Markov Processes Relat. Fields **8**, 1–36 (2002)

**Markov Processes** and **Related Fields**


# Stochastic Hamiltonian Systems: Exponential Convergence to the Invariant Measure, and Discretization by the Implicit Euler Scheme

**D. Talay**

Inria, 2004 Route des Lucioles, B.P. 93 06902 Sophia-Antipolis, France.
E-mail: denis.talay@inria.fr

**Abstract.** In this paper we carefully study the large time behaviour of

$$u(t,x,y) := \mathsf{E}_{x,y}\, f(X_t, Y_t) - \int f \, d\mu,$$

where $(X_t, Y_t)$ is the solution of a stochastic Hamiltonian dissipative system with non gbally Lipschitz coefficients, $\mu$ its unique invariant law, and $f$ a smooth function with polynomial growth at infinity. Our aim is to prove the exponential decay to 0 of $u(t,x,y)$ and all its derivatives when $t$ goes to infinity, for all $(x,y)$ in $\mathbf{R}^{2d}$.

We apply our precise estimates on $u(t,x,y)$ to analyze the convergence rate of a probabilistic numerical method based upon the implicit Euler discretization scheme which approximates $\int f \, d\mu$.

Keywords: stochastic differential equations, stochastic hamiltonian systems, parabolic partial differential equation, invariant measure, Euler method, simulation

AMS Subject Classification: Primary 60H10, 60H35, 60H30, 35K65; Secondary 60J60, 65U05

## 1. Introduction

Consider a stochastic differential system of the type

$$\begin{cases} X_t = X_0 + \int_0^t \partial_y H(X_s, Y_s)\, ds, \\ Y_t = Y_0 - \int_0^t \partial_x H(X_s, Y_s)\, ds - \int_0^t F(X_s, Y_s)\ \partial_y H(X_s, Y_s)\, ds + W_t, \end{cases} \tag{1.1}$$

where $X_t$, $Y_t$ and $W_t$ belong to $\mathbf{R}^d$. The process $(W_t)$ is a standard Brownian motion. Here and in all the paper, for all function $\phi$ defined on $\mathbf{R}^d \times \mathbf{R}^d$, we denote by $\partial_x \phi$ the vector $(\partial\phi/\partial x_i, 1 \leq i \leq d)$; the notation $\partial_y \phi$ is defined similarly.

Under the assumptions stated below, the Hamiltonian function $H$ and the function $F$ are such that there exists a unique global solution to equation (1.1) and this solution is an ergodic process. Soize [20] describes the applications of such models in Mechanics, investigates the question of existence and uniqueness of the invariant law and, under appropriate constraints on $H$ and $F$, give explicit formulae for the density $p(x,y)$ of the unique invariant probability measure $\mu$ of $(X_t, Y_t)$. If such constraints are not satisfied, one must use numerial methods to approximate quantities of the type $\int f\, d\mu$.

The objective of this paper is two-fold. We first study the large time behaviour of

$$u(t,x,y) := (x,y) \longrightarrow \mathsf{E}_{x,y}\, f(X_t, Y_t) - \int f\, d\mu,$$

for all smooth function $f$ with polynomial growth at infinity. Our aim is to prove the exponential decay to 0 of $u(t,x,y)$ and all its derivatives when $t$ goes to infinity, for all $(x,y)$ in $\mathbf{R}^{2d}$. In analytical terms, that means that we prove the exponential decay in time of the solution of a degenerate parabolic partial differential equation with non globally Lipschitz coefficients, and of the spatial derivatives of this solution. In this context, under our hypotheses (see Hypothesis 1.1 below), our result extends those of, e.g., Bakry [1], Malrieu [9], Ganidis, Roynette and Vallois [5] (see also the references in these publications). We emphasize that these authors use hypercontractivity techniques whereas here we use variational techniques. We then apply our precise estimates on $u(t,x,y)$ to analyze the convergence rate of a probabilistic numerical method which approximates

$$I := \int_{\mathrm{R}^{2d}} f(x,y) \mu(dx, dy). \tag{1.2}$$

This probabilistic procedure avoids the numerical resolution of the stationary Fokker – Planck equation

$$L^*\mu = 0, \tag{1.3}$$

where $L^*$ is formal adjoint of the infinitesimal generator $L$ of the process $(X_t, Y_t)$. Such a resolution may be impossible, or extremely long, or numerically instable, for the two following reasons. First, the dimension of the state space may be very large, especially for models used in Random Mechanics models (think of multidimensional nonlinear oscillators, e.g.). Second, the generator $L$ is degenerate since there is no noise in the dynamics of $(X_t)$.

A natural way to approximate $I$ defined in (1.2) is to apply the ergodic theorem: given any initial distribution of $(X_0, Y_0)$, one has

$$I = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_s, Y_s)\, ds \quad \mathsf{P}-a.s. \tag{1.4}$$

A time discretization with step-size $h$ of the integral in the right hand side of (1.4) and the choice of a large time $T = Nh$ ($N \in \mathbf{N} - \{0\}$) lead to the approximation

$$I \cong \frac{1}{N} \sum_{p=1}^{N} f(X_{ph}, Y_{ph}). \tag{1.5}$$

As the explicit resolution of equation (1.1) is impossible, one then has to construct a process which approximates $(X_t, Y_t)$.

Such a procedure has been studied by Talay and Tubaro [22] for general stochastic differential equations

$$Z_t = Z_0 + \int_0^t b(Z_s)\, ds + \int_0^t \sigma(Z_s)\, dW_s. \tag{1.6}$$

Talay and Tubaro [22] propose to use the Euler method to discretize the system (1.6) in time in order to construct a discrete time Markov process $(Z^h_{ph}, p \in \mathbf{N})$. For example, the Euler approximation with step-size $h$ is

$$Z^h_{(p+1)h} = Z^h_{ph} + b(Z^h_{ph})h + \sigma(Z^h_{ph})\Delta^h_{p+1}W, \tag{1.7}$$

where

$$\Delta^h_{p+1}W := W_{(p+1)h} - W_{ph}. \tag{1.8}$$

Under appropriate conditions on the functions $b$ and $\sigma$ the process $(Z^h_{ph}, p \in \mathbf{N})$ is ergodic, and the simulation of a single trajectory of $(Z^h_{ph})$ provides the following approximation of $I$:

$$I \cong I^{h,N} := \frac{1}{N} \sum_{p=1}^{N} f\big(Z^h_{ph}\big).$$

The global error $I - I^{h,N}$ depends on the two numerical parameters $h$ and $N$ and can be decomposed as follows:

