Evolution of Gaussian Concentration bounds under diffusions

J.-R. Chazottes ∗, P. Collet †, and F. Redig ‡

1CPHT, CNRS, Ecole polytechnique, IP Paris, F-91128 Palaiseau Cedex, France
2Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

July 2, 2021

Abstract

We consider the behavior of the Gaussian concentration bound (GCB) under stochastic time evolution. More precisely, we consider a Markovian diffusion process on \( \mathbb{R}^d \) and start the process from an initial distribution \( \mu \) that satisfies GCB. We then study the question whether GCB is preserved under the time-evolution, and if yes, how the constant behaves as a function of time. In particular, if for the constant we obtain a uniform bound, then we can also conclude properties of the stationary measure(s) of the diffusion process. This question, as well as the methodology developed in the paper allows to prove Gaussian concentration via semigroup interpolation method, for measures which are not available in explicit form.

We provide examples of conservation of GCB, loss of GCB in finite time, and loss of GCB at infinity. We also consider diffusions “coming down from infinity” for which we show that, from any starting measure, at positive times, GCB holds. Finally we consider a simple class of non-Markovian diffusion processes with drift of Ornstein-Uhlenbeck type, and general bounded predictable variance.

Key-words: Markov diffusions, Ornstein-Uhlenbeck process, nonlinear semigroup, coupling, Bakry-Emery criterion, non-reversible diffusions, diffusions coming down from infinity, Ginzburg-Landau diffusions, non-Markovian diffusions, Lorenz attractor with noise, Burkholder inequality.

MSC numbers: 39B62, 60J60, 47D07.

∗Email: jeanrene@cpht.polytechnique.fr
†Email: collet@cpht.polytechnique.fr
‡Email: f.h.j.redig@tudelft.nl
## Contents

1 Introduction ............................................................. 3

2 Setting and main questions .............................................. 6
   2.1 Gaussian concentration bounds: definitions .................. 6
   2.2 Equivalence between Gaussian concentration and distance Gaussian moment bounds ......................... 7
   2.3 A class of Markov diffusion processes on $\mathbb{R}^d$ .................. 9

3 Propagation of Gaussian concentration .................................. 10
   3.1 A general result ..................................................... 10
   3.2 Examples .............................................................. 14
      3.2.1 No GCB at any $t > 0$ .................................. 14
      3.2.2 GCB for all finite times but $D_t \to \infty$ as $t \to \infty$ ...... 14
      3.2.3 Blowing up of $D_t$ in finite time .......................... 14

4 Nonlinear semigroup, gradient bounds, and coupling ..................... 16
   4.1 The nonlinear semigroup ............................................ 16
   4.2 Some preparatory computations ................................... 18
   4.3 Abstract gradient bound approach ................................. 21
   4.4 Coupling approach .................................................. 23
      4.4.1 Coupling and the nonlinear semigroup .................... 24
      4.4.2 Examples ...................................................... 28

5 Further applications of the distance Gaussian moment .................. 31
   5.1 Diffusions coming down from infinity ............................ 31
   5.2 Diffusion processes with space-time dependent drift and covariance ........................................... 35
   5.3 Example: the noisy Lorenz system ................................ 38

6 Non-Markovian diffusions: Martingale moment approach ............... 38

A Proof of Theorem 2.1 ................................................. 41

B A general approximation lemma ..................................... 43
1 Introduction

Some generalities. Concentration inequalities are a well-studied subject in probability and statistics, and are very useful in the study of fluctuations of possibly complicated and indirectly defined functions of random variables, such as the Kantorovich distance between the empirical distribution and the true distribution, and various properties of random graphs. See for instance [4, 16] and references therein. Initially mostly studied in the context of independent random variables, many efforts have been done to extend concentration inequalities to the context of dependent random variables, and more generally dependent random fields. For instance, in the context of models of statistical mechanics, where the dependence is naturally encoded in the interaction potential, it is proved in [15] that a Gaussian concentration bound holds under the Dobrushin uniqueness condition, so in particular for finite-range interaction potentials at “high temperature”. An example where Gaussian concentration fails is the Ising model at sufficiently low temperature where a weaker bound holds [5]. We refer to [6] for various applications of these bounds.

In this paper we are interested in the behavior (preservation, loss, recovery) of concentration inequalities under stochastic time-evolution. To our knowledge this natural question of time evolution of concentration has not been addressed directly anywhere in the literature. There are however several motivations to be interested in this problem. First, in the context of non-equilibrium systems, non-equilibrium stationary states and transient non-equilibrium states are usually characterized rather implicitly via an underlying dynamics. If we are interested in concentration properties of such measures, we are naturally led to consider time-evolved measures, and their concentration properties.

It is also used in various contexts that a Markovian semigroup interpolates between different measures [1], [16, Section 2.3], and therefore it is of interest whether this interpolation conserves concentration properties. Notice that in the context of Gibbs measures, stochastic time-evolution (even high-temperature dynamics) can destroy the Gibbs property [10], therefore it is interesting to understand whether such measures – though not Gibbs – still enjoy concentration properties, or whether there can be phase transitions in the concentration behavior of a measure, e.g., from a GCB to a weaker concentration bound in a dynamics leading from high to low-temperature regime.

In parallel to the present work, the behavior of concentration inequalities under spin-flip dynamics of configurations in $\{-1,1\}^\mathbb{Z}^d$ was studied in [7]. For “weakly interacting” dynamics we showed that the Gaussian concentration bound is preserved as time passes, and it is satisfied by the unique stationary Gibbs measure. We also showed that, for a general class of translation-invariant spin-flip dynamics, it is not possible to go in finite time
from a low-temperature Gibbs state to a measure satisfying the Gaussian concentration bound.

**Gaussian concentration bounds and Markovian diffusions processes.** In the present paper, we focus on Markovian diffusion processes in $\mathbb{R}^d$. Suppose that a probability measure $\mu_0$ satisfies a Gaussian concentration bound (hereinafter abbreviated as GCB). This means that there exists $D_0 \geq 0$ such that

$$
\mu_0 \left( e^{f - \mu_0(f)} \right) \leq e^{D_0 \text{lip}(f)^2}
$$

for all Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ (with respect to Euclidean distance). Now we ask the following question: does the evolved probability measure $\mu_t$ at time $t > 0$ satisfy GCB($D_t$) for some $D_t$? To be more concrete, let us first look at a very simple example, namely the one-dimensional Ornstein-Uhlenbeck process, i.e., the process $(X_t)_{t \geq 0}$ solving the stochastic differential equation

$$
dX_t = -\kappa X_t \, dt + \sigma \, dW_t
$$

where $\sigma, \kappa > 0$, and $(W_t)_{t \geq 0}$ is a standard Brownian motion. Let us denote by $X^x_t$ the solution starting from $X_0 = x$. Then we have

$$
X^x_t = e^{-\kappa t} x + \sigma \int_0^t e^{-\kappa (t-s)} \, dW_s.
$$

If we start from $X_0$ which is normally distributed with expectation zero and variance $\theta^2$ (denote by $\mathcal{N}(0, \theta^2)$ the corresponding distribution) then, at time $t > 0$, $X_t$ is normally distributed with expectation zero and variance

$$
\sigma^2_t = \theta^2 e^{-2\kappa t} + \frac{\sigma^2}{2\kappa} \left( 1 - e^{-2\kappa t} \right).
$$

Because the normal distribution $\mathcal{N}(0, a^2)$ satisfies GCB($D$) with $D = a^2/2$ we conclude that for this example, with $\mu_0 = \mathcal{N}(0, \theta^2)$, $\mu_t$ satisfies GCB($D_t$) with

$$
D_t = D_\infty + (D - D_\infty) e^{-2\kappa t}
$$

with $D_\infty = \sigma^2/(2\kappa)$. Hence, $\mu_t$ satisfies GCB($D_t$) with a constant $D_t$ interpolating smoothly between the initial constant $D = D_0$ and the constant $D_\infty$ associated to the stationary normal distribution.

In case $\kappa = 0$ the process is $\sigma W_t$, and we find

$$
\sigma^2_t = \theta^2 + \sigma^2 t
$$

which implies that the constant of the Gaussian concentration bound evolves as

$$
D_t = D_0 + \sigma^2 t.
$$
In particular, $D_t \to \infty$ as $t \to \infty$. We now come back to a general Markovian diffusion process $(X_t)_{t \geq 0}$ on $\mathbb{R}^d$ which solves a stochastic differential equation of the form

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

where $b$ and $\sigma$ satisfy standard assumptions detailed later on. Given an initial probability measure $\mu_0$ satisfying GCB, does $\mu_t$, the evolved probability measure, satisfy GCB for some constant $D_t$ at time $t$? Can we approximate $D_t$? What happens when $t \to \infty$? If $\mu_t$ converges to a stationary probability measure, does it satisfy GCB? It is possible that $D_t$ blows up in finite time?

The goal of this paper is to give an answer to these questions and some other ones formulated later on.

Outline of the paper. In Section 2, we define what we mean by a Gaussian concentration bound for a probability measure on $\mathbb{R}^d$ for two classes of functions, namely, Lipschitz functions, and smooth compactly supported functions. Indeed, we will need to work with smooth compactly supported functions at some places, and we prove (in a more general context than $\mathbb{R}^d$) that having GCB for this restricted class of functions enforces GCB for Lipschitz functions. In this section, we also establish a result of independent interest (for a general metric space) providing an equivalence between GCB for Lipschitz functions and a distance Gaussian moment bound (Theorem 2.1). Finally, we state the precise assumptions on the stochastic differential equations we will work with.

Section 3 contains our first general result on propagation of Gaussian concentration for finite time. We show through examples that, besides the case of GCB for all times with bounded constant as was already illustrated by the Ornstein-Uhlenbeck process, various scenarios can occur: No GCB at any $t > 0$; or GCB for all finite times but with a constant $D_t$ going to infinity when $t \to \infty$; or blowing up of $D_t$ in finite time.

Once we have the general result of Section 3, it is natural to ask about estimating $D_t$ when it exists for all $t$, or to ask whether or not the stationary measure (if it exists) satisfies GCB for a constant $D$ which is the limit as $t \to \infty$ of $D_t$, especially when one has no information beyond its mere existence. This is the purpose of Section 4. Because we need to estimate exponential moments of a time-evolved probability measure, as we will see later on in more detail, an object popping up naturally is the so-called nonlinear semigroup $V_t(f) = \log S_t(e^f)$ where $S_t$ is the Markov semigroup of the process under consideration, as well as its associated nonlinear generator $\mathcal{H}(f) = e^{-f} \mathcal{L}(e^f)$ where $\mathcal{L}$ is the Markov generator. It is then crucial to obtain estimates for the time-dependent Lipschitz constant of $V_t f$, which, because we can restrict to smooth $f$, as mentioned above, boils down to gradient estimates. For Markovian diffusion processes, the nonlinear generator $\mathcal{H}$ is a sum of a linear and a quadratic part, where the quadratic part coincides with the “carré du champ” operator. This implies that, in
the reversible setting, one can use general results on strong gradient bounds from [1]. This is done in Section 4.3. For the non-reversible setting, we follow a second approach, based on coupling, which is pursued in Section 4.4. We give examples from non-equilibrium steady states, and non-gradient perturbations of reversible diffusions.

The two approaches of Section 4 are mainly providing complementary sufficient conditions for preservation of GCB in the course of time as well as for the (unique) stationary measure, in a context of diffusion processes where the drift and the diffusion matrix do not explicitly depend on time. A third approach, based on estimating the Gaussian moment of the distance to the origin, already used in Section 3, provides a general result of conservation of GCB in a context where explicit time dependence of the drift and the diffusion matrix is allowed. This approach is pursued in section 5 and accompanied by examples illustrating loss of GCB in finite or infinite time, as well as diffusions coming down from infinity where GCB is obtained for all positive times from any initial distribution. This applies for instance to the “noisy” Lorenz system.

Finally, in Section 6, we consider an example of a non-Markovian diffusion of Ornstein-Uhlenbeck type with general predictable bounded variance, where we use a method of estimating moments via Burkholder inequalities.

Acknowledgements. Very careful reading of the original manuscript by a referee helped us to make significant corrections and improvements, and lead us to obtain more results.

2 Setting and main questions

2.1 Gaussian concentration bounds: definitions

We denote by $\mathcal{C}_b(\mathbb{R}^d, \mathbb{R})$ the space of bounded continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$. For a probability measure $\mu$ on (the Borel $\sigma$-field of) $\mathbb{R}^d$ and $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R})$, we denote by $\mu(f) = \int f \, d\mu$ the expectation of $f$ with respect to $\mu$. $\text{Lip}(\mathbb{R}^d, \mathbb{R})$ denotes the set of real-valued Lipschitz functions. We further denote for $f \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$

$$\text{lip}(f) := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$$

the Lipschitz constant of $f$, where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. A Lipschitz function is almost surely differentiable by Rademacher’s theorem [17, p. 101], and the supremum norm of the gradient coincides with the Lipschitz constant. For $f : \mathbb{R}^d \to \mathbb{R}$ we denote by $\nabla f$ the gradient of $f$, which we view as a column vector. We denote

$$\|\nabla f\|_\infty := \text{ess sup}_{x \in \mathbb{R}^d} \|\nabla f(x)\|.$$
We can now define the notion of Gaussian concentration bound for two classes of functions.

**DEFINITION 2.1** (Gaussian concentration bounds).

Let \( \mu \) be a Borel probability measure on \( \mathbb{R}^d \).

(a) We say that \( \mu \) satisfies the smooth Gaussian concentration bound if there exists a constant \( D \geq 0 \) such that

\[
\log \mu \left( e^{f - \mu(f)} \right) \leq D \text{lip}(f)^2
\]

for all smooth (i.e., infinitely differentiable) compactly supported \( f: \mathbb{R}^d \to \mathbb{R} \). We abbreviate this property by \( \text{GCBS}(D) \).

(b) We say that \( \mu \) satisfies the Gaussian concentration bound if there exists a constant \( D \geq 0 \) such that

\[
\log \mu \left( e^{f - \mu(f)} \right) \leq D \text{lip}(f)^2
\]

for all Lipschitz functions \( f \in \text{Lip}(\mathbb{R}^d, \mathbb{R}) \). We abbreviate this property by \( \text{GCB}(D) \).

A few remarks are in order. We stress that in the above two definitions the constant \( D \) is independent of the functions \( f \). In [16, Chapter 1], \( \text{GCB}(D) \) is called “normal concentration”, and can be defined for a Borel probability measure on a metric space. In [1], \( \text{GCB}(D) \) is called a sub-Gaussian concentration bound. A more precise definition of a Gaussian concentration bound would consist in requiring \( D \) to be the smallest constant \( C \geq 0 \) such that \( \log \mu \left( e^{f - \mu(f)} \right) \leq C \text{lip}(f)^2 \) for all \( f \) (either smoothly compactly supported or Lipschitz). The case \( D = 0 \) corresponds to Dirac measures which are the “most concentrated” probability measures. In the present paper Dirac measures will be sometimes taken as initial measures.

The next proposition, which we state informally, is useful in some parts of the paper when we cannot deal directly with Lipschitz functions.

**PROPOSITION 2.1.** \( \text{GCBS}(D) \) implies \( \text{GCB}(D) \), with the same constant \( D \).

The precise statement and and the proof are given in appendix B (Lemma B.1) in a more general setting (namely, separable Banach spaces).

**2.2 Equivalence between Gaussian concentration and distance Gaussian moment bounds**

We establish that what we call “distance Gaussian moment” is equivalent to GCB. This will used in the next section, and also in Section 5. This is the only section where we work with a general separable metric space \( (\Omega, d) \). So we first generalize Definition 2.1.
DEFINITION 2.2. Let \( \mu \) be a probability measure on (the Borel \( \sigma \)-field of) \((\Omega, d)\). We say that \( \mu \) satisfies a Gaussian concentration bound with constant \( D > 0 \) on the metric space \((\Omega, d)\) if there exists \( x_* \in \Omega \) such that \( \int d(x_*, x) \, d\mu(x) < +\infty \) and for all \( f \in \text{Lip}(\Omega, \mathbb{R}) \), one has
\[
\int e^{f(x) - \mu(f)} \, d\mu(x) \leq e^{D \text{Lip}(f)^2}.
\]

For brevity we shall say that \( \mu \) satisfies \( \text{GCB}(D) \) on \((\Omega, d)\).

