 1\}} \big(\mathbb{E}\big(\big(1+|\xi_*|^{d(2p-1)+1}\big)^2 e^{-2\beta_B(\xi_*)}\big)\big)^{1/2}\, d\xi_*.$$

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 [17] (see (1.f), p. 183) that

$$\mathbb{E}\big(e^{-2\beta_B(\xi_*)}\big) = \mathbb{E}\big(e^{-2|\xi_*|^2 V_{\xi_*}}\big) = \frac{2|\xi_*|^2}{\sinh 2|\xi_*|^2}.$$

We insert this in the previous inequality and we obtain $\mathbb{E}(|\det Q|^{-p}) < \infty$. □

We are now able to give the main result in this section. We define

$$q(B) = \sum_{i=1}^{d-1} \big|B_1^i\big| + \sum_{j\neq p} \bigg|\int_0^1 B_s^j\, dB_s^p\bigg| \tag{C.3}$$

and for $\varepsilon, \rho > 0$ we denote

$$\Upsilon_{\rho,\varepsilon}(B) = \Big\{\det Q \geq \varepsilon^\rho, \sup_{t\leq 1} |B_t| \leq \varepsilon^{-\rho}, q(B) \leq \varepsilon\Big\}. \tag{C.4}$$

Proposition C.3. *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*

$$\mathbb{P}\big(\Upsilon_{\rho,\varepsilon}(B)\big) \geq c_{\rho,d} \times \varepsilon^{\frac{1}{2}d(d+1)}. \tag{C.5}$$

Proof. Using the previous proposition and Chebyshev's inequality we get

$$\mathbb{P}\big(\det Q < \varepsilon^\rho\big) \leq \varepsilon^{p\rho} \mathbb{E}|\det Q|^{-p} \leq C_{p,d}\varepsilon^{p\rho} \quad \text{and} \quad \mathbb{P}\Big(\sup_{t\leq 1}|B_t| > \varepsilon^{-\rho}\Big) \leq \exp\bigg(-\frac{1}{C\varepsilon^{2\rho}}\bigg).$$

Let $q'(B) = \sum_{i=1}^{d-1} |B_1^i| + \sum_{j<p} |\int_0^1 B_s^j \, dB_s^p|$. Since $|\int_0^1 B_s^j \, dB_s^p| \leq |B_1^j||B_1^p| + |\int_0^1 B_s^p \, dB_s^j|$ 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 $X = (X^i, X^{j,p}, i = 1, \dots, d, 1 \leq j < p \leq d)$ 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 [23] or [2]). 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

$$\mathbb{P}\big(q(B) \leq \varepsilon\big) \geq \mathbb{P}\left(q'(B) \leq \frac{\varepsilon}{3}\right) \geq \mathbb{P}\left(|X_1| \leq \frac{\varepsilon}{3\sqrt{m}}\right) \geq \frac{\varepsilon^m}{(3\sqrt{m})^m} \times \bar{c}_d,$$

with $\bar{c}_d > 0$. So finally we obtain

$$\mathbb{P}\big(\Upsilon_{\rho,\varepsilon}(B)\big) \geq \bar{c}_d \varepsilon^{\frac{1}{2}d(d+1)} - C_{p,d}\varepsilon^{p\rho} - \exp\left(-\frac{1}{C\varepsilon^{2\rho}}\right).$$

Choosing $p > \frac{1}{2\rho}d(d+1)$ and ε small we obtain our inequality. $\square$

Appendix D: Density estimates via local inversion

In this section 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. For a standard version of the inverse function theorem see [28].

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

$$c_2(\eta) = \max_{i,j=1,\dots,m} \sup_{|x|\leq 1} \left|\partial_{ij}^2 \eta(x)\right|, \qquad c_3(\eta) = \max_{i,j,k=1,\dots,m} \sup_{|x|\leq 1} \left|\partial_{ijk}^3 \eta(x)\right|, \tag{D.1}$$

and

$$h_\eta = \frac{1}{16m^2(c_2(\eta) + \sqrt{c_3(\eta)})}. \tag{D.2}$$

Lemma D.1. *Take h_η as above. If the function η is such that*

$$\eta \in C^3\big(\mathbb{R}^m, \mathbb{R}^m\big), \quad \eta(0) = 0, \qquad \nabla\eta(0) \leq \frac{1}{2},$$

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

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

and we have this quantitative estimate:

$$\forall y \in B\left(0, \frac{1}{2}h_\eta\right), \quad \frac{1}{4}\left|\Phi^{-1}(y)\right| \leq |y| \leq 4\left|\Phi^{-1}(y)\right|. \tag{D.3}$$

Remark D.2. 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

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

So

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

and

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

Therefore

$$\frac{1}{2}|x| \leq \left|\nabla\Phi(0)x\right| \leq \sqrt{3}|x|.$$

This implies $\Phi(0)$ is invertible locally around 0, and the local inverse is differentiable, using the classical inverse function theorem. We now look at the image of the inverse, and at the estimates (D.3). 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 $y \in \mathbb{R}^m$. Suppose $\Phi(\theta) = y$. Then

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

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 $|\frac{1}{2}x| \leq |\nabla\Phi(0)x|$.