$$I - I^{h,N} = \underbrace{I - \int_{\mathrm{R}^{2d}} f(x,y)\mu^h(dx,dy)}_{e_d(h)} + \underbrace{\int_{\mathrm{R}^{2d}} f(x,y)\mu^h(dx,dy) - I^{h,N}}_{e_s(h,N)},$$

where $\mu^h$ is the unique invariant probability measure of $(Z^h_{ph})$. The statistical error $e_s(h,N)$ can be estimated owing to the Central Limit Theorem for $\sqrt{N}e_s(h,N)$ which converges in distribution to a Gaussian law with zero mean (see, e.g., Revuz [17]). In addition, the variance of the limit law can usually be bounded from above uniformly in $h$ owing to results such as Lemma 4.1 below. Numerical experiments show that the discretization error $e_d(h)$, which of course is independent of the simulation time $Nh$, may be extremely large even for small values of $h$. This is explained by the following expansion obtained by Talay and Tubaro [22]:

$$e_d(h) = C_1 h + C_2 h^2 + \cdots + C_K h^K + \mathcal{O}\left(h^{K+1}\right) \qquad \text{for all } K \in \mathbf{N} - \{0\}. \tag{1.9}$$

This estimate justifies the use of the Romberg extrapolation method in order to accelerate the convergence rate: discretize (1.6) with the step-sizes $h$ and $h/2$, and consider the new approximation

$$I \cong 2I^{h/2,N} - I^{h,N} = \frac{1}{N}\sum_{p=1}^{N} f\big(Z^h_{ph}\big) - \frac{1}{2N}\sum_{p=1}^{2N} f\big(Z^{h/2}_{ph}\big). \tag{1.10}$$

The corresponding discretization error is

$$E_d(h) := I - \Big\{2\int_{\mathrm{R}^{2d}} f(x,y)\mu^{h/2}(dx,dy) - \int_{\mathrm{R}^{2d}} f(x,y)\mu^h(dx,dy)\Big\}.$$

The expansion (1.9) implies that $E_d(h) = \mathcal{O}\left(h^2\right)$.

To obtain estimate (1.9), Talay and Tubaro need several technical assumptions:

A. The functions $b^i$ and $\sigma^i_j$ are of class $\mathcal{C}^\infty(\mathbf{R}^d)$ with bounded derivatives of all order; the function $\sigma$ is bounded.

B. There exists a strictly positive real number $\nu$ such that

$$\sum_{i,j=1}^{d} \left(a^i_j(x)\xi^i\xi^j\right) \geq \nu|\xi|^2 \quad \text{for all } x \in \mathbf{R}^d \text{ and } \xi \in \mathbf{R}^d, \tag{1.11}$$

where $a(x)$ denotes the matrix $\sigma(x)\sigma(x)^*$.

C. There exist strictly positive real numbers $\beta$ and $C$ such that

$$x \cdot b(x) \leq -\beta|x|^2 + C \quad \text{for all } x \in \mathbf{R}^d. \tag{1.12}$$

As we are motivated by models in Random Mechanics, these assumptions are too stringent. First, the coefficient $\partial_x H$ of (1.1) may not be globally Lipschitz: see, e.g., the nonlinear oscillator with cubic restoring force. Second, as already mentioned, condition (1.11) is not satisfied by systems where the noise acts on certain coordinates only. In addition, the ordinary Euler scheme (1.7) maybe unstable when the coefficients of the differential system (1.1) are unbounded: see below. Therefore a new error analysis is needed. Our objective is to construct an implicit Euler scheme which has some good stability properties (see Section 4), and to prove that the discretization error satisfies (1.9).

*Remark* 1.1. In the internal report [23] the following laboratory example has been treated in detail:

$$\begin{cases} X_t = X_0 + \displaystyle\int_0^t Y_s \, ds, \\ Y_t = Y_0 - \displaystyle\int_0^t (Y_s + X_s^3) \, ds + W_t, \end{cases} \tag{1.13}$$

where $(W_t)$ is a real valued Brownian motion and the random variables $X_0$ and $Y_0$ have finite moments of all order. The present paper generalizes the technique and the result to systems of type (1.1).

We also underline that we only consider discretization schemes with a constant step-size. For other methods, see the review paper by Pagès [15] and the references therein, particularly the paper by Lamberton and Pagès [8].

**Two useful equalities and assumptions.** We define the differential operator $L$ by

$$\boxed{L\phi := \partial_y H \cdot \partial_x \phi - (\partial_x H + F \ \partial_y H) \cdot \partial_y \phi + \frac{1}{2} \sum_{i=1}^{d} \ \partial_{y_i y_i} \phi.} \tag{1.14}$$

We will often use the following obvious equalities:

$$\boxed{L(\phi\psi) = \phi \ L\psi + \psi \ L\phi + \sum_{i,j=1}^{d} \ \partial_y \phi \cdot \partial_y \psi,} \tag{1.15}$$

and, in particular,

$$\boxed{L(\phi^2) = 2\phi\ L\phi + \sum_{i,j=1}^{d} |\partial_y \phi|^2.} \tag{1.16}$$

In all the paper we make the following assumption:

**Hypothesis 1.1.** The function $F$ and $H$ are of class $\mathcal{C}^\infty(\mathbf{R}^{2d})$. The functions $F$ and $\partial_y H$ are bounded and all their derivatives are bounded. The function $H$ and its derivatives have a polynomial growth at infinity.
There exist (strictly) positive numbers $\delta$, $M$ and $\nu$, there exists a function $R$ on $\mathbf{R}^{2d}$ with second derivatives having a polynomial growth at infinity, such that

$$H(x,y) + R(x,y) + M \geq \delta\ (|x|^\nu + |y|^\nu), \tag{1.17}$$

$$0 < \nu \leq F(x,y) \leq M, \tag{1.18}$$

$$\nu|x|^2 + \nu|y|^2 \leq H(x,y) + M, \tag{1.19}$$

$$LH(x,y) + LR(x,y) \leq -\delta(H(x,y) + R(x,y)) + M, \tag{1.20}$$

$$|\partial_y H(x,y) + \partial_y R(x,y)|^2 \leq M\ (H(x,y) + R(x,y) + 1), \tag{1.21}$$

$$0 < \nu|\zeta|^2 \leq \sum_{i,j=1}^{d} (\partial_{y_i y_j} H(x,y) - \partial_{x_i y_j} H(x,y))\zeta_i \zeta_j, \tag{1.22}$$

$$0 < \nu|\zeta|^2 \leq \sum_{i,j=1}^{d} \partial_{y_i y_j} H(x,y)\zeta_i \zeta_j, \tag{1.23}$$

for all $x$, $y$, $\zeta$ in $\mathbf{R}^d$.