REMARK 2.1.
Note that if there exists \( x_* \in \Omega \) such that \( \int d(x_*, x) \, d\mu(x) < +\infty \) then all Lipschitz functions on \((\Omega, d)\) are \( \mu \)-integrable. Moreover, by the triangle inequality, \( \int d(y, x) \, d\mu(x) < +\infty \) for all \( y \in \Omega \).

THEOREM 2.1. Let \( \mu \) a probability measure on \((\Omega, d)\). Then \( \mu \) satisfies a Gaussian concentration bound if and only if it has a Gaussian moment. More precisely, we have the following:

1. If \( \mu \) satisfies \( \text{GCB}(D) \) then for all \( x_* \in \Omega \) we have
\[
\int e^{\frac{d(x_*, x)^2}{16D}} \, d\mu(x) \leq 3e^{\frac{(\int d(x_*, x) \, d\mu)^2}{8D}}.
\] (2)

2. If there exist \( x_* \in \Omega \), \( a > 0 \) and \( b \geq 1 \) such that
\[
\int e^{ad(x_*, x)^2} \, d\mu(x) \leq b
\] (3)

then \( \mu \) satisfies \( \text{GCB}(D) \) with
\[
D = \frac{b^2 e}{2a\sqrt{\pi}}.
\] (4)

Clearly, if (3) holds for some \( x_*, a, b \), then it holds for any \( \tilde{x} \) with \( a \) replaced by \( 2a \) and \( b \) replaced by \( b \exp(a \text{d}(\tilde{x}, x_*)^2) \), thus GCB holds with \( D \) modified in an obvious way. A very similar result can be found in \[11, Theorem 2.3\]. In Appendix A we provide a different proof which provides more explicit constants.

REMARK 2.2.
Note that one can find a topological space, a probability on the Borel sigma-algebra, and two distances \( d_1 \) and \( d_2 \) defining the topology such that \( \mu \) satisfies GCB on the metric space with \( d_1 \), but it does not on the metric space with \( d_2 \). For example, take \( \mathbb{R} \), \( \mu \) the Gaussian measure, \( d_1 \) the Euclidean distance, and \( d_2(x, y) = \left| \int_y^x (1 + |s|) \, ds \right| \). Then (3) is satisfied if we take \( d_1 \), but (2) is violated if we take \( d_2 \) (which cannot be equivalent to \( d_1 \) since \( d_2 \) is not induced from a norm).

8
2.3 A class of Markov diffusion processes on $\mathbb{R}^d$

We are interested in Markov diffusion processes on $\mathbb{R}^d$, i.e., stochastic processes $(X_t)_{t \geq 0}$ solving a stochastic differential equation (SDE, for short) of the form

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

(5)

where the functions $b$ and $\sigma$ are continuous on $\mathbb{R}_+ \times \mathbb{R}^d$ (with values in $\mathbb{R}^d$ and in the set of $d \times d$ real matrices, respectively), and $(W_t)_{t \geq 0}$ is a standard Brownian motion (or Wiener process) on $\mathbb{R}^d$. Letting

$$a(t, x) = \frac{1}{2} \sigma(t, x) \sigma(t, x)^T$$

(6)

we can re-write (5) as

$$dX_t = b(t, X_t) \, dt + \sqrt{2} a(t, X_t) \, dW_t$$

(7)

where $a : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow M^+_d$ (where $M^+_d$ denotes the set of $d \times d$ symmetric positive definite matrices with entries in $\mathbb{R}^d$, and the symbol $^T$ means the transpose).

Throughout this paper we will always assume (unless explicitly stated otherwise) that the hypotheses of the Corollary of Theorem 2.2 in [19] are satisfied. For the convenience of the reader and for later references we recall these hypotheses.

(H1) For any $T > 0$ and $R > 0$ there exists $C_{T,R} > 0$ depending only on $T$ and $R$ such that, for any $0 \leq t \leq T$, $\|x\| \leq R$, and $\|y\| \leq R$, we have

$$\|b(t, x) - b(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\| \leq C_{T,R} \|x - y\|$$

where $\|\cdot\|$ denotes either the Hilbert-Schmidt norm on matrices, or the Euclidean norm on vectors.

(H2) There exist two nonnegative continuous functions $\alpha$ and $\beta$ on $\mathbb{R}_+$, where $\beta$ is monotone increasing and concave, and satisfies

$$\int_0^\infty \frac{du}{1 + \beta(u)} = \infty.$$  

Moreover, for any $t \in \mathbb{R}_+$ and any $x \in \mathbb{R}^d$, we have

$$2 \langle x, b(t, x) \rangle + \|\sigma(t, x)\|^2 \leq \alpha(t) \, \beta(\|x\|^2)$$

where $\langle \cdot, \cdot \rangle$ denotes Euclidean inner product.

Under these conditions it follows from the Corollary of Theorem 2.2 in [19] that, for any given initial condition in $\mathbb{R}^d$, there exists a pathwise unique solution defined for all times of the SDE (5).
In the cases we will consider, except in Section 6, these hypotheses will apply with the function $\beta(u) = 1 + u$ (see also [3] Hypothesis 2.5.1, and [14]).

In case $a$ (or $\sigma$) and $b$ do not depend on time, we denote by $L$ the generator of the process $(X_t)_{t \geq 0}$ solving the SDE (7), and by $\mathcal{D}(L)$ its domain. It is well-known that this domain contains $C^2(\mathbb{R}^d, \mathbb{R})$ functions which are constant outside a compact subset (see for instance [3]), and the action of the generator on these functions coincides with the action of the partial-differential operator $L$ defined by

$$L = \sum_{i=1}^{d} b_i(x) \partial_i + \sum_{i,j=1}^{d} a_{ij}(x) \partial_i \partial_j$$

where $\partial_i$ denotes partial derivative w.r.t. $x_i$ (and where $a(x) = \sigma(x)\sigma(x)^\top/2$). When $a$ and $b$ depend on time, we still denote by $L$ the partial-differential operator

$$L = \sum_{i=1}^{d} b_i(t, x) \partial_i + \sum_{i,j=1}^{d} a_{ij}(t, x) \partial_i \partial_j,$$

where we have suppressed the dependence on $t$ of $L$, in order not to overload notation.

Now we can formulate the main three questions which we focus on in this paper:

(Q1) If $\mu_0$ is probability measure which satisfies GCB($D_0$), does the distribution $\mu_s$ at some time $s > 0$ of the process $(X_t)_{t \geq 0}$, starting according to $\mu_0$, satisfy GCB($D_s$) for some $D_s$? This question will be settled (under some hypotheses) in Theorem 3.1 and the examples following this theorem show that various possibilities can occur.

(Q2) The main question after Theorem 3.1 is whether $D_s$ exists globally (i.e., is finite for all $s > 0$).

(Q3) If $D_s$ exists globally, is it bounded? In that case, does the stationary measure (or stationary measures) of $(X_t)_{t \geq 0}$ satisfy GCB($D$) for some constant $D$? Can one estimate $D$?

3 Propagation of Gaussian concentration

3.1 A general result

The following theorem settles question (Q1) formulated above.

**Theorem 3.1.** Let $b$ and $\sigma$ satisfy the hypotheses of Section 2.3. Assume moreover that there exists $T \in [0, +\infty]$ (hence $T$ can be infinite) and three continuous (nonnegative) functions $m$, $\alpha$ and $\beta$ on $[0, T)$ such that, for any $0 \leq t < T$, we have
1. sup_{x \in \mathbb{R}^d} \|\sigma(t, x)\| \leq m(t),

2. \langle x, b(t, x) \rangle \leq \alpha(t) \|x\| + \beta(t) \|x\|^2, \forall x \in \mathbb{R}^d.

Then if GCB(D) holds for the initial distribution \(\mu_0\), then it holds for \(\mu_t\) for any \(0 \leq t < T\) with \(D_t < \infty\) (possibly diverging as \(t \uparrow T\)).

**Proof.** We start by defining a function \(c : \mathbb{R}_+ \to \mathbb{R}\) as the solution of the differential equation

\[
\frac{dc}{dt} = -2 c(t) \left( \beta(t) + \alpha(t) + d m(t)^2 \right) - 4 c(t)^2 m(t)^2
\]

starting with \(c(0) = c_0 > 0\). This solution is decreasing in \(t\) and remains strictly positive for \(t < T\). Indeed, if for some \(T > t_0 > 0\), \(1 \geq c(t_0) > 0\), we have \(c(t) \leq 1\) for any \(T > t \geq t_0\), hence

\[
\frac{dc}{dt} \geq -2 c(t) \left( \beta(t) + \alpha(t) + (d + 2) m(t)^2 \right)
\]

which implies for any \(T > t \geq t_0\)

\[
c(t) \geq c(t_0) e^{-2 \int_{t_0}^t (\beta(s) + \alpha(s) + (d + 2) m(s)^2) \, ds} > 0.
\]

Now, let \(\epsilon > 0\) and define

\[
\Phi(t, x, \epsilon) = \exp \left( \frac{c(t) \|x\|^2}{\tau(x)} \right)
\]

where we set \(\tau(x) := 1 + \epsilon \|x\|^2\) to alleviate notation. The function \(\Phi\) is in \(C^0_b([0, T) \times \mathbb{R}^d)\) with bounded first and second differentials in \(x\). For \(0 \leq t < T\), we have

\[
\partial_t \Phi(t, x, \epsilon) + L \Phi(t, x, \epsilon) = \frac{\dot{c}(t) \|x\|^2}{\tau(x)} \Phi(t, x, \epsilon) + \frac{2 c(t) \langle x, b(x, t) \rangle}{\tau(x)^2} \Phi(t, x, \epsilon)
\]

\[
+ \sum_{i,j=1}^d a_{i,j}(t, x) \Phi(t, x, \epsilon) \left( \frac{4 c(t)^2 x_i x_j}{\tau(x)^4} + \frac{2 c(t) \delta_{i,j}}{\tau(x)^2} - \frac{8 c(t) \epsilon x_i x_j}{\tau(x)^3} \right).
\]

Since \(\Phi(t, x, \epsilon) > 0\) and using the conditions on \(b\) and \(\sigma\), \(c(t) \geq 0\) and the
fact that the matrix $a(t, x)$ is nonnegative, we get (where $L$ is defined in (8))

$$\frac{\partial_t \Phi(t, x, \epsilon) + L \Phi(t, x, \epsilon)}{\Phi(t, x, \epsilon)} \leq \frac{\dot{c}(t) \|x\|^2}{\tau(x)} + \frac{2c(t)(\alpha(t) \|x\| + \beta(t)\|x\|^2)}{\tau(x)^2} + \frac{4c(t)^2 m(t)^2 \|x\|^2}{\tau(x)^4}$$

$$+ \frac{2c(t) d \rho(t)^2}{\tau(x)^2}$$

$$\leq \frac{\dot{c}(t) \|x\|^2}{\tau(x)} + \frac{2c(t) \beta(t) \|x\|^2}{\tau(x)} + \frac{4c(t)^2 m(t)^2 \|x\|^2}{\tau(x)}$$

$$+ 2 c(t) \frac{\alpha(t) \|x\|}{\tau(x)^2} + 2 c(t) \frac{d \rho(t)^2}{\tau(x)^2}$$

$$= (\dot{c}(t) + 2 c(t) \beta(t) + 4 c(t)^2 m(t)^2) \frac{\|x\|^2}{\tau(x)}$$

$$+ 2 c(t) \left( \frac{\alpha(t) \|x\|}{\tau(x)^2} + \frac{d \rho(t)^2}{\tau(x)^2} \right).$$

Therefore, letting $\rho(t) := 2 (\alpha(t) + d \rho(t)^2)$, we have

$$\frac{\partial_t \Phi(t, x, \epsilon) + L \Phi(t, x, \epsilon)}{\Phi(t, x, \epsilon)}$$

$$\leq (\dot{c}(t) + 2 c(t) \beta(t) + 4 c(t)^2 m(t)^2 + \rho(t) c(t)) \frac{\|x\|^2}{\tau(x)}$$

$$+ 2 c(t) \left( \frac{\alpha(t) \|x\|}{\tau(x)^2} + \frac{d \rho(t)^2}{\tau(x)^2} - \frac{\rho(t)^2}{2 \tau(x)} \right)$$

$$\leq (\dot{c}(t) + 2 c(t) \beta(t) + 4 c(t)^2 m(t)^2 + \rho(t) c(t)) \frac{\|x\|^2}{\tau(x)}$$

$$+ 2 c(t) \left( \frac{\alpha(t) \|x\| + d \rho(t)^2 - \frac{\rho(t)}{2} \|x\|^2}{\tau(x)} \right).$$

By the very definition of $c(t)$, the first term vanishes. For the second one, we have that, for any $x \in \mathbb{R}^d$ and any $0 \leq t < T$,

$$\frac{2c(t)}{\tau(x)} \left( \alpha(t) \|x\| + d \rho(t)^2 - \frac{\rho(t)}{2} \|x\|^2 \right) \Phi(t, x, \epsilon) \leq c(t) \rho(t) e^{\epsilon(t)}.$$

Indeed, consider first the case $\|x\| \geq 1$. Hence, by definition of $\rho$, the term between parentheses is negative, so the bound is trivially true. When $\|x\| < 1$, we have

$$\alpha(t) \|x\| + d \rho(t)^2 - \frac{\rho(t)}{2} \|x\|^2 \leq \alpha(t) \|x\| + d \rho(t)^2 \leq \rho(t)/2.$$

Hence, using the trivial bound $\tau(x) > 1$, we get the desired bound. Therefore we established that, for $0 \leq t < T$,

$$\partial_t \Phi(t, x, \epsilon) + L \Phi(t, x, \epsilon) \leq c(t) \rho(t) e^{\epsilon(t)}.$$
Let \( x \in \mathbb{R}^d \) be fixed and denote by \( B_R \) the ball centered in \( x \) of radius \( R > 0 \). Let \( T_{BR} \) denote the first exit time from the ball \( B_R \) of the process starting at \( x \).

Since \( \rho(t) c(t) \exp(c(t)) > 0 \), we get using Dynkin’s formula (see [20, Section 7.4])

\[
E_x \left[ \exp \left( \frac{c(t \wedge T_{BR}) X_{t \wedge T_{BR}}^2}{1 + \epsilon X_{t \wedge T_{BR}}^2} \right) \right] \\
\leq e^{c_0 \frac{\|x\|^2}{7(\epsilon)}} + E_x \left( \int_0^{t \wedge T_{BR}} \rho(s) c(s) e^{c(s)} \, ds \right) \\
\leq e^{c_0 \frac{\|x\|^2}{7(\epsilon)}} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds
\]

which is finite for all \( 0 \leq t < T \). Therefore

\[
E_x \left( 1_{\{T_{BR} > t\}} e^{\frac{c(t) X_t^2}{1 + \epsilon X_t^2}} \right) \leq e^{c_0 \frac{\|x\|^2}{7(\epsilon)}} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds.
\]

Since

\[
\sup_{y \in \mathbb{R}^d} e^{\frac{c(t) \|y\|^2}{7(\epsilon)}} \leq e^{\frac{c(t)}{7(\epsilon)}} < \infty
\]

letting \( R \) tend to infinity and using the dominated convergence theorem we get

\[
E_x \left( e^{\frac{c(t) X_t^2}{1 + \epsilon X_t^2}} \right) \leq e^{c_0 \frac{\|x\|^2}{7(\epsilon)}} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds.
\]

Since the integrant in l.h.s. is monotone decreasing in \( \epsilon \) we conclude by the monotone convergence theorem that for any \( x \) and \( 0 \leq t < T \)

\[
E_x \left( e^{c(t) X_t^2} \right) \leq e^{c_0 \|x\|^2} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds. \tag{9}
\]

Now we use part 1 of Theorem 2.1 with \( d \) given by the Euclidean norm \( \|\cdot\| \). If \( \mu_0 \) satisfies GCB(\( D \)), take \( x_* = 0 \) and

\[
c_0 = \frac{1}{16 D}
\]

and by integrating (9) over \( \mu_0 \) we get

\[
E_{\mu_0} \left( e^{c(t) X_t^2} \right) \leq 3 e^{\left( \frac{\|x\| d_{\mu_0}(x)}{8D} \right)^2} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds.
\]

Therefore, by part 2 of Theorem 2.1 we know that \( \mu_t \) satisfies GCB(\( D_t \)) where

\[
D_t = \frac{e}{2 c(t) \sqrt{\pi}} \left( 3 e^{\left( \frac{\|x\| d_{\mu_0}(x)}{8D} \right)^2} + \int_0^t \rho(s) c(s) e^{c(s)} \, ds \right)^2 \tag{10}
\]

which is finite for all \( 0 \leq t < T \). \( \square \)
3.2 Examples

We give examples for initial distributions which are Dirac masses. These distributions trivially satisfy Gaussian concentration (with $D = 0$).