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

Now, from (D.2), for $\theta_1, \theta_2 \in B(0,h_\eta)$

$$\left|\nabla^2\eta(t\theta_1)[\theta_1,\theta_1-\theta_2]\right| \leq m^2 c_2(\eta)h_\eta|\theta_1-\theta_2| \leq \frac{1}{16}|\theta_1-\theta_2|,$$

and

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

and therefore

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

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

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

Define now the sequence

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

We know that $\theta_k\in B(0,2h_\eta)$ for any $k\in\mathbb{N}$, therefore inequality (D.4) implies

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

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 $B(0,\frac{1}{2}h_\eta)$, 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

$$|\theta|=\left|U_y(\theta)\right|\le\frac{1}{2}\theta+2|y|\quad\Rightarrow\quad|\theta|\le 4|y|,$$
$$|\theta|=U_y(\theta)\ge\left|U_y(0)\right|-\left|U_y(\theta)-U_y(0)\right|\ge\frac{1}{2}|y|-\frac{1}{2}|\theta|\quad\Rightarrow\quad|\theta|\ge\frac{1}{4}|y|.\qquad\square$$

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

$$c_*(\eta,h)=\sup_{|x|\le 2h}\max_{i,j}\left|\partial_i\eta^j(x)\right| \tag{D.5}$$

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

Take $r>0$ such that

$$c_*(\eta,16r)\le\frac{1}{2m}\sqrt{\frac{\underline{\lambda}}{\overline{\lambda}}},\qquad r\le h_\eta=\frac{1}{16m^2(c_2(\eta)+\sqrt{c_3(\eta)})}. \tag{D.6}$$

We take a localizing function as in (E.2):

$$U=\prod_{i=1}^m\psi_r(\Theta_i). \tag{D.7}$$

Lemma D.3. *Let Q be non degenerate. Let r such that* (D.6) *holds and set U as in* (D.7). *Then the density $p_{G,U}$ of*

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

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

$$\frac{1}{C \det Q^{1/2}} \exp\left(-\frac{C}{\underline{\lambda}}|z|^2\right) \le p_{G,U}(z) \le \frac{C}{\det Q^{1/2}} \exp\left(-\frac{1}{C\overline{\lambda}}|z|^2\right). \tag{D.8}$$

Proof. For a general nonnegative, measurable function $f : \mathbb{R}^m \to \mathbb{R}$ with support included in $B(0,4r)$, we compute $\mathbb{E}(f(G)1_{\{\Theta\in\Phi^{-1}B(0,4r)\}})$. 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 D.1 for $G = \Phi(\Theta)$

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

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

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

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

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

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

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

Moreover, using Lemma D.1

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

Therefore

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

Now we define, as in (E.2) the localization variables

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

Notice that

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

so that we have

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

The following bounds for the local densities follow:

$$p_{G,U_1}(z) \ge \frac{1}{(8\pi)^{m/2} \det Q^{1/2}} \exp\left(-\frac{8}{\underline{\lambda}}|z|^2\right),$$
$$p_{G,U_2}(z) \le \frac{2^{m/2}}{\pi^{m/2} \det Q^{1/2}} \exp\left(-\frac{1}{32\overline{\lambda}}|z|^2\right).$$

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

Appendix E: Localization and density estimates

We first recall some notations and basic notions in Malliavin calculus. Our main reference is [26]. Given a d-dimensional Brownian motion $W = (W_t^1, \dots, W_t^d)_{t \ge 0}$ we denote its Malliavin derivative as

$$DF = \big(DF^1, \dots, DF^d\big)^T.$$

We introduce the Sobolev norm of F:

$$\|F\|_{1,p} = \big[\mathbb{E}|F|^p + \mathbb{E}|DF|^p\big]^{\frac{1}{p}} \quad \text{where } |DF| = \left(\int_0^T |D_s F|^2 \, ds\right)^{\frac{1}{2}}.$$

For any $k \in \mathbb{N}$, for a multi-index $\alpha = (\alpha_1, \dots, \alpha_k) \in \{1, \dots, d\}^k$ and $(s_1, \dots, s_k) \in [0, T]^k$, we denote the higher order derivative as