**Example 1.1.** For the system (1.13) one has $H(x,y) = \frac{1}{4}x^4 + \dfrac{1}{2}y^2$. One can choose $R(x,y) := \alpha\ x\ y$ with $\alpha$ positive and small enough.

**Example 1.2.** Consider the case where $F \equiv 1$ and

$$H(x,y) := V(x) + \frac{1}{2}|y|^2.$$

Suppose that $\nabla V$ is a Lipschitz function and satisfies

$$-\nabla V(x) \cdot x \leq -a|x|^2 + b, \tag{1.24}$$

$$0 \leq V(x) \leq k|x|^2 + \ell, \tag{1.25}$$

for some strictly positive real numbers $a$, $b$, $k$ and $\ell$. Let $0 < \alpha$ small enough (in particular, suppose $2k\alpha < a$). The inequalities (1.24) and (1.25) imply

$$-\nabla V(x) \cdot x \leq (-a + 2k\alpha)|x|^2 - 2\alpha V(x) + 2\alpha\ell + b. \tag{1.26}$$

To ensure (1.20) and (1.21) one can choose $M$ large enough and

$$R(x, y) := \frac{1-\alpha}{4}|x|^2 + \frac{1}{2}x \cdot y. \tag{1.27}$$

If, in addition, $V$ is such that (1.19) holds true, then (1.17) is also satisfied. The other inequalities obviously hold true.

## 2. Moments and ergodicity of the Hamiltonian system

In this section, we prove that $(X_t, Y_t)$ solution to (1.13) has a unique invariant measure. We have to face two dificulties:

- The drift coefficient of (1.1) does not satisfy a condition of the type (1.12). As it may even not be globally Lipschitz we introduce a Lyapunov function. This is a classical argument (see, e.g., Hasminskii [6]), but it seems somewhat new to apply it to examples such as 1.2.
- The differential operator $L$ is not uniformly strictly elliptic. To overcome this difficulty, we also use classical arguments (see, e.g., Campillo [3] for systems with discontinuous coefficients having a linear growth at infinity, and the introductory paper by Pagès [15]).

**Lemma 2.1.** (i) *The solution of the system* (1.1) *has moments of all order which are bounded uniformly in time.*

(ii) *The process $(X_t, Y_t)$ is a Feller process, which means that, for each bounded continuous function $g$ from $\mathbf{R}^2$ to $\mathbf{R}$, the map*

$$(x, y) \mapsto \mathsf{E}_{x,y}[g(X_t, Y_t)]$$

*is continuous.*

*Proof.* To prove (i), we introduce the positive Lyapunov function

$$\boxed{\Gamma(x, y) := H(x, y) + R(x, y) + M,} \tag{2.1}$$

where $R$ and $M$ are as in (1.20), $M$ being possibly increased to get the positivity of $\Gamma$. We aim to prove the sufficient condition

$$\boxed{\sup_{t\geq 0} \mathsf{E}\,\Gamma(X_t,Y_t)^m < \infty \quad \text{for all } m\in\mathbf{N}.} \tag{2.2}$$

We start by proving (2.2) for $m=1$. Let $\tau_N$ be the first exit time of the process $(X_t,Y_t)$ from the ball of radius $N$. From Itô's formula and (1.20) we deduce that there exist $\lambda>0$ and $M>0$ such that

$$\begin{aligned}\mathsf{E}\,\Gamma(X_{t\wedge\tau_N},Y_{t\wedge\tau_N}) &= \mathsf{E}\,\Gamma(X_0,Y_0) + \mathsf{E}\int_0^{t\wedge\tau_N} L\Gamma(X_\theta,Y_\theta)\,d\theta \\ &\leq \mathsf{E}\,\Gamma(X_0,Y_0) + \mathsf{E}\int_0^{t\wedge\tau_N} \big\{-\delta\Gamma(X_\theta,Y_\theta)+M\big\}\,d\theta.\end{aligned}$$

Let $N$ tend to infinity and use Fatou's lemma. It comes

$$\mathsf{E}\,\Gamma(X_t,Y_t) \leq \mathsf{E}\,\Gamma(X_0,Y_0) + Mt < \infty \quad \text{for all } t>0,$$

which, in view of (1.21), implies that

$$\mathsf{E}\int_0^t \partial_y\Gamma(X_s,Y_s)\,dW_s = 0 \qquad \text{for all } t\geq 0.$$

Thus

$$\frac{d}{dt}\,\mathsf{E}\,\Gamma(X_t,Y_t) = \mathsf{E}\,L\Gamma(X_t,Y_t) \leq -\delta\,\mathsf{E}\,\Gamma(X_t,Y_t) + M. \tag{2.3}$$

Differentiating $\exp(\delta t)\ \mathsf{E}\,\Gamma(X_t,Y_t)$ we finally get

$$\sup_{t\geq 0}\mathsf{E}\,\Gamma(X_t,Y_t) < \infty.$$

We have thus proved (2.2) for $m=1$. We now prove (2.2) by induction on $m$. The induction hypothesis, inequality (1.21) and Itô's formula applied to $\Gamma(X_t)$ allow us to write

$$\begin{aligned}&\frac{d}{dt}\big\{\,\mathsf{E}\,\Gamma(X_t,Y_t)^{m+1}\big\} \\ &\qquad = (m+1)\,\mathsf{E}\,\Gamma(X_t,Y_t)^m L\Gamma(X_t,Y_t) \\ &\qquad\quad + \frac{m(m+1)}{2}\,\mathsf{E}\big\{\Gamma(X_t,Y_t)^{m-1}\big|\partial_y\Gamma(X_t,Y_t)\big|^2\big\} \\ &\qquad \leq (m+1)\,\mathsf{E}\big\{\Gamma(X_t,Y_t)^m(-\delta\Gamma(X_t,Y_t)+M)\big\} \\ &\qquad\quad + \frac{m(m+1)}{2}\,\mathsf{E}\big\{\Gamma(X_t,Y_t)^{m-1}\big|\partial_y\Gamma(X_t,Y_t)\big|^2\big\}.\end{aligned} \tag{2.4}$$

In view of (1.21), one has

$$\begin{aligned} &\mathsf{E}\left\{\Gamma(X_t,Y_t)^{m-1}\left|\partial_y\Gamma(X_t,Y_t)\right|^2\right\} \\ &\quad\leq \left(\mathsf{E}\,\Gamma(X_t,Y_t)^m\right)^{(m-1)/m}\left(\mathsf{E}\left|\partial_y\Gamma(X_t,Y_t)\right|^{2m}\right)^{1/m} \\ &\quad\leq C_m\,\mathsf{E}\,\Gamma(X_t,Y_t)^m \end{aligned}$$

for some positive constant $C_m$. We thus have proved (i).