3.2.1 No GCB at any $t > 0$

Consider the stochastic differential equation in $\mathbb{R}^d$

$$dX_t = X_t dW_t.$$ 

The hypotheses of Section 2.3 are satisfied. The solution starting at $x_0$ at time $t = 0$ is

$$X_t = x_0 e^{W_t - \frac{t}{2}}.$$ 

All exponential moments of $X_t$ are infinite, therefore $\mu_t$ cannot satisfy GCB, for any $t > 0$.

3.2.2 GCB for all finite times but $D_t \to \infty$ as $t \to \infty$

Consider the stochastic differential equation in $\mathbb{R}$

$$dX_t = dW_t.$$ 

The hypotheses of Section 2.3 are satisfied. The solution starting at $x_0$ at time $t = 0$ is

$$X_t = x_0 + W_t.$$ 

From Theorem 3.1 we know that GCBS($D_t$) holds for any $t > 0$, with $D_t < \infty$ (which may depend on $x_0$). (Of course, we can apply the second part of Theorem 2.1.) Now, taking the function $f(x) = x$, since the law of $X_t$ is a normal law with mean $x_0$ and variance $t$, we have $D_t \geq t/2$, hence $D_t \to \infty$ as $t \to \infty$.

3.2.3 Blowing up of $D_t$ in finite time

Let $\alpha(t)$ be a $C^1$ non decreasing function such that

$$\alpha(t) = \begin{cases} 
0 & \text{for } 0 \leq t \leq 2 \\
1/3 & \text{for } t \geq 2.
\end{cases}$$

Consider the stochastic differential equation on $\mathbb{R}$

$$dX_t = \left(1 + X_t^2\right)^{\alpha(t)} dW_t.$$ 

The hypotheses of Section 2.3 are satisfied. Therefore for a given initial condition $x_0$ the solution is unique and does not explode at any finite time. It
follows from Theorem 3.1 (applied with $T = 1$) that Gaussian concentration holds for any $0 \leq t < 1$.

We now prove by contradiction that Gaussian concentration cannot hold beyond some $t^* > 0$. Applying stochastic calculus to the function $x \mapsto 1 + x^2$ it follows easily that

$$\sup_{t \geq 0} E_{x_0}(X_t) < \infty.$$  

Observe that if $D_t < \infty$, it follows from Theorem 2.1, part 1, that

$$E_{x_0} \left( e^{\sqrt{1 + X_t^2}} \right) < \infty.$$  

Letting $f(x) = e^{\sqrt{1 + x^2}}$, we have

$$f'(x) = \frac{x}{\sqrt{1 + x^2}} f \quad \text{and} \quad f''(x) = \left( \frac{x^2}{1 + x^2} + \frac{1}{(1 + x^2)^{3/2}} \right) f \geq uf$$

where

$$u = \inf_{x \in \mathbb{R}} \left( \frac{x^2}{1 + x^2} + \frac{1}{(1 + x^2)^{3/2}} \right) = \frac{23}{27}.$$  

Using Itô’s formula we get

$$df(X_t) = \frac{X_t}{\sqrt{1 + X_t^2}} f(X_t) \, dW(t)$$

$$+ \frac{(1 + X_t^2)^{2\alpha(t)}}{2} \left( \frac{X_t^2}{1 + X_t^2} + \frac{1}{(1 + X_t^2)^{3/2}} \right) f(X_t).$$

Therefore

$$E_{x_0} \left( e^{\sqrt{1 + X_t^2}} \right)$$

$$= \int_0^t E_{x_0} \left[ \left( \frac{1 + X_s^2}{2} \right)^{2\alpha(s)} \left( \frac{X_s^2}{1 + X_s^2} + \frac{1}{(1 + X_s^2)^{3/2}} \right) e^{\sqrt{1 + X_s^2}} \right] \, ds$$

$$\geq \frac{u}{2} \int_0^t E_{x_0} \left[ \left( \sqrt{1 + X_s^2} \right)^{4\alpha(s)} e^{\sqrt{1 + X_s^2}} \right] \, ds$$

$$\geq \frac{u}{2} \int_0^t E_{x_0} \left[ e^{\sqrt{1 + X_s^2}} \right] \left( \log E_{x_0} \left[ e^{\sqrt{1 + X_s^2}} \right] \right)^{4\alpha(s)} \, ds$$

where the last inequality follows from Jensen’s inequality applied to the random variable $Z(s)(\log Z(s))^{4\alpha(s)}$ with $Z(s) := e^{\sqrt{1 + X_s^2}}$, for each $s$.

Now consider the ordinary differential equation on $\mathbb{R}_+$

$$\frac{dy}{dt} = \frac{u}{2} y \left( \log(y) \right)^{4\alpha(t)}, \quad y(0) = e^{\sqrt{1 + x^2}}.$$  

We have

$$E_{x_0} \left[ e^{\sqrt{1 + X_t^2}} \right] \geq y(t).$$
This solution \( y(t) \) blows up in finite time. The proof is by contradiction. A solution with \( y(0) > 1 \) is monotonically increasing. If \( y(t) \) is forever finite we have for \( t \geq 2 \)

\[
\frac{dy}{dt} = \frac{u}{2} \ y \ (\log(y))^{1+\alpha(t)}.
\]

We get a contradiction by Osgood’s criterion [21] since for \( t \geq 2 \) we have

\[
\int_{y(2)}^{\infty} \frac{dz}{z (\log(z))^{1/3}} = \frac{3}{(\log(y(2)))^{1/3}} < \infty.
\]

4 Nonlinear semigroup, gradient bounds, and coupling

In some sense Theorem 3.1 gives an answer to the basic questions about preservation in time of Gaussian concentration bounds, but we have little information on \( D_t \), when it is finite (see (10)). The goal of this section is to obtain a more explicit \( D_t \), and also to possibly obtain GCB for the stationary measure, when there is one, by taking \( t \to +\infty \). We restrict to time-homogeneous Markovian diffusions, i.e., the drift and diffusion matrix do not depend on time, so (5) takes the form

\[
dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t
\]

where otherwise the \( b \) and \( \sigma \) satisfy the assumptions in Section 2.3.

We develop an abstract approach based on the so-called nonlinear semigroup in the next two subsections. Then, we first combine it with the Bakry-Emery \( \Gamma_2 \) criterion in the reversible context. We show that if the strong gradient bound is satisfied, then the Gaussian concentration bound is conserved in the course of the time evolution, and in the limit \( t \to +\infty \). Second, we develop a coupling approach to treat non-reversible degenerate situations (which are out of reach of the first approach).

Before going on, we recall that we can look for Gaussian concentration bounds for smooth compactly supported functions, which will automatically give Gaussian concentration bounds for Lipschitz functions by virtue of Proposition 2.1.

4.1 The nonlinear semigroup

Let \( (X_t)_{t \geq 0} \) be a Markov diffusion process on \( \mathbb{R}^d \) as defined in (7), with drift and diffusion matrix not depending on time. Denote by \( (S_t)_{t \geq 0} \) its semigroup acting on \( \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}) \). As usual, the generator is denoted by

\[
\mathcal{L} f(x) = \lim_{t \downarrow 0} \frac{S_t f(x) - f(x)}{t}
\]
on its domain $\mathcal{D}(\mathcal{L})$ of functions $f$ such that $(S_t f(x) - f(x))/t$ converges uniformly in $x$ when $t \downarrow 0$. The nonlinear semigroup is denoted by

$$V_t(f) = \log S_t(e^f).$$

This is indeed a semigroup since

$$V_{t+s}(f) = \log (S_{t+s}(e^f)) = \log S_t(S_s(e^f)) = \log S_t(\log e^{V_s(f)}) = V_t(V_s(f)).$$

We denote by $\mathcal{H}$ its generator, i.e., for all $x \in \mathbb{R}^d$,

$$\mathcal{H}(f)(x) = \lim_{t \downarrow 0} \frac{V_t(f)(x) - f(x)}{t} \quad (11)$$

defined on the domain $\mathcal{D}(\mathcal{H})$ where the defining limit in (11) converges uniformly. The relation between $\mathcal{H}$ and $V_t$ is more subtle than the relation between $\mathcal{L}$ and $S_t$. We will restrict ourselves to the case of diffusions on $\mathbb{R}^d$, although what follows can be formulated in a more abstract setting. Thanks to the approximation results found in Appendix B, it is enough to restrict ourselves to adequate subsets of the domains $\mathcal{D}(\mathcal{L})$ and $\mathcal{D}(\mathcal{H})$. Denote by $C^\infty_c(\mathbb{R}^d, \mathbb{R})$ the space of infinitely differentiable real-valued functions on $\mathbb{R}^d$ with compact support.

**Proposition 4.1.** The following properties hold:

1. $C^\infty_c(\mathbb{R}^d, \mathbb{R}) \subset \mathcal{D}(\mathcal{L})$ and constant functions also belong to $\mathcal{D}(\mathcal{L})$;

2. $C^\infty_c(\mathbb{R}^d, \mathbb{R}) \subset \mathcal{D}(\mathcal{H})$, and for $f \in C^\infty_c(\mathbb{R}^d, \mathbb{R})$ we have

   $$\mathcal{H}(f) = e^{-f} \mathcal{L} e^f;$$

3. $\forall f \in C^\infty_c(\mathbb{R}^d, \mathbb{R})$, $V_t(f) \in \mathcal{D}(\mathcal{H})$ for each $t \geq 0$, and we have

   $$\frac{dV_t(f)}{dt} = \mathcal{H}(V_t(f)).$$

**Proof.** The first statement is well-known, see for instance [1]. In order to prove the second one, we first observe that $\exp(-\|f\|_\infty) \leq S_t(\exp(f)) \leq \exp(\|f\|_\infty)$ and $\exp(f) \in \mathcal{D}(\mathcal{L})$ (see again [1]). Now the statement follows from the definition of $\mathcal{H}$. The last two statements are proved as follows. From the semigroup property of $(V_t)_{t \geq 0}$ and $(S_t)_{t \geq 0}$, for each $t > 0$ and $|\epsilon| < t$, we have

$$\frac{V_\epsilon(V_t(f)) - V_t(f)}{\epsilon} = \frac{1}{\epsilon} \log \left( \frac{S_\epsilon(S_t(e^f))}{S_t(e^f)} \right). \quad (12)$$
Moreover, since \( e^f = 1 + \tilde{f} \) with \( \tilde{f} \in C_c^\infty(\mathbb{R}^d, \mathbb{R}) \subset D(\mathcal{L}) \), we obtain \( S_t(\exp(f)) \in D(\mathcal{L}) \) for each \( t \geq 0 \). Therefore

\[
S_t \left( S_t(e^f) \right) = S_t(e^f) + \varepsilon \mathcal{L} S_t(e^f) + o(\varepsilon)
\]

uniformly. The explicit formula for the limit in (12) gives the last statement.

\[\square\]

### 4.2 Some preparatory computations

We now show how the nonlinear semigroup naturally enters the picture. For all \( t \geq 0 \), we have

\[
\mu_t \left( e^f - \mu_t(f) \right) = \mu_0 \left( S_t(e^f) \right) e^{-\mu_0(S_t(f))}
\]

\[
= \mu_0 \left( e^{V_t(f) - \mu_0(V_t(f))} \right) e^{\mu_0(V_t(f) - S_t(f))}.
\]

Therefore, if \( \mu_0 \) satisfies GCBS(\( D_0 \)), then we can estimate the first factor in the r.h.s. of (13)

\[
\mu_0 \left( e^{V_t(f) - \mu(V_t(f))} \right) \leq e^{D_0 \text{lip}(V_t(f))^2}
\]

and so we have to estimate \( \text{lip}(V_t(f)) \), which in the case of diffusion processes will boil down to estimating \( \nabla V_t(f) \). Concerning the second factor in (13) we define first the “truly nonlinear” part of the nonlinear generator as follows

\[
\mathcal{H}_{nl}(f) = \mathcal{H}(f) - \mathcal{L}(f)
\]

for \( f \in D(\mathcal{L}) \cap D(\mathcal{H}) \). In the case of diffusion processes, this operator exactly contains the quadratic term of \( \mathcal{H} \), which coincides in turn with the so-called “carré du champ operator” (see Section 4.3 below).

We have the following proposition.

**Proposition 4.2.** If \( \partial_t - \mathcal{L} \) is a hypoelliptic diffusion on \( \mathbb{R}^d \) with \( C^\infty \) coefficients, then

\[
\mathcal{H}_{nl}(f) = \Gamma(f) = \frac{1}{2} \mathcal{L}(f^2) - f \mathcal{L}(f).
\]

Moreover, assume that, for any \( f \in C_c^\infty(\mathbb{R}^d, \mathbb{R}) \), and any \( T > 0 \), we have

\[
\sup_{0 \leq t \leq T} \| \Gamma(V_t(f)) \|_{C^0(\mathbb{R}^d, \mathbb{R})} < \infty.
\]

Then for any probability measure \( \mu_0 \) on \( \mathbb{R}^d \), and for all \( t \geq 0 \), we have

\[
\mu_0(\| V_t(f) - S_t(f) \|) \leq \| V_t(f) - S_t(f) \|_{\infty} \leq \int_0^t \| \Gamma(V_s(f)) \|_{\infty} \, ds.
\]

(15)
PROOF. It follows from Proposition 4.1 that
\[
\frac{d(V_t(f) - S_t(f))}{dt} = \mathcal{H}(V_t(f)) - \mathcal{L}S_t(f) \\
= \mathcal{H}(V_t(f)) - \mathcal{L}V_t(f) + \mathcal{L}(V_t(f) - S_t(f)) \\
= \mathcal{H}_n(V_t(f)) + \mathcal{L}(V_t(f) - S_t(f)).
\]

As a consequence, we obtain by the variation-of-constant method
\[
V_t(f) - S_t(f) = \int_0^t S_{t-s}(\mathcal{H}_n(V_s(f))) \, ds.
\]

Notice that the use of this method needs appropriate justification (for instance that \(V_t(f)\) belongs to the domain of \(\mathcal{L}\)). See also [1] pages 144-145 for comments about such worries. We now provide full details. It will be convenient to distinguish between the generator \(\mathcal{L}\) defined before and the associated second order differential operator denoted by \(\hat{\mathcal{L}}\). For \(f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})\) we have \(\mathcal{L}f = \hat{\mathcal{L}}f\) and for any \(t \geq 0\)
\[
\frac{d}{dt}S_t(f) = \mathcal{L}S_t(f) = \hat{\mathcal{L}}S_t(f).
\]

We observe that for \(f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})\)
\[
e^f = 1 + g
\]
with \(g \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})\) and
\[
e^f \geq e^{-\|f\|_\infty} > 0.
\]

In particular \(V_t(f) = \log(1 + S_t(g))\). By hypoellipticity it follows that \(V_t(f)(x)\) (as well as \(S_t(f)(x)\)) is \(\mathcal{C}_c^\infty\) in \(t\) and \(x\). We have for \(s \geq 0\)
\[
\partial_s V_s(f)(x) = \frac{\mathcal{L}S_s(g)(x)}{1 + S_s(g)(x)} = \frac{\hat{\mathcal{L}}S_s(g)(x)}{1 + S_s(g)(x)} = \hat{\mathcal{L}}V_s(f)(x) + \Gamma(V_s(f))(x).
\]