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

We denote by $|\alpha| = k$ the length of the multi-index. We define the Sobolev norm of $D^{\alpha}_{s_1,\dots,s_k} F$ as

$$\|F\|_{k,p} = \left[\mathbb{E}|F|^p + \sum_{j=1}^{k} \mathbb{E}\big|D^{(j)}F\big|^p\right]^{\frac{1}{p}} \quad \text{where } \big|D^{(j)}F\big| = \left(\sum_{|\alpha|=j} \int_{[0,T]^j} \big|D^{\alpha}_{s_1,\dots,s_j} F\big|^2 \, ds_1 \dots ds_j\right)^{1/2}.$$

Notice that with this notation $|DF| = |D^{(1)}F|$. Also notice that $D^{(j)}$ means "derivative of order j" and D^j means "derivative with respect to W^j".

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 by L the Ornstein–Uhlenbeck operator, i.e. $L = -\delta \circ D$, where δ is the adjoint operator of D.

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

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

The following notion of localization is introduced in [2]. Consider a random variable $U \in [0,1]$ and denote

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

$\mathbb{P}_U$ is a non-negative measure (not a probability measure, in general). We also set $\mathbb{E}_U$ the expectation (integral) w.r.t. $\mathbb{P}_U$, and denote

$$\|F\|_{p,U}^p = \mathbb{E}_U\big(|F|^p\big) = \mathbb{E}\big(|F|^p U\big),$$

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

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

$$m_U(p) := 1 + \mathbb{E}_U|D\ln U|^p < \infty. \tag{E.1}$$

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

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

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

$$U = \prod_{i=1}^n \psi_{a_i}(\Theta_i). \tag{E.2}$$

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

$$m_U(p) \le C\frac{\|\Theta\|_{1,p}^p}{|a|}. \tag{E.3}$$

The proof of (E.3) follows from standard computations and inequality

$$\sup_x\big|(\ln\psi_a)(x)\big|^p\psi_a(x) \le \frac{C}{a^p}.$$

In the following proposition we state the general lower and upper bound that we use in our density estimate. These results are proved in [2] and [3].

Proposition E.1. *Let $F \in (\mathbb{D}^{2,\infty})^d$.*

1. *Suppose that for every $p \in \mathbb{N}$: $\mathbb{E}_U|\lambda_*(\gamma_F)|^{-p} < \infty$, $U \in \mathbb{D}^{1,\infty}$ and $m_U(p) < \infty$. Let $G \in (\mathbb{D}^{2,\infty})^d$ such that for every $p \in \mathbb{N}$*

$$\mathbb{E}_U\big|\lambda_*(\gamma_G)\big|^{-p} < \infty.$$

Then for every $p > d$

$$\begin{aligned} &p_{F,U}(y)\\ &\quad\ge p_{G,U}(y) - Cm_U(p)^b\max\big\{1, \big(\mathbb{E}_U\big|\lambda_*(\gamma_G)\big|^{-p}\big)^b\big(\|F\|_{2,p,U} + \|G\|_{2,p,U}\big)\big\}\|F-G\|_{2,p,U}, \end{aligned} \tag{E.4}$$

where C, b are constants depending only on d, p and $m_U(p)$ is given by (E.1).

2. *Assume* $\mathbb{E}|\lambda_*(\gamma_F)|^{-p} < \infty, \forall p$. *Then* $\exists C, p, b$ *constants depending only on the dimension* d *such that*

$$\left|p_F(y)\right| \le C \max\left\{1, \mathbb{E}\left|\lambda_*(\gamma_F)\right|^{-p} \|F\|_{2,p}\right\} \mathbb{P}\left(|F-y|<2\right)^b. \tag{E.5}$$

Proof.

1. The lower bound (E.4) for $p_{F,U}$ is a version of Proposition 2.5. in [2] with the lowest eigenvalue instead of the determinant.

2. The upper bound (E.5) for p_F is a version of Theorem 2.14, point A., in [3]. We take therein $q = 0$, so there is no derivative, and $\Theta = 1$, that means that we are not localizing.

□

Acknowledgements

The authors gratefully acknowledge Antoine Ledent for discussion and useful comments.

L.C. acknowledges the MIUR Excellence Department Project awarded to the Department of Mathematics, University of Rome Tor Vergata, CUP E83C18000100006.