In view of the Lebesgue Dominated Convergence Theorem, the part (ii) of Lemma 2.1 results from the continuity of the stochastic flow

$$(x,y)\mapsto (X_\cdot^{x,y},Y_\cdot^{x,y}),$$

where $(X_t^{x,y},Y_t^{x,y})$ is the solution to (1.13) with initial condition $(X_0,Y_0)=(x,y)$, and this continuity results from the local Lipschitz property of the coefficients and the non explosion of the solution (see Protter [16, Chapter V, Theorem 38]). □

As the law of $(X_t^{x,y},Y_t^{x,y})$ has moments of all order which are uniformly bounded in time, the sequence of measures

$$\mu_n(\bullet) := \frac{1}{n}\int_0^n \mathsf{P}\{(X_s^{x,y},Y_s^{x,y})\in\bullet\}\,ds$$

is tight (see, e.g., Billingsley [2, Chapter 1]). It then is straightforward (see, e.g., Ethier and Kurtz [4]) to get

**Corollary 2.1.** *The process $(X_t,Y_t)$ has at least one invariant probability measure.*

We now examine the question of uniqueness of the invariant measure. To this end we first prove

**Lemma 2.2 (Step 2).** *For all law of $(X_0,Y_0)$ and all strictly positive time $t$, the law of $(X_t,Y_t)$ has a density $p(t,x,y)$ with respect to the Lebesgue measure, and this density is everywhere strictly positive.*

*Proof.* We use the two following arguments. First, we get the existence of the density $p(t,x,y)$ by using the Girsanov transformation. Second, we prove the strict positivity of $p(t,x,y)$ by studying the controllability of a deterministic system obtained by substituting deterministic controls to the 'noise' $(W_t)$.

It is easy to see that we do not restrict the problem by supposing that $(X_0,Y_0)$ is a deterministic vector $(x_0,y_0)$: if not, we use the Markov property of $(X_t,Y_y)$ and integrate the density corresponding to the initial condition $(x_0,y_0)$ with respect to the initial law of $(X_0,Y_0)$.

Set

$$\begin{aligned}
X_t^0 &:= x_0 + \int_0^t \partial_y H(X_s^0, Y_s^0)\, ds, \\
Y_t^0 &:= y_0 + W_t, \\
M_t^0 &:= -\int_0^t \partial_x H(X_s^0, Y_s^0)\, dW_s - \int_0^t F(X_s^0, Y_s^0)\ \partial_y H(X_s^0, Y_s^0)\, dW_s, \\
Z_t^0 &:= \exp\Big(M_t^0 - \frac{1}{2}\langle M^0\rangle_t\Big).
\end{aligned}$$

Classical arguments (see, e.g., Ustünel and Zakaï [24, Theorem 2.4.2]) show that the law of the random variable $(X_t, Y_t)$ is absolutely continuous with respect to the law of the pair $(X_t^0, Y_t^0)$, and that

$$\mathsf{E}_{x,y}\, \Psi(X_t, Y_t) = \mathsf{E}_{x,y}\left[\Psi(X_t^0, Y_t^0)\ Z_t^0\right] \tag{2.5}$$

for all bounded measurable real valued function $\Psi$. In view of (1.23) the generator of $(X_t^0, Y_t^0)$ is hypo-elliptic and thus the law of $(X_t^0, Y_t^0)$ has a density. Therefore equality (2.5) shows that the random variable $(X_t, Y_t)$ has a density $p(t, x, y)$ with respect to Lebesgue's measure.

To get the everywhere positivity of $p(t, x, y)$ we use a result of Michel and Pardoux [12, Section 3.3.6.1]: it is sufficient to show that for all $t$ in $\mathbf{R}_+ - \{0\}$ and all $(x, y)$ in $\mathbf{R}^{2d}$, the set

$$A(t, x, y) := \{(X_t^{x,y}(u), Y_t^{x,y}(u)),\ u \in H^1(\mathbf{R}_+; \mathbf{R}^d)\}$$

is equal to $\mathbf{R}^{2d}$, where $(X_t^{x,y}(u), Y_t^{x,y}(u))$ solves

$$\begin{cases}
X_t^{x,y}(u) = x + \displaystyle\int_0^t \partial_y H(X_s^{x,y}(u), Y_s^{x,y}(u))\, ds, \\
Y_t^{x,y}(u) = y - \displaystyle\int_0^t \partial_x H(X_s^{x,y}(u), Y_s^{x,y}(u))\, ds \\
\qquad - \displaystyle\int_0^t F(X_s^{x,y}(u), Y_s^{x,y}(u))\ \partial_y H(X_s^{x,y}(u), Y_s^{x,y}(u))\, ds \\
\qquad + u_t.
\end{cases} \tag{2.6}$$

Owing to Hypothesis 1.1 it is easy to check that the set $A(t, x)$ is equal to the whole space (e.g., one can adapt the argument used in the proof of the lemma 3.4 in Mattingly, Stuart and Higham [10]). □

**Corollary 2.2.** *The process $(X_t, Y_t)$ has a unique invariant probability measure. This probability measure has an everywhere strictly positive density $p(x, y)$ with respect to the Lebesgue measure.*

*Proof.* Lemma 2.2 implies that any invariant probability measure $\mu$ has an everywhere strictly positive density $p(x, y)$ with respect to the Lebesgue measure since $\mu$ is left invariant by the transition operator of $(X_t, Y_t)$. Thus two invariant probability measures are equivalent. The ergodic theorem (see, e.g., Skorokhod [19, Chapter 1,Theorem 7]) implies that they are identical since one has

$$\mu(A) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathsf{E}_{x,y}\, \mathbf{I}_A(X_s, Y_s)\, ds, \tag{2.7}$$

for all Borel subset $A$ of $\mathbf{R}^{2d}$ and $\mu$-almost $(x, y)$. □

We are going to prove that the invariant measure $\mu$ has finite moments of all order. To this end we prove the following statement which will be useful for other purposes in the sequel.