In particular for \(s = 0\) we get \(\mathcal{H}f = \mathcal{L}f + \Gamma(f)\), which is the first statement. For \(f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})\), we define a real function \(\Lambda(t, x)\) on \(\mathbb{R}^{d+1}\) by
\[
\Lambda(t, x) = \begin{cases} 
V_t(f)(x) - S_t(f)(x) - \int_0^t S_{t-s}(\Gamma(V_s(f))) \, ds & \text{if } t \geq 0 \\
0 & \text{if } t < 0.
\end{cases}
\]

Note that this function is continuous in \(t\) and \(x\), bounded on the set \([0, T] \times \mathbb{R}^d\) for any \(T > 0\) and satisfies \(\Lambda(0, x) = 0\) for all \(x \in \mathbb{R}^d\). We will now prove that \(\partial_t \Lambda - \hat{\mathcal{L}}\Lambda = 0\) in the sense of distributions. Observe that for fixed \(s \geq 0\), the function \(S_{t-s}(\Gamma(V_s(f)))\) is defined through Theorem 2.2.5 in [3]
(see also Sections 2.4 and 2.5 therein). Let \( u \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}) \). We have by Fubini’s Theorem
\[
\int \int (\partial_t u - \hat{L}^\dagger u)(t, x) \Lambda(t, x) \, dt \, dx
= \int \int (\partial_t u - \hat{L}^\dagger u)(t, x) (V_t(f)(x) - S_t(f)(x)) \, dt \, dx
- \int ds \int_s^\infty \int (\partial_t u - \hat{L}^\dagger u)(t, x) S_{t-s}(\Gamma(V_s(f))) \, dt \, dx.
\]
By the definition of derivatives in the sense of distributions, denoting \( \hat{L}^\dagger \) the adjoint of \( \hat{L} \), we obtain
\[
\int \int (\partial_t u - \hat{L}^\dagger u)(t, x) (V_t(f)(x) - S_t(f)(x)) \, dt \, dx
= -\int \int u(t, x) \Gamma(V_t(f))(x) \, dt \, dx
\]
and for each \( s \)
\[
\int_s^\infty \int (\partial_t u - \hat{L}^\dagger u)(t, x) S_{t-s}(\Gamma(V_s(f))) \, dt \, dx
= -\int u(s, x) \Gamma(V_s(f))(x) \, dx.
\]
Therefore, we have
\[
\int \int (\partial_t u - \hat{L}^\dagger u)(t, x) \Lambda(t, x) \, dt \, dx = 0
\]
and since this holds for any \( u \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}) \), we have
\[
\partial_t \Lambda - \hat{L} \Lambda = 0
\]
in the sense of distributions. By hypoellipticity, this also holds in the sense of functions.

Since the function \( \beta \) occurring in hypothesis (H2) is concave, for any \( u > 0 \) we have \( \beta(2u) \leq 2\beta(u) - \beta(0) \). Note that from our hypothesis (H2) we have \( \beta(0) > 0 \) since \( \sigma(0) \neq 0 \). Therefore for any \( u \geq 1 \), \( \beta(u) \leq 2u\beta(1) \) and from the monotonicity of \( \beta \) we get for any \( u \geq 0 \)
\[
\beta(u) \leq (2u + 1)\beta(1).
\]
Let
\[
\varphi(x) = \log \left( 1 + \|x\|^2 \right).
\]
It is left to the reader to check that the above bound on \( \beta \) implies that there exists \( \lambda_0 > 0 \) and \( C > 0 \) such that for any \( x \in \mathbb{R}^d \)
\[
\hat{L}\varphi - \lambda_0\varphi \leq C.
\]
Since $\varphi(x)$ diverges with $\|x\|$, it follows from Theorem 4.1.3 (iii) in [3] that $\Lambda = 0$ on $\mathbb{R} \times \mathbb{R}^d$ since $\Lambda$ is bounded on $[0, T] \times \mathbb{R}^d$ for any $T > 0$.

Because $(S_t)_{t \geq 0}$ is a Markov semigroup, it is a contraction semigroup in the supremum norm and because $\mu_0$ is a probability measure, we obtain the desired inequality.

Assuming that $\mu_0$ satisfies GCBS($D_0$), when we combine (13), (14) and (15), we obtain, for all $t \geq 0$,

$$\mu_t \left( e^{\mu t} \right) \leq \exp \left( D_0 \text{lip}(V_t(f))^2 + \int_0^t \| \Gamma(V_s(f)) \|_{\infty} \mathrm{d}s \right). \quad (16)$$

In particular, in the case of diffusion processes on $\mathbb{R}^d$, $\Gamma(g)$ is bounded in terms of $(\nabla g)^2$. Hence, if we have estimates for $\text{lip}(V_t(f))$ and $\nabla V_t(f)$, we can plug them in immediately.

### 4.3 Abstract gradient bound approach

In this section we study the questions formulated in Section 2.3 in the context of Markovian diffusion triples, in the sense of [1], i.e., reversible diffusion processes for which we have the integration by parts formula relating the Dirichlet form and the carré du champ bilinear form. Let $(X_t)_{t \geq 0}$ be a Markov diffusion, i.e., a solution of the SDE of the form (7). Moreover, we will assume in this subsection that the covariance matrix $a(x)$ is not degenerate, and is bounded, uniformly in $x \in \mathbb{R}^d$, i.e., for some $C_1, C_2 > 0$,

$$C_1^{-2} \|v\|^2 \leq \langle v, a(x)v \rangle \leq C_2^2 \|v\|^2. \quad (17)$$

To the generator $L$ is associated the carré du champ bilinear form $\Gamma(\cdot, \cdot)$ on $C^\infty(\mathbb{R}^d, \mathbb{R})$ functions which are constant outside a compact set given by

$$\Gamma(f, g) = \frac{1}{2} \left( L(fg) - gL(f) - fL(g) \right) = \langle \nabla f, a \cdot \nabla g \rangle. \quad (18)$$

Notice that $\Gamma$ satisfies the so-called diffusive condition, i.e., for all smooth functions $\psi : \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}^d$,

$$\Gamma(\psi(f), \psi(f))(x) = (\psi')^2(f(x)) \Gamma(f, f)(x). \quad (19)$$

We will further assume that there exists a reversible measure $\nu$ such that the integration by parts formula

$$\int f(-Lg) \mathrm{d}\nu = \int \Gamma(f, g) \mathrm{d}\nu$$

holds. The triple $(\mathbb{R}^d, \Gamma, \nu)$ is then a Markov diffusion triple in the sense of [1, Section 3.1.7].
The second-order carré du champ bilinear form is given by
\[ \Gamma_2(f,g) = \frac{1}{2}(\mathcal{L}\Gamma(f,g) - \Gamma(\mathcal{L}f,g) - \Gamma(f,\mathcal{L}g)). \]
In what follows, we abbreviate, as usual, \( \Gamma(f,f) =: \Gamma(f) \), \( \Gamma_2(f,f) = \Gamma_2(f) \).
An important example is when \( b = -\nabla U \) and \( a = I_d \), in which case the second-order carré du champ bilinear form is given by
\[ \Gamma_2(f,f) = \|\nabla^2 f\|^2 + \langle \nabla f, \nabla^2 U(\nabla f) \rangle \]
where \( \nabla^2 U \) denotes the Hessian of \( U \), i.e., the matrix of the second derivatives. By the non-degeneracy and boundedness condition (17), we have, for all \( x \in \mathbb{R}^d \)
\[ C_1^{-2}\|\nabla f(x)\|^2 \leq \Gamma(f)(x) \leq C_2\|\nabla f(x)\|^2. \tag{20} \]
Following [1] we say that the strong gradient bound is satisfied with constant \( \rho \in \mathbb{R} \) if for all \( t \in \mathbb{R}_+ \)
\[ \sqrt{\Gamma(S_t f)} \leq e^{-\rho t} S_t(\sqrt{\Gamma(f)}). \tag{21} \]
This condition is fulfilled when, e.g., the Bakry-Emery curvature bound,
\[ \Gamma_2(f) \geq \rho \Gamma(f) \]
is satisfied. We refer to [1, Chapter 3] for the proof and more background on this formalism. We then have the following general result.

**Theorem 4.1.** Let \( (X_t)_{t \geq 0} \) be a reversible diffusion process such that (21) is fulfilled. Assume that \( \mu_0 \) satisfies GCBS(\( D \)). Then, for every \( t \geq 0 \), \( \mu_t \) satisfies GCBS(\( D_t \)) with
\[ D_t = D C_1^2 C_2^2 e^{-2\rho t} + \frac{C_1^2 C_2^4}{2\rho} (1 - e^{-2\rho t}) \tag{22} \]
Hence, in particular, if \( \rho > 0 \), then the unique reversible measure \( \nu \) satisfies GCBS(\( D_\infty \)) with \( D_\infty = \frac{C_1^2 C_2^4}{2\rho} \).

**Proof.** Using (21) we start by estimating \( \|\nabla V_t f\| \) for \( f : \mathbb{R}^d \to \mathbb{R} \) which is \( C^\infty \) with compact support
\[
\|\nabla V_t(f)\| = \frac{\|\nabla(S_t(e^f))\|}{S_t(e^f)} \leq C_1 \frac{\sqrt{\Gamma(S_t(e^f))}}{S_t(e^f)} \]
\[
\leq C_1 e^{-\rho t} \frac{S_t(\sqrt{\Gamma(e^f)})}{S_t(e^f)} = C_1 e^{-\rho t} \frac{S_t(e^f \sqrt{\Gamma(f)})}{S_t(e^f)} \]
\[
\leq C_1 e^{-\rho t} \|\sqrt{\Gamma(f)}\|_\infty \leq C_1 C_2 e^{-\rho t} \|\nabla f\|_\infty .
\]
The first inequality comes from (20). The second comes from (21). The second equality is the diffusive condition (19). The third inequality follows by

\[ \frac{|S_t(e^f g)|}{S_t(e^f)} \leq \frac{S_t(e^f |g|)}{S_t(e^f)} \leq \|g\|_\infty \]

where we used that \((S_t)_{t \geq 0}\) is a Markov semigroup, and the fourth inequality follows from (20). As a consequence we obtain

\[ \text{lip}(V_t(f)) = \|\nabla V_t f\|_\infty \leq C_1 C_2 e^{-\rho t} \|\nabla f\|_\infty. \] (23)

Using Proposition 4.2, (20) and (23), we obtain

\[ \|V_t(f) - S_t(f)\|_\infty \leq \int_0^t \|\Gamma(V_s(f))\|_\infty ds \]
\[ \leq C_2^2 \int_0^t \|\nabla V_s(f)\|_\infty^2 ds \]
\[ \leq C_1^2 C_2^2 \|\nabla f\|_\infty^2 \int_0^t e^{-2\rho s} ds. \] (24)

Combining (23), (25) with (16) we obtain that \(\mu_t\) satisfies GCBS\((D_t)\) with

\[ D_t = D C_1^2 C_2^2 e^{-2\rho t} + C_1^2 C_2^4 \int_0^t e^{-2\rho s} ds \]

which is the claim of the theorem. □

**Remark 4.1.**

(a) In case \(\Gamma(f) = a^2 \|\nabla f\|^2\), we can take \(C_1 = a^{-2}, C_2 = a^2\), so \(D_0 = D\). In general \(C_1^2 C_2^2 > 1\), which means that at time \(t = 0\) we do not recover the constant \(D\) in (22), but a larger constant. This is an artifact of the method where we estimate the norm of the gradient via the carré du champ.

(b) In case we have an exact commutation relation of the type

\[ \nabla S_t(f) = e^{-\rho t} S_t(\nabla f) \]

such as is the case for the Ornstein-Uhlenbeck process, we obtain directly

\[ \|\nabla V_t(f)\| \leq e^{-\rho t} \|\nabla f\|_\infty \]

i.e., without using the bilinear form \(\Gamma\).

**4.4 Coupling approach**

In this section we develop a coupling approach to obtain a bound on the Lipschitz constant of \(V_t(f)\) for \(f \in \mathcal{C}_0^\infty\), allowing to apply Proposition 4.2. This approach will work in some nonreversible situations and also for some degenerate diffusions, where the approach of Section 4.3 fails.
4.4.1 Coupling and the nonlinear semigroup

In the previous section, the essential input coming from the strong gradient bound is the estimate (23) which implies that for all \( x, y \in \mathbb{R}^d \) and all \( t \in \mathbb{R}_+ \)

\[
\|V_t(f)(x) - V_t(f)(y)\| \leq C_t \|\nabla f\|_\infty \|x - y\| e^{-\rho t}.
\]

(26)

Once we have the bound (26), we can use it to further estimate the r.h.s. of (15), provided we have a control on \( \mathcal{H}_{nl} \). Instead of starting from the curvature bound, in this subsection we start from a coupling point of view. This has the advantage that reversibility is no longer necessary, and moreover we can include degenerate diffusions such as the Ginzburg-Landau diffusions (see below). We denote by \( X^x_t \) the process \((X^x_t)_{t \geq 0}\) solving the SDE (5) started at \( X_0 = x \) (with \( b \) and \( \sigma \) not depending explicitly on time, as assumed for the whole section).

As an important example to keep in mind, consider the Ornstein-Uhlenbeck process on \( \mathbb{R}^d \), with generator

\[-\langle Ax, \nabla \rangle + \Delta\]

where \( \Delta \) denotes the Laplacian in \( \mathbb{R}^d \), and where \( A \) is a \( d \times d \) matrix. In that case we have

\[
X^x_t = e^{-At} x + \int_0^t e^{-2A(t-s)} \, dW_s
\]

(27)

which depends deterministically, and in fact linearly, on \( x \).

**Definition 4.1.** Let \( \gamma : [0, +\infty) \rightarrow [0, +\infty) \) be a measurable function such that \( \gamma(0) = 1 \). We say that the process \((X^x_t)_{t \geq 0}\) can be coupled at rate \( \gamma \) if, for all \( x, y \in \mathbb{R}^d \), there exists a coupling of \((X^x_t)_{t \geq 0}\) and \((X^y_t)_{t \geq 0}\) such that almost surely in this coupling

\[
d(X^x_t, X^y_t) \leq d(x, y) \gamma(t)
\]

(28)

where \( d \) denotes the Euclidean distance.

In the case of the Ornstein-Uhlenbeck process in \( \mathbb{R}^d \), we have from (27) (which implicitly defines a coupling, because we use (27) for all \( x \) with the same Brownian realization)

\[
\|X^x_t - X^y_t\| \leq \|e^{-At}\| \|x - y\|
\]

hence \( \gamma(t) = \|e^{-At}\| \). Notice that \( \gamma(t) \) can be “expanding” or “contracting”, depending on the spectrum of \( A \). More precisely, \( \gamma \) will be eventually contracting if the numerical range of \( A \) lies in the half-plane of complex numbers with non-positive real part.
Remark 4.2. In the context of Brownian motion on a Riemannian manifold, it is proved in [22] that (28) for $\gamma(t) = e^{-Kt/2}$ is equivalent with having $K$ as a lower bound for the Ricci curvature, which in that context is equivalent with the Bakry-Emery curvature bound. In general however, the relation between the coupling condition (28) and the Bakry-Emery curvature bound is not so simple. In particular, the coupling approach applies beyond reversibility, in the context of degenerate diffusions, and beyond the setting of exponential decay of $\gamma(t)$ in (28).

We have the following result. Let $W_1$ be the space of probability measures $\mu$ such that $\int d(0, x) d\mu(x) < +\infty$ equipped with the distance

$$d_{W_1}(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu : \text{lip}(f) \leq 1 \right\}$$

$$= \inf \left\{ \int d(x, y) dP : P \text{ coupling of } \mu, \nu \right\}$$

(29)

where the second equality follows from the Kantorovich-Rubinstein theorem [9, Theorem 11.8.2, p. 420].