References

[1] D. Bakry, F. Baudoin, M. Bonnefont and D. Chafai. On gradient bounds for the heat kernel on the Heisenberg group. *J. Funct. Anal.* **255** (8) (2008) 1905–1938. MR2462581

[2] V. Bally and L. Caramellino. Positivity and lower bounds for the density of Wiener functionals. *Potential Anal.* **39** (2013) 141–168. MR3078335

[3] V. Bally and L. Caramellino. On the distances between probability density functions. *Electron. J. Probab.* **19** (110) (2014) 1–33. MR3296526

[4] V. Bally, L. Caramellino and P. Pigato. Diffusions under a local strong Hörmander condition. Part I: Density estimates. *ArXiv e-prints* (2016). Available at 1607.04542.

[5] V. Bally, L. Caramellino and P. Pigato. Diffusions under a local strong Hörmander condition. Part II: Tube estimates. *ArXiv e-prints* (2016). Available at 1607.04544. MR3765891

[6] V. Bally and S. De Marco. Some estimates in extended stochastic volatility models of Heston type. *Risk Decis. Anal.* **2** (4) (2011) 195–206.

[7] V. Bally, B. Fernandez and A. Meda. Estimates for the probability that a Itô process remains near a path. In *Stochastic Processes and Their Applications* 2087–2113, **121**, 2011. MR2819243

[8] V. Bally and A. Kohatsu-Higa. Lower bounds for densities of Asian type stochastic differential equations. *J. Funct. Anal.* **258** (9) (2010) 3134–3164. MR2595738

[9] F. Baudoin. *An Introduction to the Geometry of Stochastic Flows*. Imperial College Press, London, 2004. MR2154760

[10] G. Ben Arous, M. Gradinaru and M. Ledoux. Hölder norms and the support theorem for diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **30** (1994) 415–436. MR1288358

[11] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. I. *Probab. Theory Related Fields* **90** (2) (1991) 175–202. MR1128069

[12] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. II. *Probab. Theory Related Fields* **90** (3) (1991) 377–402. MR1133372

[13] K. Bichteler, J.-B. Gravereaux and J. Jacod. *Malliavin Calculus for Processes with Jumps, Volume Stochastic Mongraphs*, **2**. Gordon and Breach Science Publishers, New York, 1987. MR1008471

[14] M. Capitaine. On the Onsager–Machlup functional for elliptic diffusion processes. In *Sminaire de Probabilits XXXIV* 313–328. *Lecture Notes in Mathematics* **1729**. Springer, Berlin Heidelberg, 2000. MR1768071

[15] P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions. *Probab. Theory Related Fields* **123** (2002) 453–483. MR1921010

[16] F. Delarue and S. Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *J. Funct. Anal.* **259** (6) (2010) 1577–1630. MR2659772

[17] 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** (1991) 181–200. MR1118933

[18] B. K. Driver and T. Melcher. Hypoelliptic heat kernel inequalities on the Heisenberg group. *J. Funct. Anal.* **221** (2) (2005) 340–365. MR2124868

[19] P. Friz, T. Lyons and D. Stroock. Lvy's area under conditioning. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** (2006) 89–101. MR2196973

[20] A. Guillin. Averaging principle of sde with small diffusion: Moderate deviations. *Ann. Probab.* **31** (1) (2003) 413–443. MR1959798

[21] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*, 2nd edition. *North-Holland Mathematical Library* **24**. North-Holland Publishing Co., Amsterdam, 1989. MR0637061

[22] D. Jerison and A. Sánchez-Calle *Subelliptic, Second Order Differential Operators*, 46–77. Springer, Berlin, Heidelberg, 1987.

[23] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. III. *J. Fac. Sci., Univ. Tokyo, Sect. 1A, Math.* **34** (2) (1987) 391–442. MR0914028

[24] H.-Q. Li. Estimations asymptotiques du noyau de la chaleur sur les groupes de Heisenberg. *C. R. Math. Acad. Sci. Paris* **344** (8) (2007) 497–502. MR2324485

[25] A. Nagel, E. M. Stein and S. Wainger. Balls and metrics defined by vector fields. I. Basic properties. *Acta Math.* **155** (1–2) (1985) 103–147. MR0793239

[26] D. Nualart. *Malliavin Calculus and Related Topics*. Springer, Berlin, 2006. MR2200233

[27] P. Pigato. Tube estimates for diffusion processes under a weak Hörmander condition. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** (1) (2018) 299–342. MR3765891

[28] W. Rudin. *Principles of Mathematical Analysis*, 3rd edition. *International Series in Pure and Applied Mathematics*. McGraw-Hill Book Co., New York, 1976. MR0385023

[29] A. Sánchez-Calle. Fundamental solutions and geometry of the sum of squares of vector fields. *Invent. Math.* **78** (1) (1984) 143–160. MR0762360

[30] 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 III: Probability Theory* 333–359. Univ. California Press, Berkeley, 1972. MR0400425