**Lemma 2.3.** *The process $(X_t, Y_t)$ has finite moments of all order. In addition, for all integer $m$, there exist integers $K_m$, $k_m$ and a strictly positive real number $\lambda_m$ such that*

$$\begin{aligned} &\mathsf{E}_{x_0,y_0}\left\{|X_t|^m + |Y_t|^m\right\} \\ &\qquad \le \quad K_m\big(1 + |x_0|^{k_m} \exp(-\lambda_m t) + |y_0|^{k_m} \exp(-\lambda_m t)\big) \end{aligned} \tag{2.8}$$

*for all $t > 0$.*

*Proof.* Let the function $\Gamma$ be defined as in (2.1). Remember that $\Gamma(X_t, Y_t)$ has finite moments of all order. We again use an induction to prove the following inequality which implies (2.8):

$$\begin{aligned} &\forall m \in \mathbf{N},\ \exists Q_m(x, y),\ \exists C_m > 0,\ \exists \lambda_m > 0, \\ &\qquad\qquad \mathsf{E}_{x,y}\, \Gamma(X_t, Y_t)^m \le C_m + Q_m(x, y) \exp(-\lambda_m t), \end{aligned} \tag{2.9}$$

where the $Q_m(x, y)$'s are polynomial functions. For $m = 1$, Inequality (2.9) readily follows from (2.3). Now suppose that (2.9) holds for all integer up to $m$. From (2.4) it follows that

$$\frac{d}{dt}\, \mathsf{E}\, \Gamma(X_t, Y_t)^{m+1} \le -\delta(m+1)\, \mathsf{E}\, \Gamma(X_t, Y_t)^{m+1} + C_m(1 + \mathsf{E}\, \Gamma(X_t, Y_t)^m)$$

for some integer $C_m$ depending on $m$ only. The desired result follows. □

**Lemma 2.4.** *The probability measure $\mu$ has finite moments of all order.*

*Proof.* Let $K$ be a compact subset $K$ of $\mathbf{R}^{2d}$. We deduce from (2.7) and Lemma 2.3 that

$$\int_K (|x|^m + |y|^m)\mu(dx, dy) = \lim_{t\to+\infty} \frac{1}{t} \int_0^t \mathsf{E}\left[(|X_s|^m + |Y_s|^m)\mathbf{I}_K(X_s, Y_s)\right] ds.$$

In view of Lemma 2.3, we then have

$$\int_K (|x|^m + |y|^m)\mu(dx, dy) \le C,$$

for some real number $C$ independent of the set $K$. The conclusion is obtained by letting $K$ increase to $\mathbf{R}^{2d}$ and applying Fatou's lemma. □

## 3. Exponential decay in time of the solution of a degenerate parabolic equation with non globally Lipschitz coefficients

We set

$$\boxed{u(t, x, y) := \mathsf{E}_{x,y}\, f(X_t, Y_t) - \int_{\mathrm{R}^{2d}} f\, d\mu.} \tag{3.1}$$

The objective of this section is to prove the following result which is interesting in itself.

**Theorem 3.1.** *Suppose that $f$ is a smooth function, and that all its derivatives have a polynomial growth at infinity.*
*Let $D^m u(t)$ denote the vector of the derivatives of order $m$ of the mapping*

$$(x, y) \mapsto u(t, x, y).$$

*For any integer $m$ there exist an integer $s$ and strictly positive real numbers $C$ and $\gamma$ such that*

$$|D^m u(t)| \le C(1 + |x|^s + |y|^s)\exp(-\gamma t) \tag{3.2}$$

*for all $t > 0$ and $(x, y) \in \mathbf{R}^{2d}$.*

The proof of Theorem 3.1 is long. One needs to carefully adapt the method of proof of Talay [21], where the above hypotheses A–C (see the Introduction) where used in force to get a similar statement.

We first need to prove that $u(t, x, y)$ is a smooth function and that its growth at infinity w.r.t. to the spatial variables is polynomial. This step may be difficult for a general Hamiltonian system because the polynomial growth of the coefficients combined with the degeneracy of $L$ does not allow the use of classical results on parabolic partial differential equations. The Girsanov transformation that we use here is one way to get rid of the difficulty.

**Lemma 3.1.** *Suppose that $f$ is a function of class $\mathcal{C}^\infty(\mathbf{R}^{2d})$, and that all its derivatives have a polynomial growth at infinity. Then $u(t,x,y)$ is infinitely differentiable with respect to $(x,y)$. There exist strictly positive real numbers $C_0$ and $\gamma_0$ and an integer $s_0$ such that*

$$|u(t,x,y)| \le C_0(1 + |x|^{s_0}\exp(-\gamma_0 t) + |y|^{s_0}\exp(-\gamma_0 t)). \tag{3.3}$$

*In addition, for all integer $m$, there exists an integer $s_m$ such that*

$$\forall T > 0,\ \exists C_m(T),\ |D^m u(t,x,y)| \le C_m(T)(1 + |x|^{s_m} + |y|^{s_m}) \tag{3.4}$$

*for all $(x,y) \in \mathbf{R}^{2d}$ and $t \in [0,T]$.*

*Proof.* Inequality (3.3) follows from Lemma 2.3 and the assumptions made on $f$. In addition, one can permute expectation and differentiations with respect to $(x,y)$ in the right hand side of equality (2.5) with $\Psi = f$. Inequality (3.4) follows readily. □

The proof of Theorem 3.1 then proceeds as follows.

1. We first show that, for any ball $B$, there exist strictly positive real numbers $C$ and $\lambda$ such that

$$\int_B |u(t)|^2 d\mu \le C\exp(-\gamma t) \quad \text{for all } t > 0. \tag{3.5}$$

2. We then show that the preceding inequality also holds for any spatial derivative of $u(t)$ (possibly with different real numbers $C$ and $\gamma$). As $\mu$ has a smooth and strictly positive density with respect to Lebesgue's measure, we deduce from the Sobolev imbedding theorem that, for any ball $B$ in $\mathbf{R}^{2d}$, there exist strictly positive real numbers $C$ and $\gamma$ such that

$$\forall (x,y) \in B,\ |u(t,x,y)| \le C\exp(-\gamma t) \quad \text{for all } t > 0.$$

3. Then we show that there exists strictly positive real numbers $C$ and $\gamma$ such that

$$\int |u(t)|^2 \pi_s(x,y)\,dx\,dy \le C\exp(-\gamma t) \quad \text{for all } t > 0,$$

where the function $\pi_s$ is defined as

$$\pi_s(x,y) = \frac{1}{\Gamma(x,y)^s}$$

for some integer $s$.

4. Finally, we prove that the prededing inequality also holds for any spatial derivative of $u(t)$ (possibly with different real numbers $s$, $C$ and $\gamma$), and then we conclude by applying the Sobolev imbedding theorem again.