Theorem 4.2. Assume that $(X_t)_{t \geq 0}$ can be coupled at rate $\gamma$, where $\gamma$ is square integrable on compacts subsets of $\mathbb{R}_+$. Assume that $\mu_0$ satisfies GCBS($D$). Then for all $t \geq 0$, and for all $f$ smooth with compact support we have the estimate

$$\log \mu_t(e^{f} - \mu_t(f)) \leq D \text{lip}(f)^2 \gamma(t)^2 + C_2^2 \text{lip}(f)^2 \int_0^t \gamma(s)^2 ds$$

(30)

where $C_2$ is defined in (17). As a consequence, for every $t > 0$, $\mu_t$ satisfies GCBS($D_t$) with

$$D_t = D \gamma(t)^2 + C_2^2 \int_0^t \gamma(s)^2 ds.$$  

(31)

In particular, if $\int_0^{+\infty} \gamma(s)^2 ds < \infty$, then every weak limit point of $\{\mu_t, t \geq 0\}$ satisfies GCBS($D_\infty$) with

$$D_\infty = C_2^2 \int_0^{+\infty} \gamma(s)^2 ds.$$  

Moreover, the stationary probability measure $\nu \in W_1$ satisfies GCBS($D_\infty$).

Before giving the proof of this theorem, we establish two lemmas. The first one provides a general estimate on the variation of $V_t f$.

Lemma 4.1. Let $f$ be Lipschitz and assume that $(X_t)_{t \geq 0}$ can be coupled at rate $\gamma$. Then for all $t \geq 0$ and $x, y \in \mathbb{R}^d$ we have

$$V_t(f)(x) - V_t(f)(y) \leq \text{lip}(f) \gamma(t) d(x, y).$$

As a consequence, for all $t \geq 0$,

$$\text{lip}(V_t(f)) \leq \text{lip}(f) \gamma(t).$$
Let us denote by $\hat{E}$ expectation in the coupling of $(X^x_t)_{t \geq 0}$ and $(X^y_t)_{t \geq 0}$ for which (28) holds (which exists by assumption). Then we have

$$
\exp(V_t(f)(x) - V_t(f)(y)) = \frac{\hat{E}\left(e^{f(X^x_t)}\right)}{\hat{E}\left(e^{f(X^y_t)}\right)} = \frac{\hat{E}\left(e^{f(X^x_t)} e^{\text{lip}(f) d(X^x_t, X^y_t)}\right)}{\hat{E}\left(e^{f(X^y_t)}\right)} \\
\leq \frac{\hat{E}\left(e^{f(X^y_t)} e^{\text{lip}(f) d(x,y) \gamma(t)}\right)}{\hat{E}\left(e^{f(X^x_t)}\right)} \\
\leq \frac{e^{\text{lip}(f) d(x,y) \gamma(t)}}{\hat{E}\left(e^{f(X^x_t)}\right)}
$$

where in the last inequality we used (28).

Notice that in lemma 4.1 it is not required that $\gamma(t) \to 0$ as $t \to +\infty$, i.e., the coupling does not have to be successful. However if one wants to pass to the limit $t \to +\infty$ then it is important that $\gamma(t) \to 0$ as $t \to +\infty$. This in turn implies, as we see in the next lemma that, among all probability measures in the Wasserstein space $W_1$, there is a unique invariant probability measure $\nu$, and for all $\mu_0 \in W_1$, $\mu_t \to \nu$ weakly as $t \to +\infty$.

**Lemma 4.2.** Assume that $(X_t)_{t \geq 0}$ can be coupled at rate $\gamma$ and $\gamma(t) \to 0$ as $t \to +\infty$. Then there exists a unique invariant probability measure $\nu$ in $W_1$. Moreover, for all $\mu_0 \in W_1$, $\mu_t \to \nu$ as $t \to +\infty$.

**Proof.** Let $\mu_0$, $\nu_0$ be elements of $W_1$ and let $f$ be a Lipschitz function with $\text{lip}(f) \leq 1$. Because $\mu_0$, $\nu_0$ are elements of $W_1$, there exists a coupling $\mathbb{P}$ such that

$$
\int d(x,y) \, d\mathbb{P}(x,y) = d_{W_1}(\mu_0, \nu_0) < +\infty.
$$

The infimum in definition (29) is attained, see [9, Theorem 11.8.2, p.420]. Then

$$
\int f \, d\mu_t - \int f \, d\nu_t = \int \hat{E}\left(f(X^x_t) - f(X^y_t)\right) \, d\mathbb{P}(x,y) \\
\leq \int \hat{E}\left(d(X^x_t, X^y_t)\right) \, d\mathbb{P}(x,y) \\
\leq \gamma(t) \, d_{W_1}(\mu_0, \nu_0).
$$

This shows that for all $\mu_0, \nu_0 \in W_1$, and for all $t \geq 0$,

$$
d_{W_1}(\mu_t, \nu_t) \leq \gamma(t) \, d_{W_1}(\mu_0, \nu_0).
$$

(32)
Existence of an invariant measure $\nu \in \mathcal{W}_1$ now follows via a standard contraction argument. If $\mu, \nu \in \mathcal{W}_1$ are both invariant then (32) gives, after taking $t \to +\infty$: $d_{\mathcal{W}_1}(\mu, \nu) = 0$, which shows uniqueness of the invariant measure $\nu \in \mathcal{W}_1$. The fact that $\mu_0 \in \mathcal{W}_1$, implies $\mu_t \to \nu$ as $t \to +\infty$ then also follows from (32).

**Proof of Theorem 4.2.** By Proposition 4.2 (and by Rademacher’s theorem recalled above) we get for any $f : \mathbb{R}^d \to \mathbb{R}$ which is $C^\infty$ with compact support

$$\mathcal{H}_{al}(f) = \Gamma(f) \leq C_2^2 \|\nabla f\|_\infty^2 \leq C_2^2 \text{lip}(f)^2.$$  

Hence, using (16) and Lemma 4.1, this establishes (30) and (31). The last statement follows from Lemma 4.2.

As an application we have the following result on Markovian diffusions with covariance matrix $a$ not depending on the location $x$. We will prove later on a generalization of this result by using different techniques (Theorem 5.3). Nevertheless, the following theorem provides a better bound for $D_t$.

**Theorem 4.3.** Let $(X_t)_{t \geq 0}$ denote a diffusion process on $\mathbb{R}^d$ with generator of type (8), and where the drift does not depend on time, and where the covariance matrix $a$ does not depend on time and location $x,t$. Let $\|a\|$ denote Euclidean norm on matrices. Assume furthermore that the function $b : \mathbb{R}^d \to \mathbb{R}^d$ is continuously differentiable and the differential $D_x b$ satisfies the estimate

$$\langle D_x b(x)(u), u \rangle \leq -\kappa \|u\|^2 \quad (33)$$

for all $x, u \in \mathbb{R}^d$ and some $\kappa \in \mathbb{R}$. Let $\mu_0$ satisfy GCBS($D_0$), then, for all $t > 0$, $\mu_t$ satisfies GCBS($D_t$) with

$$D_t = D_0 e^{-2\kappa t} + \frac{\|a\|}{2\kappa} \left(1 - e^{-2\kappa t}\right). \quad (34)$$

Moreover, if $\kappa > 0$, then $\mu_t \to \nu$ as $t \to +\infty$ where $\nu$ is the unique invariant probability measure, and it satisfies GCBS($\|a\|/(2\kappa)$). In particular, if $b = -\nabla U$, where the potential $U : \mathbb{R}^d \to \mathbb{R}$ is $C^2$, then (33) reduces to the convexity condition

$$\langle \nabla \nabla U u, u \rangle \geq \kappa \|u\|^2.$$  

**Proof.** We have $\|\mathcal{H}_{al}(f)\| = \Gamma(f) \leq \|a\|(\nabla f)^2$. Therefore by Theorem 4.2 it suffices to see that we have a coupling rate $\gamma(t) = e^{-\kappa t}$. We couple $X_t^x, X_t^y$ by using the same realization of the underlying Brownian motion $(W_t)_{t \geq 0}$, and as a consequence, because $a$ does not depend on $x$, the difference between $X_t^x$ and $X_t^y$ is evolving according to

$$\frac{d}{dt}(X_t^x - X_t^y) = b(X_t^x) - b(X_t^y).$$

27
We have
\[ b(X^x_t) - b(X^y_t) = \int_0^1 \frac{d}{ds}(b(sX^x_t + (1 - s)X^y_t)) \, ds \]
\[ = \int_0^1 D_x b(\xi(s))(X^x_t - X^y_t) \, ds \]
with
\[ \xi(s) = sX^x_t + (1 - s)X^y_t. \]
Therefore
\[ \frac{d}{dt}(\|X^x_t - X^y_t\|^2) = 2 \int_0^1 (X^x_t - X^y_t, D_x b(\xi(s))(X^x_t - X^y_t)) \, ds \leq -2\kappa \|X^x_t - X^y_t\|^2 \]
which implies by Gronwall’s lemma
\[ \|X^x_t - X^y_t\| \leq e^{-\kappa t} \|x - y\| \]
for all \( x, y \in \mathbb{R}^d \) and for all \( t \in \mathbb{R}_+ \).

**Remark 4.3.**

a) Notice that in the approach based on the strong gradient bound, we needed non-degeneracy of the covariance matrix \( \mathbf{a} \) in (7), cf. Condition (17). In the coupling setting, we do allow the matrix \( \mathbf{a} \) to be degenerate, but not depending on \( x \), and the condition is only on the drift \( b \).

b) Unlike the time-dependent constant \( D_t \), given via the strong gradient bound (22), the bound (34) yields the correct constant \( D \) at time zero. Remark that the constant of the limiting stationary distribution, which is \( \|a\|/2\kappa \), is invariant under linear rescaling of time, as it should. More precisely, if we multiply the generator by a factor \( \alpha \), \( \|a\| \) is multiplied by this same factor \( \alpha \), and so is the constant \( \kappa \).

c) Note that inequality (33) for all \( x \) and \( u \) is equivalent to
\[ \langle b(x) - b(y), x - y \rangle \leq -\kappa \|x - y\|^2 \]
for all \( x, y \). Better and more general bounds under condition (33) will be obtained in Theorem 5.3, assuming however non-degeracy.

**4.4.2 Examples**

**Example 1: Ornstein-Uhlenbeck process and Brownian motion.**
Coming back to the simple example of the Ornstein-Uhlenbeck process (1), we have coupling rate
\[ \gamma(t) = e^{-\kappa t} \]
and we find (30), i.e., the time evolution of the constant in the Gaussian concentration bound is the same in general as for the special case of a Gaussian
starting measure, as we may take \( C_2 = 1 \). If we have a standard Brownian motion, then \( \gamma(t) = 1 \) and the formula (31) reads (since \( \|a\| = 1 \))

\[
D_t = D_0 + t
\]

which is sharp if the starting measure is the normal law \( \mu_0 = \mathcal{N}(0, \sigma^2) \), which at time \( t \) gives \( \mu_t = \mathcal{N}(0, \sigma^2 + t) \).

**Example 2: Ginzburg-Landau dynamics with boundary reservoirs.**

We consider the process \( (X_t)_{t \geq 0} \) on \( \mathbb{R}^N \) with generator

\[
L = \sum_{i=1}^{N-1}(\partial_i - \partial_{i+1})^2 - (x_{i+1} - x_i)(\partial_{i+1} - \partial_i) + L_1 + L_N
\]

where \( \partial_i \) denotes partial derivative with respect to \( x_i \), and where the extra operators \( L_1 \) and \( L_N \) model the reservoirs and are given by

\[
L_1 = b_1(x_1) \partial_1 + \frac{\sigma_1^2}{2} \partial_1^2, \quad L_N = b_N(x_N) \partial_N + \frac{\sigma_N^2}{2} \partial_N^2.
\]

Here, \( \sigma_1, \sigma_N > 0 \), and the drifts associated to the reservoirs \( \tilde{b}_1, \tilde{b}_N : \mathbb{R} \to \mathbb{R} \) are smooth functions. Such diffusion processes are studied in the literature on hydrodynamic limits, see e.g. [12].

This models a non-equilibrium transport system driven by reservoirs with drift \( \tilde{b}_1, \tilde{b}_N \). In absence of the reservoir driving, the bulk system with generator \( \sum_{i=1}^{N}(\partial_i - \partial_{i+1})^2 - (x_{i+1} - x_i)(\partial_{i+1} - \partial_i) \) has reversible measures which are products of mean zero Gaussian random variables. If \( \tilde{b}_1(x_1) = -\alpha x_1, \tilde{b}_N(x_N) = -\alpha x_N \), with \( \alpha_1, \alpha_N > 0 \), then the system is in equilibrium with reversible Gaussian product measure \( C \exp(-\alpha \sum_{i=1}^{N} x_i^2) \) (where \( C \) is the normalizing constant).

In all other cases, by the coupling to the reservoirs, a non-equilibrium steady state is created.

For the choice \( \tilde{b}_1(x) = -\alpha_1 x, \tilde{b}_N(x) = -\alpha_N x \), with \( \alpha_1 \neq \alpha_N \), this corresponds to a “non-equilibrium” Ornstein-Uhlenbeck process, for which it can be shown that the unique stationary measure \( \mu \) is a product of mean zero Gaussians, with variance given by

\[
\int x_i^2 \, d\mu(x) = \frac{1}{\alpha_1} + \left( \frac{1}{\alpha_N} - \frac{1}{\alpha_1} \right) \frac{i}{N+1}
\]

linearly interpolating between the left and right reservoirs.

The noise in the system is degenerate, but does not depend on \( x \), which means that the coupling condition is satisfied. The covariance matrix \( a \) of \((7)\) is given by

\[
a_{ii} = 2, \quad 2 \leq i \leq N - 1, \quad a_{11} = 1, \quad a_{NN} = 1, \quad a_{i,i+1} = -1, \quad 1 \leq i \leq N - 1
\]
the other entries being equal to 0. The drift \( b(x) \) is given by

\[
b_i(x) = x_{i+1} + x_{i-1} - 2x_i, \quad 1 < i < N \]
\[
b_1(x) = x_2 - x_1 + \tilde{b}_1(x_1),
\]
\[
b_N(x) = x_{N-1} - x_N + \tilde{b}_N(x_N).
\]

If the drifts associated to the reservoirs \( \tilde{b}_1, \tilde{b}_N \) are not linear, then the stationary non-equilibrium state is unknown and not Gaussian. A direct application of Theorem 4.3 then gives the following result. Let \( -\Delta \) denote the discrete laplacian defined via

\[
(\Delta u)_i = u_{i+1} + u_{i-1} - 2u_i \quad \text{for} \quad 2 < i < N - 1,
\]
\[
(\Delta u)_1 = u_2 - u_1 + \tilde{b}_1(u_1),
\]
\[
(\Delta u)_N = u_{N-1} - u_N + \tilde{b}_N(u_N).
\]

**Proposition 4.3.** If the reservoir drifts are such that

\[
\langle u, -\Delta u \rangle + u_1^2 \tilde{b}_1(x_1) + u_N^2 \tilde{b}_N(x_N) \leq -\kappa_N \|u\|^2, \quad \forall u \in \mathbb{R}^N
\]

for some \( \kappa_N > 0 \), then the unique stationary measure of the process with generator \( L \) satisfies GCBS(\( D \)), with \( D = \|a\|/(2\kappa_N) \leq 2/\kappa_N \).

**Example 3: Perturbation of the drift.**

Remark that if (33) is satisfied for the drift \( b \) with constant \( \kappa > 0 \) and \( \tilde{b} \) is such that \( \langle D_x (\tilde{b} - b)(u), u \rangle \leq \epsilon \|u\|^2 \), for some \( 0 < \epsilon < \kappa \), then obviously, (33) is satisfied for the drift \( \tilde{b} \) with constant \( \tilde{\kappa} = \kappa - \epsilon \). For instance, if \( \tilde{b}(x) = -\nabla W(x) + \eta(x) \), where \( W(x) \) is a strictly convex potential, then if \( \|D_x \eta\|_\infty \) is sufficiently small, there is a unique invariant probability measure \( \nu \) which satisfies GCBS. However, \( \eta \) is allowed to be of non-gradient form, which implies that \( \nu \) is not known in explicit form. The same applies to systems where one adds sufficiently weak “boundary” reservoirs as long as the noise associated to these reservoirs does not depend on \( x \).