*Remark* 3.1. In all the computations thereafter, time integrations over finite intervals of $u(t)$ and of its spatial derivatives are allowed owing to inequality (3.4).

### 3.1. Estimates on $u(t)$ and its first derivatives in $L^2(\mu)$

The aim of this subsection is to prove the following inequalities: there exists a positive real number $\gamma > 0$ such that

$$\int |u(t)|^2 d\mu \quad \leq \quad C\exp(-\gamma t) \quad \text{for all } t > 0, \tag{3.6}$$

$$\int \left|\partial_x u(t)\right|^2 d\mu \quad \leq \quad C\exp(-\gamma t) \quad \text{for all } t > 0, \tag{3.7}$$

$$\int \left|\partial_y u(t)\right|^2 d\mu \quad \leq \quad C\exp(-\gamma t) \quad \text{for all } t > 0. \tag{3.8}$$

The degeneracy of the operator $L$ (observe that, in its definition (1.14), there is no second derivative with respect to $x$) induces a deep adaptation of the method of proof introduced in Talay [21] for the case of uniform strictly elliptic operators. The new tool is the Lemma 3.4 below: the specific structure of the drift of the ergodic Hamiltonian system under consideration is used to compensate the degeneracy of the generator $L$.

In view of Lemma 3.1, Itô's formula and standard calculation show that $u(t,x,y)$ satisfies

$$\boxed{\begin{cases} \dfrac{\partial u}{\partial t}(t,x,y) = Lu(t,x,y) \quad \text{for all } t > 0 \text{ and } (x,y) \in \mathbf{R}^{2d}, \\ u(0,x,y) = f(x,y) - \displaystyle\int_{\mathrm{R}^{2d}} f\, d\mu. \end{cases}} \tag{3.9}$$

**Lemma 3.2.** *There exist strictly positive real numbers $\gamma_0$ and $C$ such that*

$$\int |u(t)|^2 d\mu \leq C\exp(-\gamma_0 t) \quad \textit{for all } t \geq 0. \tag{3.10}$$

*In addition, for all positive polynomial function $P$, there exist strictly positive real numbers $\gamma_0^P$ and $C$ such that*

$$\int P\ |u(t)|^2\, d\mu \leq C\exp(-\gamma_0^P t) \quad \textit{for all } t \geq 0. \tag{3.11}$$

*Proof.* We first observe that, in view of Lemma 3.1, inequality (3.10) and Cauchy – Schwarz inequality imply inequality (3.11). We thus have now to prove (3.10) only.

From equality (1.16) one has

$$\frac{d}{dt}|u(t)|^2 - L(|u(t)|^2) = -|\partial_y u(t)|^2, \tag{3.12}$$

so that, in view of (1.3), one has

$$\frac{d}{dt}\int |u(t)|^2 d\mu \le 0.$$

Therefore it is sufficient to prove (3.10) for all $t = n\rho$ for some strictly positive real number $\rho$.

We now proceed as in Section 6.1.1 of Talay [21]. Consider the ergodic Markov chain $(X_{n\rho}, Y_{n\rho})$. In view of (1.17) and (1.20), for $\rho$ small enough we easily obtain that

$$\begin{aligned} &\exists\alpha > 0,\ \exists R > 0, \\ &\sup_{|(x,y)|\ge R} \mathsf{E}\left[(1+\alpha\rho)\Gamma(X_{(n+1)\rho}, Y_{(n+1)\rho}) - \Gamma(x,y) \;\middle|\; (X_{n\rho}, Y_{n\rho}) = (x,y)\right] \quad < \quad 0. \end{aligned} \tag{3.13}$$

Therefore (see, e.g., Nummelin [14, Chap.5,6]) the chain is geometrically recurrent, so that (see, e.g., Meyn and Tweedie [11]) there exist strictly positive real numbers $C$ and $\gamma$ such that, for all integer $n$,

$$\int \left|\mathsf{E}_{x,y}\, f(X_{n\rho}, Y_{n\rho}) - \int f d\mu\right| \mu(dx, dy) \le C\exp(-\gamma n\rho).$$

In view of inequality (3.3), one therefore has

$$\begin{aligned} \int |u(n\rho)|^2 d\mu &\le C_0 C\exp(-\gamma n\rho) \\ &\qquad + C_0 \int \left\{|x|^{k_0} + |y|^{k_0}\right\} \exp(-\gamma_0 n\rho) u(n\rho)\, d\mu, \end{aligned}$$

from which, using (3.3) again and Lemma 2.4, we deduce

$$\forall n \in \mathbf{N},\ \int |u(n\rho)|^2 d\mu \le C\exp(-\gamma n\rho) \tag{3.14}$$

for some new strictly positive real numbers $C$ and $\gamma$. That ends the proof of inequality (3.10). □

We are now going to prove that inequalities similar to (3.10) hold for the first order spatial derivatives of $u(t, x, y)$. To this end, we need two preliminary lemmas.

**Lemma 3.3.** *There exist strictly positive real numbers $C$ and $\gamma_1 < \gamma_0$ such that*

$$\exp(\gamma_1 T) \int \left|u(T)\right|^2 d\mu + \int_0^T \exp(\gamma_1 t) \int \left|\partial_y u(t)\right|^2 d\mu\, dt \leq C \tag{3.15}$$

*for all $T > 0$. Let $P(x, y)$ be a positive polynomial function. There exist strictly positive real numbers $C_P$ and $\gamma_1^P$ such that*

$$\exp(\gamma_1^P T) \int P\, u(T)^2 d\mu + \int_0^T \exp(\gamma_1^P t) \int P \left|\partial_y u(t)\right|^2 d\mu\, dt \leq C_P \tag{3.16}$$

*for all $T > 0$.*

*Proof.* We fix $\gamma_1 < \gamma_0$, where $\gamma_0$ is as in Lemma 3.2. Using (1.16) we get

$$\begin{aligned}&\frac{d}{dt}\big[\exp(\gamma_1 t)\ |u(t)|^2\big]\\ &\qquad = \gamma_1 \exp(\gamma_1 t)\ |u(t)|^2 + \exp(\gamma_1 t) L\big(|u(t)|^2\big) - \exp(\gamma_1 t)\ |\partial_y u(t)|^2,\end{aligned}$$

from which, integrating with respect to $t$,

$$\begin{aligned}\exp(\gamma_1 T) u(T)^2 - |u(0)|^2 = {}&\gamma_1 \int_0^T \exp(\gamma_1 t)|u(t)|^2 dt + \int_0^T \exp(\gamma_1 t) L\big(|u(t)|^2\big) dt\\ &- \int_0^T \exp(\gamma_1 t)|\partial_y u(t)|^2\, dt.\end{aligned}$$

We now integrate with respect to $\mu$. We use inequality (3.10) and the fact that $\mu$ solves equation (1.3). Choose $\gamma_1 \leq \gamma_0$. It comes:

$$\begin{aligned}&\exp(\gamma_1 T) \int u(T)^2 d\mu + \int_0^T \exp(\gamma_1 t) \int \left|\partial_y u(t)\right|^2 d\mu\, dt\\ &\qquad \leq \int |u(0)|^2 d\mu + \gamma_1 \int_0^T \exp(\gamma_1 t) \int |u(t)|^2 d\mu\, dt \ \leq\ C.\end{aligned}$$

Similarly, for $\gamma_1^P < \min(\gamma_0, \gamma_1)$, successively using equality (1.15) with $\phi = |u(t)|^2$ and $\psi = P$ and equality (1.3), one gets

$$
\begin{aligned}
\exp(\gamma_1^P T) \int P\, u(T)^2 d\mu + \int_0^T \exp(\gamma_1^P t) \int P \left|\partial_y u(t)\right|^2 d\mu\, dt \\
= \quad & \int P\, |u(0)|^2 d\mu + \gamma_1^P \int_0^T \exp(\gamma_1^P t) \int P\, |u(t)|^2 d\mu\, dt \\
& - \int_0^T \exp(\gamma_1^P t) \int LP\, |u(t)|^2 d\mu\, dt \\
& - 2 \int_0^T \exp(\gamma_1^P t) \int u(t)\, \partial_y u(t)\ \partial_y P\, d\mu\, dt \\
=: \quad & \int P\, |u(0)|^2 d\mu + A_1 + A_2 + A_3.
\end{aligned}
$$

We use inequality (3.11) to bound $A_1$ and $A_2$ from above uniformly in $T$. In addition, we observe that

$$
|A_3| \leq \int_0^T \exp(\gamma_1^P t) \{ \left|\partial_y P\right|^2 |u(t)|^2 + \left|\partial_y u(t)\right|^2 \} d\mu\, dt,
$$

and we use (3.11) and (3.15), from which (3.16) follows. □

**Lemma 3.4.** *There exist strictly positive real numbers $C$ and $\gamma_2$ such that*

$$
\begin{aligned}
& \int_0^T \exp(\gamma_2 t) \int |\partial_y u(t) - \partial_x u(t)|^2\, d\mu\, dt \\
& \qquad + \int_0^T \exp(\gamma_2 t) \int \left|\partial_y (\partial_{y_i} u(t) - \partial_{x_i} u(t))\right|^2 d\mu\, dt \leq C
\end{aligned}
\tag{3.17}
$$

*for all $T > 0$.*

*Proof.* In the calculation which follows, we take advantage of the inequality (3.16) and of the assumption (1.22) to substitute $(\partial_{y_i} u(t) - \partial_{x_i} u(t))\ (\partial_{x_j} u(t) - \partial_{y_j} u(t))$ to $(\partial_{y_i} u(t) - \partial_{x_i} u(t))\ \partial_{x_j} u(t)$.

One has[1]

$$
\begin{aligned}
&\frac{d}{dt}|\partial_{y_i}u(t)-\partial_{x_i}u(t)|^2\\
&= \quad 2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,L(\partial_{y_i}u(t)-\partial_{x_i}u(t))\\
&\qquad +2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\big(\partial_{y_iy_j}H\,\partial_{x_j}u(t)-\partial_{y_i}(\partial_{x_j}H+F\,\partial_{y_j}H)\,\partial_{y_j}u(t)\big)\\
&\qquad -2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\big(\partial_{x_iy_j}H\,\partial_{x_j}u(t)-\partial_{x_i}(\partial_{x_j}H+F\,\partial_{y_j}H)\,\partial_{y_j}u(t)\big)\\
&= \quad L\big(\big|\partial_{y_i}u(t)-\partial_{x_i}u(t)\big|^2\big)-\big|\partial_y(\partial_{y_i}u(t)-\partial_{x_i}u(t))\big|^2\\
&\qquad -2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,(\partial_{y_j}u(t)-\partial_{x_j}u(t))\,\partial_{y_iy_j}H\\
&\qquad +2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,\partial_{y_j}u(t)\,\partial_{y_iy_j}H\\
&\qquad -2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,\partial_{y_i}(\partial_{x_j}H+F\,\partial_{y_j}H)\,\partial_{y_j}u(t)\\
&\qquad +2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,(\partial_{y_j}u(t)-\partial_{x_j}u(t))\,\partial_{x_iy_j}H\\
&\qquad -2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,\partial_{y_j}u(t)\,\partial_{x_iy_j}H\\
&\qquad +2(\partial_{y_i}u(t)-\partial_{x_i}u(t))\,\partial_{x_i}(\partial_{x_j}H+F\,\partial_{y_j}H)\,\partial_{y_j}u(t).
\end{aligned}
$$

All the terms of the type

$$|(\partial_{y_i}u(t)-\partial_{x_i}u(t))\ \partial_{y_j}u(t)\ \mathcal{H}|$$

(where $\mathcal{H}$ is a function depending on the derivatives of $H$ only) are bound from above by

$$\varepsilon\big|\partial_y u(t)-\partial_x u(t)\big|^2+\frac{1}{\varepsilon}\ P\,|\partial_y u(t)|^2,$$

where $P$ is an appropriate positive polynomial function and $\varepsilon$ is a positive real number. One then uses the assumption (1.22). For some positive constant $\rho$ independent from $\varepsilon$ and new polynomial function $P$ one has

$$
\begin{aligned}
&\exp(\gamma_2 T)\int|\partial_y u(T)-\partial_x u(T)|^2\,d\mu+\int_0^T\exp(\gamma_2 t)\int\big|\partial_y(\partial_{y_i}u(t)-\partial_{x_i}u(t))\big|^2\,d\mu\,dt\\
&\qquad\le\quad \int|\partial_y u(0)-\partial_x u(0)|^2 d\mu+\frac{1}{\varepsilon}\int_0^T\exp(\gamma_2 t)\int P\,|\partial_y u(t)|^2\,d\mu\,dt\\
&\qquad\qquad +(\gamma_2+\rho\,\varepsilon-2\nu)\int_0^T\exp(\gamma_2 t)\int|\partial_y u(t)-\partial_x u(t)|^2\,d\mu\,dt.
\end{aligned}
$$

[1] In the right hand side of the next equalities we do not explicitly write the summation over the index $j$.