**Example 4: A degenerate diffusion.**

We consider the diffusion on \( \mathbb{R}^2 \) given by

\[
\begin{align*}
\frac{dq}{dt} &= v dt - \alpha(q) dt \\
\frac{dv}{dt} &= -\gamma v dt + \beta(q) dt + dW_t
\end{align*}
\]

where \( \gamma > 0 \), \( \alpha \) and \( \beta \) are \( C^\infty \) functions with bounded derivatives satisfying

\[
\inf_{q \in \mathbb{R}} \frac{\alpha'(q)}{(1 + \beta'(q))^2} > \frac{1}{4\gamma}.
\]

This diffusion satisfies our hypotheses (H1) and (H2).

The drift is the vector field \( b \) on \( \mathbb{R}^2 \) given by

\[
b(q, v) = \begin{pmatrix} v - \alpha(q) \\ -\gamma v + \beta(q) \end{pmatrix}
\]

30
and we leave to the reader to check that under the hypothesis (35), condition (33) of Theorem 4.3 is satisfied for some $\kappa > 0$. The matrix $\sigma$ is given by

$$\sigma(q, v) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

which is obviously degenerate. The generator $L$ of this diffusion is given by

$$L = \frac{1}{2} \partial_v^2 + (\beta(q) - \gamma v) \partial_v + (v - \alpha(q)) \partial_q$$

and it is left to the reader to check that hypoellipticity holds which follows easily from Hörmander’s condition. Therefore the conclusions of Theorem 4.3 hold.

5 Further applications of the distance Gaussian moment

In this section, we come back to the SDE (5) with drift $b$ and covariance $\sigma$ can possibly depend explicitly on time (in contrast with the previous section) In Section 3, we used Theorem 2.1 to obtain a general result on the propagation of Gaussian concentration bounds. Here we apply it to diffusions “coming down from infinity”, and to diffusions with space-time dependent drift and covariance.

5.1 Diffusions coming down from infinity

As a first example of application, we consider diffusions “coming down from infinity” for which we show that from any starting measure, at positive times $t > 0$, GCBS($D$) holds. We refer to [2] for more on diffusions “coming down from infinity” in the one-dimensional case.

We consider a diffusion process on $\mathbb{R}^d$ which solves the SDE (7), and we further assume that for some $C_1, C_2 > 0$ and for any $t \geq 0$, $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$

$$C_1 \|v\|^2 \leq \langle v, a(t, x)v \rangle \leq C_2 \|v\|^2. \tag{36}$$

We introduce the following condition on the drift.

**CONDITION 5.1.** There exists a real, non-negative, non-decreasing and $\mathcal{C}^1$ function $h$ and a constant $A > 0$ such that for all $x \in \mathbb{R}^d$ and all $t \geq 0$

$$\frac{\langle x, b(t, x) \rangle}{\|x\|} \leq A - h(\|x\|).$$

**THEOREM 5.1.** Under Condition 5.1, if additionally we have the integrability condition

$$\int_0^{+\infty} \frac{du}{h(u)} < +\infty \tag{37}$$
then there exists $t_*>0$, a non-negative bounded function $C$ on $[0,t_*]$, and a constant $\alpha>0$ such that for all $0 \leq t \leq t_*$

$$
\sup_{x \in \mathbb{R}^d} E_x \left( e^{\alpha \|X_t\|^2} \right) \leq C(t).
$$

(38)

We deduce the following result showing that immediate Gaussian concentration arises for diffusions coming down from infinity.

**THEOREM 5.2.** Assume that Condition 5.1 and (37) hold. Let $\mu_0$ be any probability measure on (the Borel field of) $\mathbb{R}^d$. Let $t_*$, $\alpha$, and $C(t)$ be as in Theorem 5.1. Then, for all $t > 0$, the probability measure $\mu_t$ defined by

$$
\mu_t(f) = E_{\mu_0}(f(X_t)), \forall f \in C_b(\mathbb{R}^d)
$$

satisfies GCB($D_t$) where

$$
D_t = \begin{cases}
    \frac{C(t)^2 e}{2\alpha\sqrt{\pi}} & \text{if } 0 < t < t_* \\
    \frac{C(t_*)^2 e}{2\alpha\sqrt{\pi}} & \text{if } t \geq t_*.
\end{cases}
$$

**PROOF.** For $0 < t \leq t_*$, the result follows from Theorems 5.1 and 2.1. For $t > t_*$ the result follows recursively. Namely, assume that for any $0 < t \leq kt_*$, where $k > 0$ is an integer we have

$$
\sup_{x \in \mathbb{R}^d} E_x \left( e^{\alpha \|X_t\|^2} \right) \leq C(t)
$$

where $C(t) = C(t_*)$ for $t \geq t_*$. For any $t \in [kt_*, (k+1)t_*]$, we can apply Theorem 5.1 to $\mu_{t-t_*}$ since $0 < t - t_* \leq kt_*$, and we extend the previous bound to the time interval $[0, (k+1)t_*]$, and hence recursively to any $t > 0$.

$$
\square
$$

**REMARK 5.1.** Theorem 5.2 implies tightness of the family $(\mu_t)_{t>0}$. Therefore the stochastic process $(X_t)$ has invariant probability measures, each of them satisfying GCB($D_{t_*}$). By standard arguments one can show that there is a unique invariant probability measure and it is absolutely continuous with respect to the Lebesgue measure.

**REMARK 5.2.** Let us denote by $\zeta(t)$ the left-hand side of (38), for $t \in [0,t_*]$. The proof of Theorem 5.2 implies that $\zeta(t)$ extends to all $t \geq 0$, and the Markov property implies that it is non-increasing.

Now we prove Theorem 5.1.

**PROOF of Theorem 5.1.** Define

$$
u(t,x) = \varphi(t) e^{\alpha \|x\|^2}$$

32
where $\alpha$ and $\varphi$ will be chosen later on. We have, using Condition 5.1 and (36),

\[
\partial_t u(t, x) + Lu(t, x) = e^{\alpha \|x\|^2} \left( \dot{\varphi}(t) + \varphi(t) \left[ 2 \alpha \text{Tr}(a(t, x)) + 4 \alpha^2 \langle x, a(t, x) x \rangle + 2 \alpha \langle x, b(t, x) \rangle \right] \right) \\
\leq e^{\alpha \|x\|^2} \left( \dot{\varphi}(t) + \varphi(t) \left[ 2 d \alpha C_2 + 4 \alpha^2 C_2 \|x\|^2 + 2 \alpha A \|x\| - 2 \alpha h(\|x\|) \|x\| \right] \right).
\]

Using integration by parts we get

\[
\int_0^z \frac{du}{h(u)} = \frac{z}{h(z)} + \int_0^z \frac{uh'(u)}{h(u)^2} du
\]

and using that $h$ is non-decreasing we obtain

\[
\liminf_{z \to +\infty} \frac{h(z)}{z} \geq \frac{1}{\int_0^{+\infty} \frac{du}{h(u)}} > 0.
\]

Therefore, choosing $1/2 > \alpha > 0$ sufficiently small and $y_*>0$ sufficiently large, we have for $u \geq y_*$

\[
2 \alpha h(u) - 4\alpha^2 u C_2 - 2 \alpha A - \frac{d C_2 \alpha \|x\|}{u} > \alpha h(u).
\]

Hence if $\|x\| > y_*$ we obtain

\[
\partial_t u(t, x) + Lu(t, x) \leq u(t, x) \left( \dot{\varphi}(t) - \alpha \varphi(t) h(\|x\|) \|x\| \right) (39)
\]

Define the function $\omega$ on $[0, 1/y^*]$ by $\omega(0) = 0$ and for $v > 0$

\[
\omega(v) = \int_{v^{-1}}^{+\infty} \frac{ds}{h(s)}.
\]

One check easily is strictly increasing and $C^1$ on $[0, 1/y^*]$. Hence it is a bijection between $[0, 1/y^*]$ and $[0, \alpha t_*/2]$ where $t_*$ is defined via

\[
\int_{y_*}^{+\infty} \frac{du}{h(u)} = \frac{\alpha t_*}{2}.
\]

The inverse function $\omega^{-1}$ is continuous on $[0, \alpha t_*/2]$, and $C^1$ on $[0, \alpha t_*/2]$. Now define $y(s)$ for $s \in [0, t_*)$ by

\[
y(s) = \frac{1}{\omega^{-1}(\frac{s}{t_*})}.
\]

One checks that this function is $C^1$ decreasing on $[0, t_*]$. Moreover $\lim_{s \downarrow 0} y(s) = +\infty$. Now define the function $\varphi(s)$ for $s \in [0, t_*]$ by

\[
\varphi(s) = e^{-y(s)^2/2}.
\]
Note that $\varphi$ is continuous on $[0, t_\ast]$ and $C^1$ on $]0, t_\ast]$, and $\varphi(0) = 0$. Observe that for $s \in [0, t_\ast]$ we have
\[
\frac{\varphi'}{\varphi} = -\frac{\dot{y}(s)}{y(s)} = \alpha \frac{y(s) h(y(s))}{2}.
\]
Note that for $0 \leq t \leq t_\ast$ we have $y(s) \geq u(s)$. Given $x \in \mathbb{R}^d$, let $B > 2 \max \{y_\ast, \|x\|\}$. Using Itô’s formula, with $T_B$ the hitting time of the boundary of the ball centered at $x$ with radius $B$, we get
\[
E_x(u(t \wedge T_B, X_{t \wedge T_B})) = E_x \left( \int_0^{t \wedge T_B} (\partial_t u + Lu)(s, X_s) ds \right).
\]
For $0 < s \leq t_\ast$, if $\|X_s\| \leq y(s) \geq y_\ast$ we have, using (39) and the monotonicity of $h$
\[
(\partial_t u + Lu)(s, X_s) \leq e^{\alpha \|X_s\|^2} \left( \dot{\varphi}(s) - \alpha \varphi(s) \frac{h(y(s)) y(s)}{2} \right) = 0.
\]
For $0 < s \leq t_\ast$, if $\|X_s\| < y(s)$ we have
\[
(\partial_t u + Lu)(s, X_s) \leq C e^{\alpha y(s)^2} (\dot{\varphi}(s) + \varphi(s) (1 + y(s)^2))
\]
for some (computable) constant $C > 0$ independent of $s$. Therefore if $0 \leq t \leq t_\ast$ we obtain
\[
E_x \left( \int_0^{t \wedge T_B} (\partial_t u + Lu)(s, X_s) ds \right)
\]
\[
\leq C E_x \left( \int_0^{t \wedge T_B} e^{\alpha y(s)^2} (\varphi(s) + \varphi(s) (1 + y(s)^2)) ds \right)
\]
\[
\leq C \int_0^t e^{\alpha y(s)^2} (\varphi(s) + \varphi(s) (1 + y(s)^2)) ds
\]
\[
= -C \int_0^t e^{\alpha y(s)^2} \dot{y}(s) y(s) e^{-\frac{y(s)^2}{2}} ds + C \int_0^t e^{\alpha y(s)^2} e^{-\frac{y(s)^2}{2}} (1 + y(s)^2) ds
\]
and since $\alpha < 1/2$ we obtain
\[
E_x \left( \int_0^{t \wedge T_B} (\partial_t u + Lu)(s, X_s) ds \right) \leq C \int_{y(0)}^{+\infty} e^{\alpha y^2} y e^{-\frac{y^2}{2}} dy + \frac{2 C t}{1 - 2\alpha} \int_0^t ds
\]
\[
= \frac{C e^{-(1-2\alpha)\frac{y(0)^2}{2}}}{1 - 2\alpha} + \frac{2 C t}{1 - 2\alpha}.
\]
We now observe that since $u \geq 0$, for any $0 \leq t \leq t_\ast$ we have
\[
E_x(u(t, X_t) 1_{\{T_B > t\}}) \leq E_x(u(t \wedge T_B, X_{t \wedge T_B})) \leq \frac{C e^{-(1-2\alpha)\frac{y(0)^2}{2}}}{1 - 2\alpha} + \frac{2 C t}{1 - 2\alpha}.
\]
Therefore by the monotone convergence theorem (letting \( B \) tend to infinity)

\[
E_x(u(t, X_t)) \leq \frac{C e^{-(1-2\alpha)\frac{y(t)^2}{2}}}{1-2\alpha} + \frac{2C t}{1-2\alpha}.
\]

The result follows with

\[C(t) = \frac{C}{1-2\alpha} \left( e^{\alpha y(t)^2} + 2 te^{\frac{y(t)^2}{2}} \right).\]

5.2 Diffusion processes with space-time dependent drift and covariance

In this section, we consider diffusions which do not come down from infinity as strongly as in Theorem 5.1 (for example the Ornstein-Uhlenbeck process). We use a confining condition on the drift, which allows to give better estimates on the constants \( D_t \) and \( D_\infty \) than in Section 3.

**Theorem 5.3.** Assume that \( \alpha > 0 \), \( \beta > 0 \) and \( \theta > 0 \) such that, for all \( x \in \mathbb{R}^d \) and \( t \geq 0 \)

\[
\langle x, b(x, t) \rangle \leq \alpha \|x\| - \beta \|x\|^2
\]

and

\[
\sigma(x, t) \sigma(x, t)^\top \leq \theta I_d
\]

where the second inequality we use the order induced by positive definite matrices. Then, for every initial probability measure \( \mu_0 \) on \( \mathbb{R}^d \) satisfying GCBS\((D_0)\), the evolved probability measure \( \mu_t \) satisfies GCB\((D_t)\) for all \( t \geq 0 \), where \( D_t \) is given by formula (4), with

\[
a = a_0 = \frac{\beta}{2\theta} \wedge \frac{1}{16D_0}
\]

\[
b = b_t = b_0 \exp\left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right)
+ 2 e^{\frac{4a_0}{\theta} \left( \theta d + \frac{2\alpha^2}{\beta} \right)} \left( 1 - \exp\left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \right)
\]

and

\[
b_0 = 3 e^{\frac{\mu_0(d)^2}{8D}}
\]

where \( \mu_0(d) = \int \|x\| \, d\mu_0(x) \).

The same conclusions as in Remark 5.1 hold with GCB\((D_\infty)\) where

\[
D_\infty = \frac{b_\infty^2 e}{2a\sqrt{\pi}}
\]
The conditions of the previous theorem generalize the coupling setting of Theorem 4.3, i.e., we impose a more general confining condition on the drift $b(x,t)$ and allow the covariance matrix $\sigma(x,t)$ to depend on time and location. However we get a worse estimate on $D_t$.

**Proof.** Let $a_0 = \frac{\beta}{2\theta} \wedge \frac{1}{16D_0}$ and define $u(x) = e^{a_0\|x\|^2}$. Using the assumptions we get

$$Lu(x) \leq (2a_0^2 \theta \|x\|^2 + a_0 \theta d + 2a_0 \alpha \|x\| - 2a_0 \beta \|x\|^2)u(x)$$

$$\leq a_0 (\theta d + 2\alpha \|x\| - \beta \|x\|^2)u(x)$$

$$\leq a_0 \left( \theta d + \frac{2\alpha^2}{\beta} - \frac{\beta}{2} \|x\|^2 \right) u(x).$$

For any $A > 0$, let $T_A = \inf \{ t \geq 0 : \|X_t\| \geq A \}$. Using Dynkin’s formula and Theorem 2.1, we get

$$E_{\mu_0} \left( e^{a_0\|X_t\|^2} \right) \leq b_0 + a_0 E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} \left( \theta d + \frac{2\alpha^2}{\beta} - \frac{\beta}{2} \|X_s\|^2 \right) d\sigma_s \right)$$

(41)

where, via (2)

$$b_0 = \int e^{\mu_0\|x\|^2} d\mu_0(x) \leq 3 e^{\frac{\mu_0(d)^2}{\kappa D}}$$

where $\mu_0(d) = \int \|x\| d\mu_0(x)$. We now estimate the expectation on the right-hand side of (41). Define, for $s > 0$, the event

$$E_s = \left\{ \|X_s\|^2 > \frac{4}{\beta} \theta d + \frac{8\alpha^2}{\beta} \right\}.$$

We have

$$E_{\mu_0} \left( e^{a_0\|X_t\|^2} \right) \leq b_0 + 2a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} 1_{E_s} d\sigma_s \right)$$

$$- a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} d\sigma_s \right)$$

$$\leq b_0 + 2a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) e^{\frac{4a_0}{\beta} \left( \theta d + \frac{8\alpha^2}{\beta} \right) t} \left( \theta d + \frac{2\alpha^2}{\beta} \right) E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} d\sigma_s \right).$$

We have

$$E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} d\sigma_s \right) = E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} d\sigma_s \right)$$

$$\leq E_{\mu_0} \left( \int_0^{T_A} e^{a_0\|X_s\|^2} d\sigma_s \right).$$

36
Therefore
\[
E_{\mu_0} \left( e^{a_0 \|X_t \wedge T_A\|^2} \right) \leq b_0 + 2a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) \frac{4a_0}{e} \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \\
- a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) E_{\mu_0} \left( \int_0^t e^{a_0 \|X_s \wedge T_A\|^2} \, ds \right) .
\]

Using Grönwall’s lemma, and the fact that \( xe^{-x} \leq 1 - e^{-x} \) for \( x \geq 0 \), we obtain
\[
E_{\mu_0} \left( e^{a_0 \|X_t \wedge T_A\|^2} \right) \leq b_0 \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \\
+ 2 e^{a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right)} \left( 1 - \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \right) .
\]

This implies
\[
E_{\mu_0} \left( 1_{\{t<T_A\}} e^{a_0 \|X_t\|^2} \right) \leq b_0 \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \\
+ 2 e^{a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right)} \left( 1 - \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \right) .
\]

By the monotone convergence theorem, letting \( A \uparrow \infty \), we get
\[
E_{\mu_0} \left( e^{a_0 \|X_t\|^2} \right) \leq b_0 \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \\
+ 2 e^{a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right)} \left( 1 - \exp \left( -a_0 \left( \theta d + \frac{2\alpha^2}{\beta} \right) t \right) \right) .
\]

By Theorem 2.1, we deduce that \( \mu_t \) satisfies GCBS(\( D_t \)) with the announced constant \( D_t \). \( \square \)

**Remark 5.3.**

In the case of diffusions coming down from infinity, we saw in Theorem 5.1 that GCB develops out of the time evolution of any initial distribution. In Theorem 5.3 we required that the initial distribution satisfies GCB. In the case of the one-dimensional Ornstein-Uhlenbeck process, if one starts for example with the initial probability distribution
\[
d\mu_0(x) = \frac{\sqrt{2}}{\pi} \frac{dx}{1 + x^4}
\]
it is easy to verify by an explicit computation that for any \( t \geq 0 \) and any \( a > 0 \)
\[
\int_{-\infty}^{+\infty} d\mu_0(x) \, E_x \left( e^{a X_t^2} \right) = \infty.
\]
5.3 Example: the noisy Lorenz system

As an application of Theorem 5.3, we consider the famous Lorenz system

\[
\begin{align*}
\frac{dx_1}{dt} &= \sigma(x_2 - x_1) \\
\frac{dx_2}{dt} &= rx_1 - x_2 - x_1x_3 \\
\frac{dx_3}{dt} &= x_1x_2 - bx_3
\end{align*}
\]

which, for a certain range of (strictly positive) parameters, has a strange attractor [13, Chapter 14].

Adding a noise which satisfies the condition of Theorem 5.3, this leads to a unique invariant probability measure whose properties are largely unknown. However, this measure satisfies GCBS. This can be proved observing that the stochastic process \(X_t = (X_t^{(1)}, X_t^{(2)}, X_t^{(3)})\) satisfies (40) using the squared norm \(\| (x_1, x_2, x_3) \|^2 = rx_1^2 + \sigma x_2^2 + \sigma x_3^2\) with

\[
\beta = \inf \frac{rx_1^2 + x_2^2 + bx_3^2}{rx_1^2 + \sigma x_2^2 + \sigma x_3^2} \geq \min \{ 1, \sigma^{-1}, b\sigma^{-1} \}
\]

where the infimum is taken over \(x_1, x_2, x_3\) with \((x_1, x_2, x_3) \neq (0, 0, 0)\). Indeed, we have

\[
rx_1^2 + x_2^2 + bx_3^2 \geq \min \{ 1, \sigma^{-1}, b\sigma^{-1} \} \times (rx_1^2 + \sigma x_2^2 + \sigma x_3^2)
\]

whence \(\beta \geq \min \{ 1, \sigma^{-1}, b\sigma^{-1} \}\). Moreover

\[
(rx_1^2 + x_2^2 + bx_3^2)/(rx_1^2 + \sigma x_2^2 + \sigma x_3^2) = \min \{ 1, \sigma^{-1}, b\sigma^{-1} \}
\]

on some coordinate axis outside the origin. Note that the expression for \(\beta\) also follows from the Rayleigh-Riesz principle.

6 Non-Markovian diffusions: Martingale moment approach

In this section we consider the simplest context beyond the Markov case, where we can no longer rely on methods based on generators. We will again exploit Theorem 2.1.

We consider the stochastic differential equation on \(\mathbb{R}\) given by

\[
\begin{align*}
\text{d}X_t = -\kappa X_t \text{d}t + \sigma_t \text{d}W_t
\end{align*}
\]

where we assume that the process \((\sigma_t)_{t \geq 0}\) is uniformly bounded and predictable. An example of this setting is

\[
\begin{align*}
\begin{cases}
\text{d}Y_t &= -\theta Y_t + \text{d}W_t \\
\text{d}X_t &= -\kappa X_t + \sigma(Y_t) \text{d}W_t
\end{cases}
\end{align*}
\]

38
Then the couple \((X_t, Y_t)_{t \geq 0}\) is a Markov process, but \((X_t)_{t \geq 0}\) is not a Markov process, and satisfies a SDE of the form (42).

Because the process \((X_t)_{t \geq 0}\) is no longer a Markov process (unless \(\sigma_t\) depends only on \(X_t\)) we can no longer use techniques based on the generator as we did before for processes of Ornstein-Uhlenbeck type. The main point is that as a consequence, \(X_t^x\) equals a deterministic process of bounded variation plus a stochastic integral w.r.t. \(dW_t\). As a consequence, the Gaussian concentration bound can be obtained from estimating the stochastic integral, which can be done with the help of Burkholder’s inequalities.

The assumption (42) allows us to write the solution in the form

\[
X_t = X_0 e^{-\kappa t} + \int_0^t e^{-\kappa(t-s)} \sigma_s \, dW_s. \tag{43}
\]

We have the following result.

**Theorem 6.1.** Assume that there exists \(M > 0\) such that

\[
\sup_{t \geq 0} \|\sigma_t\|_{L^\infty} \leq M.
\]

Assume \(X_0\) is distributed according to a probability measure \(\mu_0\) satisfying GCBS\((D_0)\). Then we have that for all \(t > 0\) there exists \(D_t > 0\) such that \(X_t\) satisfies GCBS\((D_t)\). Moreover, if \(\kappa > 0\) then all weak limit points of \((X_t)_{t \geq 0}\) satisfy GCBS\((D_\infty)\) for some \(D_\infty > 0\).

**Proof.** We use Theorem 2.1, and will prove that there exist \(a > 0, b > 0\) such that

\[
E_{\mu_0} \left( e^{aX_t^2} \right) \leq b.
\]

Then we can conclude via Theorem 2.1, that the distribution of \(X_t\) satisfies GCBS\((C)\) with \(C \leq \frac{k^2a^2}{2\kappa^2} \). We start from (43) from which we derive the inequality

\[
X_t^2 \leq 2X_0^2 e^{-2\kappa t} + 2 \left( \int_0^t e^{-\kappa(t-s)} \sigma_s \, dW_s \right)^2. \tag{44}
\]

Let \(u > 0\). We start by estimating

\[
E_{\mu_0} \left[ \exp \left( u \left( \int_0^t e^{-\kappa(t-s)} \sigma_s \, dW_s \right)^2 \right) \right] = \sum_{n=0}^{+\infty} \frac{u^n}{n!} E_{\mu_0} \left[ \left( \int_0^t e^{-\kappa(t-s)} \sigma_s \, dW_s \right)^{2n} \right].
\]

Next use Burkholder’s inequality [8] which states that for a martingale \((Z_t)_{t \geq 0}\) w.r.t. Brownian filtration, with quadratic variation \([Z, Z]_t\), we have the estimate

\[
E_{\mu_0} (Z_t^{2n}) \leq A(2n)^n E_{\mu_0} ([Z, Z]_t^n)
\]

39
with $A$ an absolute constant. As a consequence, we get

$$
E_{\mu_0} \left[ \left( \int_0^t e^{-\kappa(t-s)} \sigma_s dW_s \right)^{2n} \right] = e^{-2nt} E_{\mu_0} \left[ \left( \int_0^t e^{\kappa s} \sigma_s dW_s \right)^{2n} \right] \\
\leq e^{-2nt} A(2n)^n E_{\mu_0} \left[ \left( \int_0^t e^{2\kappa s} \sigma_s^2 ds \right)^n \right] \\
\leq e^{-2nt} AM^{2n}(2n)^n E_{\mu_0} \left[ \left( \int_0^t e^{2\kappa s} ds \right)^n \right] \\
\leq AM^{2n}(2n)^n \left( \frac{1 - e^{-2nt}}{2\kappa} \right)^n.
$$

As a consequence we obtain

$$
E_{\mu_0} \left[ \exp \left( u \left( \int_0^t e^{-\kappa(t-s)} \sigma_s dW_s \right)^2 \right) \right] \leq A \sum_{n=0}^{+\infty} u^n M^{2n}(2n)^n \left( \frac{1 - e^{-2nt}}{2\kappa} \right)^n.
$$

The right-hand side of this inequality is a convergent series provided

$$
u < \left( 2e M^2 \left( \frac{1 - e^{-2nt}}{2\kappa} \right) \right)^{-1}.
$$

Then, by (44) and the Cauchy-Schwarz inequality, we get

$$
E_{\mu_0} \left( e^{aX_0^2} \right) \leq \left( E_{\mu_0} \left[ e^{4aX_0^2_0} e^{-2nt} \right] \right)^{1/4} \left( E_{\mu_0} \left[ e^{4a \left( \int_0^t e^{-\kappa(t-s)} \sigma_s dW_s \right)^2} \right] \right)^{1/4}.
$$

(45)

Because by assumption the distribution of $X_0$ satisfies GCBS($C$), we have that the first factor in the r.h.s. in (45) is finite as soon as $4a e^{-2nt} < a_0$ where $a_0$ is such that $E_{\mu_0} \left( e^{a_0 X_0^2} \right) < +\infty$. The second factor is finite as soon as

$$
a < \left( 8e M^2 \left( \frac{1 - e^{-2nt}}{2\kappa} \right) \right)^{-1}.
$$

Therefore, $E_{\mu_0} \left( e^{aX_t^2} \right)$ is finite for

$$
a < \left( 8e M^2 \left( \frac{1 - e^{-2nt}}{2\kappa} \right) \right)^{-1} \wedge a_0 e^{2nt}
$$

which, combined with Theorem 2.1, concludes the proof of the theorem. \qed
A Proof of Theorem 2.1

Throughout this proof, we set \( \mu(d; x_*) := \int d(x, x_*) \, d\mu(x) \).

**Statement 1.** Choose \( x_* \in \Omega \) arbitrarily. Since \( x \mapsto d(x, x_*) \) is 1-Lipschitz, \( \text{GCB}(D) \) (see Definition 2.2) implies that for all \( r \geq 0 \) we have

\[
\mu \{ x \in \Omega : d(x, x_*) \geq \mu(d; x_*) + r \} \leq e^{-\frac{r^2}{4D}} \tag{46}
\]

where

\[
\mu(d; x_*) = \int d(x, x_*) \, d\mu(x).
\]

Indeed, \( \text{GCB}(D) \) gives for any \( \lambda > 0 \)

\[
\mu \left( e^{\lambda(d(\cdot, x_*)) - \mu(d; x_*)} \right) \leq e^{D \lambda^2}
\]

whence by Markov’s inequality we have for any \( r \geq 0 \)

\[
\mu \{ x \in \Omega : d(x, x_*) \geq \mu(d; x_*) + r \} = \mu \{ x \in \Omega : \lambda(d(x, x_*) - \mu(d; x_*)) \geq \lambda r \} \leq \mu \left( e^{\lambda(d(\cdot, x_*)) - \mu(d;x_*)} \right) e^{-\lambda r} \leq e^{D \lambda^2 - \lambda r}
\]

which gives (46) by minimizing over \( \lambda > 0 \). Now take \( a > 0 \) to be chosen later on. We have

\[
\int e^{ad(x, x_*)^2} \, d\mu(x)
\]

\[
= \int e^{ad(x, x_*)^2} \mathbf{1}_{\{d(x, x_*) \leq \mu(d; x_*)\}} \, d\mu(x) + \int e^{ad(x, x_*)^2} \mathbf{1}_{\{d(x, x_*) > \mu(d; x_*)\}} \, d\mu(x)
\]

\[
\leq e^{a \mu(d; x_*)^2} + e^{2a \mu(d; x_*)^2} \int e^{2a(d(x, x_*) - \mu(d; x_*)^2)} \mathbf{1}_{\{d(x, x_*) > \mu(d; x_*)\}} \, d\mu(x).
\]

Now we use the fact that

\[
\int e^{2a(d(x, x_*) - \mu(d; x_*))^2} \mathbf{1}_{\{d(x, x_*) > \mu(d; x_*)\}} \, d\mu(x)
\]

\[
= \mu \{ x \in \Omega : d(x, x_*) > \mu(d; x_*) \} + \int_{1}^{+\infty} \mu \left\{ x : e^{2a(d(\cdot, x_*) - \mu(d; x_*))^2} > u \right\} \, du
\]

\[
\leq 1 + \int_{1}^{+\infty} \mu \left\{ x : d(x, x_*) - \mu(d; x_*) > \sqrt{\log u/(2a)} \right\} \, du.
\]

The result follows using (46) with \( r = \sqrt{\log u/(2a)} \), and choosing \( a = 1/(16D) \) (which makes the last integral bounded above by 1).

**Statement 2.** Since for all \( x \) and for all \( a > 0 \)

\[
d(x, x_*) \leq \frac{1}{\sqrt{a}} e^{ad(x, x_*)^2}
\]
it follows that \( x \mapsto d(x, x^*) \) is \( \mu \)-integrable. We also have that \( e^f \) is \( \mu \)-integrable for any Lipschitz function. Now, using Jensen’s inequality and then the triangle inequality, we obtain
\[
\int e^{f - \mu(f)} \, d\mu \leq \int \int e^{f(x) - f(y)} \, d\mu(x) \, d\mu(y) \quad (47)
\]
\[
\leq \int \int e^{\text{lip}(f) d(x,y)} \, d\mu(x) \, d\mu(y)
\]
\[
\leq \left( \int e^{\text{lip}(f) d(x,x^*)} \, d\mu(x) \right)^2.
\]
Combining the elementary inequality
\[
\text{lip}(f) d(x, x^*) \leq \frac{\text{lip}(f)^2}{4a} + ad(x, x^*)^2
\]
with (3), we obtain
\[
\int e^{\text{lip}(f) d(x,x^*)} \, d\mu(x) \leq b e^{\frac{1}{4a} \text{lip}(f)^2}.
\]
This implies
\[
\int \int e^{f(x) - f(y)} \, d\mu(x) \, d\mu(y) \leq b^2 e^{\frac{1}{4a} \text{lip}(f)^2}.
\quad (48)
\]
We now show how the pre-factor of the exponential can be changed to 1. We first establish the following lemma.

**Lemma A.1.** Let \( Z \) be a random variable with all odd moments vanishing and such that there exist \( C_1 \geq 1 \) and \( C_2 > 0 \) such that for all \( \lambda \in \mathbb{R} \)
\[
\mathbb{E}(e^{\lambda Z}) \leq C_1 e^{C_2 \lambda^2}.
\quad (49)
\]
Then for all \( \lambda \in \mathbb{R} \) we have
\[
\mathbb{E}(e^{\lambda Z}) \leq e^{\frac{C_1 C_2 e}{\sqrt{\pi}} \lambda^2}.
\]

**Proof.** Let \( q \in \mathbb{N} \). Then for any \( \theta > 0 \)
\[
\mathbb{E}(Z^{2q}) = \mathbb{E}(Z^{2q} e^{-\theta Z} e^{\theta Z})
\]
\[
= \mathbb{E}((Z^{2q} e^{\theta Z}) e^{-\theta Z} 1_{\{Z < 0\}}) + \mathbb{E}((Z^{2q} e^{-\theta Z}) e^{\theta Z} 1_{\{Z \geq 0\}})
\]
\[
\leq 2C_1 (2q)^2 \theta^{-2q} e^{-2q} e^{C_2 \theta^2}
\]
where the inequality follows by maximizing \( x^{2q} e^{-\theta x} \) over \( x < 0 \) in the first term and over \( x \geq 0 \) in the second one, and then (49). Now we can minimize the bound over \( \theta > 0 \) to get
\[
\mathbb{E}(Z^{2q}) \leq 2C_1 4^q q^q e^{-q} C_2^q.
\quad (50)
\]
Using the bounds
\[ \sqrt{2\pi n^{n+\frac{1}{2}} e^{-n}} \leq n! \leq e n^{n+\frac{1}{2}} e^{-n}, \quad n \geq 1 \]
we get
\[ \frac{2C_1 q^q e^{-q} C_2^q}{(2q)!} = 2C_1 q^q e^{-q} C_2^q \frac{q!}{(2q)!} \times \frac{1}{q!} \leq \frac{\left( \frac{C_1 e}{\sqrt{\pi}} \right)^q C_2^q}{q!}, \quad q \in \mathbb{N}. \]
Therefore using the assumption that $Z$ has all its odd moments equal to 0, (50), and the previous bound we have
\[ \mathbb{E} \left( e^{\lambda Z} \right) = 1 + \sum_{q=1}^{+\infty} \frac{\mathbb{E}(Z^{2q})\lambda^{2q}}{(2q)!} \leq 1 + \sum_{q=1}^{+\infty} \frac{\left( \frac{C_1 C_2 e \lambda^2}{\sqrt{\pi}} \right)^q}{q!} = e^{\frac{C_1 C_2 e \lambda^2}{\sqrt{\pi}}}. \]
The proof is finished.

We now apply Lemma A.1 to the random variable $Z = f(X) - f(Y)$, where $(X,Y)$ is distributed according to the product probability measure $d\mu(x) d\mu(y)$. It is easy to verify that all odd moments vanish ($Z$ is antisymmetric with respect to the exchange of $X$ and $Y$) and the bound on the exponential moments follow by replacing $f$ by $\lambda f$ in (48). We use the constants $C_1 = b^2$ and $C_2 = \frac{\text{lip}(f)^2}{2a}$. Finally, the second statement of Theorem 2.1 follows from (47).

## B A general approximation lemma

In this appendix, $(\Omega, \| \cdot \|)$ is a separable Banach space equipped with a sigma-algebra of Borel sets. We denote by $\text{Lip}(\Omega, \mathbb{R})$ the space of real-valued Lipschitz functions on $(\Omega, \| \cdot \|)$, by $\text{Lip}_b(\Omega, \mathbb{R})$ the space of real-valued Lipschitz functions with bounded support, and by $\text{Lip}_b(\Omega, \mathbb{R})$ the space of real-valued bounded Lipschitz functions. We denote by $\mathcal{C}^\infty(\Omega, \mathbb{R})$ the space of real-valued infinitely differentiable functions, and by $\mathcal{C}^{\infty}_s(\Omega, \mathbb{R})$ the space of real-valued infinitely differentiable functions with bounded support.

Let $\mathcal{C}$ be a class of real-valued functions on $\Omega$. We say that $\mu$ satisfies $\text{GCB}(\mathcal{C}; D)$ if there exists $D > 0$ such that
\[ \log \mu \left( e^{f - \mu(f)} \right) \leq D \text{lip}(f)^2 \]
for all $f \in \mathcal{C}$.

**Lemma B.1.** Let $\mu$ be a probability measure on $\Omega$. Then

1. If $\mu$ satisfies $\text{GCB}(\mathcal{C}^{\infty}_s(\Omega, \mathbb{R}); D)$, then it satisfies $\text{GCB}(\text{Lip}_b(\Omega, \mathbb{R}); D)$.  

43
2. If \( \mu \) satisfies GCB(\( \text{Lip}_s(\Omega, \mathbb{R}); D) \), then it satisfies GCB(\( \text{Lip}(\Omega, \mathbb{R}); D) \).

**Proof.** Let \( \nu \) be a \( \mathcal{C}^\infty \) (in the sense of distributions) probability measure on \( \Omega \) with bounded support. For every \( \lambda > 0 \) we define the rescaled measure \( \nu_\lambda \) by

\[

\nu_\lambda(f) := \nu(f_\lambda)

\]

for any \( f \) continuous with bounded support, where \( f_\lambda(x) := f(\lambda x) \). For \( f \in \text{Lip}_s(\Omega, \mathbb{R}) \), we have \( \nu_\lambda * f \in \mathcal{C}^\infty(\Omega, \mathbb{R}) \) and \( \text{lip}(\nu_\lambda * f) \leq \text{lip}(f) \). Since \( \mu \) is assumed to satisfy GCB(\( \mathcal{C}^\infty_s(\Omega, \mathbb{R}); D) \), it follows that

\[

\mu \left( e^{\nu_\lambda * f - \mu(\nu_\lambda * f)} \right) \leq e^{D \text{lip}(f)^2}.

\]

The first statement then follows by dominated convergence.

For the second statement, as an intermediate step, we prove that if \( \mu \) satisfies GCB(\( \text{Lip}_s(\Omega, \mathbb{R}); D) \) then it satisfies GCB(\( \text{Lip}_b(\Omega, \mathbb{R}); D) \). Let \( \psi : \mathbb{R}_+ \to \mathbb{R}_+ \) be defined by

\[

\psi(u) = \begin{cases} 
1 & \text{if } u \leq 1 \\
2 - u & \text{if } 1 \leq u \leq 2 \\
0 & \text{if } u \geq 2.
\end{cases}

\]

For any \( A > 0 \) define \( \psi_A : \Omega \to \mathbb{R}_+ \) by

\[

\psi_A(x) = \psi \left( \frac{\|x\|}{A} \right).

\]

We have \( \psi_A \in \text{Lip}_s(\Omega, \mathbb{R}) \) and \( \text{lip}(\psi_A) \leq 1/A \). Without loss of generality, we can take \( f \in \text{Lip}_b(\Omega, \mathbb{R}) \) such that \( f(0) = 0 \). Then define the function \( F_A \) by

\[

F_A(x) = f(x)\psi_A(x).

\]

We show that \( F_A \in \text{Lip}_s(\Omega, \mathbb{R}) \). We have

\[

F_A(x) - F_A(y) = f(x)(\psi_A(x) - \psi_A(y)) + \psi_A(y)(f(x) - f(y)).

\]

Since \( \|\psi_A\|_\infty \leq 1 \) we get

\[

\text{lip}(F_A) \leq \frac{\|f\|_\infty}{A} + \text{lip}(f).

\]

Since \( \mu \) is assumed to satisfy GCB(\( \text{Lip}_s(\Omega, \mathbb{R}); D) \), we have

\[

\mu \left( e^{F_A - \mu(F_A)} \right) \leq \exp \left( D \left( \frac{\|f\|_\infty}{A} + \text{lip}(f) \right)^2 \right).

\]

Using the Dominated Convergence Theorem, we take the limit \( A \to +\infty \) and get

\[

\mu \left( e^{f - \mu(f)} \right) \leq e^{D \text{lip}(f)^2}.

\]
Finally, let us prove that if $\mu$ satisfies GCB($\text{Lip}_b(\Omega, \mathbb{R}); D$) then it satisfies GCB($\text{Lip}(\Omega, \mathbb{R}); D$). Define for $M > 0$

$$f_M(x) = (f(x) \land M) \lor (-M).$$

By observing that $\text{lip}(f_M) \leq \text{lip}(f)$ and since $\mu$ satisfies GCB($\text{Lip}_b(\Omega, \mathbb{R}); D$) by assumption, we have

$$\mu \left( e^{f_M - \mu(f_M)} \right) \leq e^{D \text{lip}(f)^2}. \quad (51)$$

We are going to take the limit $M \to +\infty$ and prove that the left-hand side converges to $\mu (\exp(f - \mu(f)))$. We first prove that $\sup_{M > 0} |\mu(f_M)| < +\infty$. We start by proving that $\inf_{M > 0} \mu(f_M) > -\infty$. Take a ball $B$ such that $\mu(B) > 0$. Denote by $x_B$ its center and by $r_B$ its radius. Assume that for all $x \in B$ we have

$$\mu(B) e^{f_M(x) - \mu(f_M)} > e^{D \text{lip}(f)^2}.$$

Integrating over $B$ with respect to $\mu$ we get

$$\mu(B) \mu \left( e^{f_M - \mu(f_M)} 1_B \right) > e^{D \text{lip}(f)^2} \mu(B)$$

which contradicts (51). Hence there exists $B' \subset B$ such that $\mu(B') > 0$ and such that for any $x \in B'$ we have

$$\mu(B) e^{f_M(x) - \mu(f_M)} \leq e^{D \text{lip}(f)^2}. \quad (52)$$

Hence, picking an arbitrary $x \in B'$ and using that $\text{lip}(f_M) \leq \text{lip}(f)$, we get

$$f_M(x_B) \leq \mu(f_M) + D \text{lip}(f)^2 - \log \mu(B) + \text{lip}(f) r_B.$$

Since $f_M(0) = 0$, we have $f_M(x_B) \geq -\text{lip}(f) ||x_B||$, which implies $\inf_{M > 0} \mu(f_M) > -\infty$.

A similar argument applies to $-f$, therefore

$$A_f := \sup_{M > 0} |\mu(f_M)| < +\infty.$$

We now prove that $e^f$ is integrable with respect to $\mu$. We have

$$\mu \left( e^{f_M} \right) = \mu \left( 1_{\{f \geq 0\}} e^{f_M} \right) + \mu \left( 1_{\{f < 0\}} e^{f_M} \right). \quad (52)$$

If $x \in \Omega$ is such that $f(x) \geq 0$, then $f_M(x) \uparrow f(x)$ as $M \uparrow +\infty$, then

$$\mu \left( 1_{\{f \geq 0\}} e^{f_M} \right) \leq \mu \left( e^{f_M} \right) \leq e^{D \text{lip}(f)^2 + A_f}.$$

By the Monotone Convergence Theorem we thus get

$$\mu \left( 1_{\{f \geq 0\}} e^f \right) = \lim_{M \to +\infty} \mu \left( 1_{\{f \geq 0\}} e^{f_M} \right) \leq e^{D \text{lip}(f)^2 + A_f}. \quad (53)$$
Now we deal with the second term in the right-hand side of (52). Since the function $\mathbb{1}_{\{f < 0\}} e^{f_M}$ is nonnegative and bounded above by 1 and converges pointwise to $\mathbb{1}_{\{f < 0\}} e^f$ as $M$ tends to $+\infty$, we apply the Dominated Convergence Theorem to get that

$$\lim_{M \to +\infty} \mu \left( \mathbb{1}_{\{f < 0\}} e^{f_M} \right) = \mu \left( \mathbb{1}_{\{f < 0\}} e^f \right).$$

Therefore, using this inequality, (53) and (52) we conclude that

$$\lim_{M \to +\infty} \mu (e^{f_M}) = \mu (e^f) < +\infty. \tag{54}$$

By a similar argument one shows that $\mu (e^{-f}) < +\infty$.

We now prove that $\mu (f_M)$ converges to $\mu (f)$ as $M$ tends to $+\infty$. We observe that $|f_M| \leq e^f + e^{-f}$. Hence by the Dominated Convergence Theorem we conclude that

$$\lim_{M \to +\infty} \mu (f_M) = \mu (f). \tag{55}$$

Using (55) and (54), we can take the limit $M \to +\infty$ in inequality (51) and obtain

$$\mu \left( e^{f - \mu (f)} \right) \leq e^{D \text{lip}(f)^2}.$$

The lemma is proved. \[\Box\]

Acknowledgment

F. Redig thanks Ecole Polytechnique and CNRS for financial support and hospitality. We thank R. Kraaij for useful discussions in an early stage of this work.

References

[1] D. Bakry, I. Gentil, M. Ledoux. Analysis and geometry of Markov diffusion operators. Grundlehren der Mathematischen Wissenschaften, 348. Springer, 2014.

[2] V. Bansaye, P. Collet, S. Martínez, S. Méléard, J. San Martín. Diffusions from infinity. Trans. AMS 372, 5781–5823 (2019).

[3] M. Bertoldi, L. Lorenzi. Analytical Methods for Markov Semigroups. Chapman & Hall/CRC, 2007.

[4] S. Boucheron, G. Lugosi, P. Massart. Concentration inequalities. A nonasymptotic theory of independence. Oxford University Press, 2013.
[5] J.-R. Chazottes, P. Collet, C. Külske, F. Redig. Concentration inequalities for random fields via coupling. Probab. Theory Related Fields 137 (2007), no. 1-2, 201–225.

[6] J.-R. Chazottes, P. Collet, F. Redig. On concentration inequalities and their applications for Gibbs measures in lattice systems. J. Stat. Phys. 169 (2017), no. 3, 504–546.

[7] J.-R. Chazottes, P. Collet, F. Redig. Evolution of concentration under lattice spin-flip dynamics. Preprint 2021 (arXiv:2103.08264 [math.PR]).

[8] B. Davis. On the $L_p$ norms of stochastic integrals and other martingales, Duke Math. J. 43, 697-704 (1976).

[9] R. M. Dudley. Real Analysis and Probability. Cambridge Studies in Advanced Mathematics vol. 74. Cambridge University Press, 2002

[10] A.C.D. van Enter, R. Fernández, F. den Hollander, F. Redig. Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures. Comm. Math. Phys. 226 (2002), no. 1, 101–130.

[11] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. Ann. Probab., Volume 32, Number 3B (2004), 2702-2732.

[12] M.Z. Guo, G.C. Papanicolaou, S.R.S. Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. Comm. Math. Phys. 118 (1988), no. 1, 31–59.

[13] M. Hirsch, S. Smale, R. Devaney. Differential equations, dynamical systems, and an introduction to chaos. Second edition. Pure and Applied Mathematics (Amsterdam), 60. Elsevier/Academic Press, Amsterdam, 2004.

[14] R. Khasminskii. Stochastic Stability of Differential Equations. Springer, 2012.

[15] C. Külske. Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. Comm. Math. Phys. 239 (2003), no. 1-2, 29-51.

[16] M. Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs, 89. American Mathematical Society, 2001.

[17] P. Mattila. Geometry of Sets and Measures in Euclidean Spaces. Cambridge University Press, 1995.
[18] I. Miyadera. Nonlinear semigroups. Translations of Mathematical Monographs 109. American Mathematical Society, Providence, RI, 1992.

[19] K. Narita. No explosion criteria for stochastic differential equations. J. Math. Soc. Japan 34, No. 2, 1982.

[20] B. Øksendal. Stochastic differential equations. An introduction with applications. Sixth edition. Universitext. Springer, 2003.

[21] W. F. Osgood. Beweis der Existenz einer Lösung der Differentialgleichung $dy/dx = f(x, y)$ ohne Hinzunahme der Cauchy-Lipschitzschen Bedingung, Monatshefte für Mathematik 9 (1898) 331–345.

[22] M.K. von Renesse, K.T. Sturm, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure and Applied Math., LVIII, (2005), 923-940.