One then chooses $\varepsilon = \nu/\rho$ and $\gamma_2 < \min(\nu, \gamma_1)$. The conclusion follows from Lemma 3.3. □

**Lemma 3.5.** *There exist strictly positive real numbers $C$ and $\gamma_3 < \gamma_2$ such that*

$$\int \left|\partial_y u(T)\right|^2 d\mu \le C \exp(-\gamma_3 T) \quad \textit{for all } T > 0. \tag{3.18}$$

*Proof.* One has

$$\frac{d}{dt}\left(\left|\partial_{y_i} u(t)\right|^2\right) = L\left(\left|\partial_{y_i} u(t)\right|^2\right) - |\partial_y(\partial_{y_i} u(t))|^2 + 2\sum_{j=1}^{d} \partial_{y_i y_j} H\ \partial_{x_j} u(t)\ \partial_{y_i} u(t)$$
$$- 2\sum_{j=1}^{d} \partial_{y_i}(\partial_{x_j} H + F\ \partial_{y_j} H)\ \partial_{y_i} u(t)\ \partial_{y_j} u(t).$$

Notice that

$$|\partial_{x_j} u(t)\ \partial_{y_i} u(t)| \le |\partial_{x_j} u(t) - \partial_{y_j} u(t)|^2 + |\partial_y u(t)|^2.$$

Differentiate $\exp(\gamma_3 t)\ \left|\partial_{y_i} u(t)\right|^2$, and then integrate with respect to $\mu$ and $t$ successively. Choose $\gamma_3$ small enough and then use (3.17). The conclusion follows. □

**Lemma 3.6.** *There exist strictly positive real numbers $C$ and $\gamma_4 < \gamma_3$ such that*

$$\int \left|\partial_x u(T)\right|^2 d\mu \le C \exp(-\gamma_4 T) \quad \textit{for all } T > 0. \tag{3.19}$$

*Proof.* Similar computations as before lead to

$$\frac{d}{dt}|\partial_{x_i} u(t)|^2 = L\left(|\partial_{x_i} u(t)|^2\right) - |\partial_y(\partial_{x_i} u(t))|^2 + 2\sum_{j=1}^{d} \partial_{x_i y_j} H\ \partial_{x_j} u(t)\ \partial_{x_i} u(t)$$
$$- 2\sum_{j=1}^{d} \partial_{x_i}(\partial_{x_j} H + F\ \partial_{y_j} H)\ \partial_{y_j} u(t)\ \partial_{x_i} u(t).$$

Use:

$$\left|\partial_{x_i} u(t)\ \partial_{y_j} u(t)\ \partial_{x_i}(\partial_{x_j} H + F\ \partial_{y_j} H)\right| \le C\left|\partial_{x_i} u(t) - \partial_{y_i} u(t)\right|^2 + P\left|\partial_y u(t)\right|^2$$

for some polynomial function $P$ and real number $C$, and

$$|\partial_{x_i} u(t)\ \partial_{x_j} u(t)| \le C|\partial_x u(t)|^2 \le C\left|\partial_x u(t) - \partial_y u(t)\right|^2 + C|\partial_y u(t)|^2.$$

It comes:

$$\exp(\gamma_5 T) \int |\partial_{x_i} u(T)|^2 \, d\mu$$
$$\leq \int |\partial_{x_i} u(0)|^2 \, d\mu + C\gamma_5 \int_0^T \exp(\gamma_5 t) \int |\partial_x u(t) - \partial_y u(t)|^2 \, d\mu \, dt$$
$$+ \int_0^T \int P \, |\partial_y u(t)|^2 \, d\mu \, dt.$$

We now choose $\gamma_5$ small enough and use (3.16), (3.17). The conclusion follows.
□

### 3.2. Estimates on all order derivatives of $u(t)$ in $L^2(\mu)$

Our objective now is to extend Lemma 3.2, 3.5 and 3.6: we want to show that for all integer $m$ there exist strictly positive real numbers $C_m$ and $\gamma_m$ such that

$$\boxed{\int \left|D^m u(T)\right|^2 d\mu \leq C_m \exp(-\gamma_m T) \quad \text{for all } T > 0,} \tag{3.20}$$

where $D^m u(T)$ denotes the vector of all the spatial derivatives of $u(T)$ of order $m$. We proceed by induction on the order $m$. We suppose that inequality (3.20) holds up to $m$ and we want to obtain it for $m+1$. To this end, we have to prove several estimates (Lemmas 3.7–3.10 below) whose proofs are only sketched since they closely follow those of Lemmas 3.3–3.6.

**Lemma 3.7.** *Suppose that the induction hypothesis* (3.20) *holds. Let*

$$\partial_y D^m u(t)$$

*denote the vector whose coordinates are the derivatives in all the $y$ directions of all the coordinates of $D^m$. There exist strictly positive real numbers $C$ and $\gamma$ such that*

$$\exp(\gamma T) \int |D^m u(T)|^2 d\mu + \int_0^T \exp(\gamma t) \int \left|\partial_y D^m u(t)\right|^2 d\mu \, dt \leq C \tag{3.21}$$

*for all $T > 0$. Let $P(x, y)$ be a positive polynomial function. There exist strictly positive real numbers $C_P$ and $\gamma_P$ such that*

$$\exp(\gamma_P T) \int P \, |D^m u(T)|^2 d\mu + \int_0^T \exp(\gamma_P t) \int P \left|\partial_y D^m u(t)\right|^2 d\mu \, dt \leq C_P \tag{3.22}$$

*for all $T > 0$.*

*Proof.* In view of equality (1.16), we have

$$\frac{d}{dt}\big[\exp(\gamma t)\big|D^m u(t)\big|^2\big]$$
$$= \gamma \exp(\gamma t)\big|D^m u(t)\big|^2 + \exp(\gamma t) L\big(\big|D^m u(t)\big|^2\big) - \exp(\gamma t)\big|\partial_y D^m u(t)\big|^2.$$

We then proceed as in the proof of Lemma 3.3: we first integrate with respect to $t$, and then with respect to $\mu$. Finally, we use the induction hypothesis (3.20). $\square$

The next lemma is analogous to Lemma 3.4.

**Lemma 3.8.** *Suppose that the induction hypothesis* (3.20) *holds. Let* $D_x^m u(t)$ *denote an arbitrary partial derivative of* $u(t)$ *of the type* $\partial_{x_{i_1}} \dots \partial_{x_{i_m}}$. *There exist strictly positive real numbers* $\lambda$, $\gamma$ *and* $C$ *depending on* $m$ *such that*

$$\int_0^T \exp(\gamma t) \int \big|\lambda\, \partial_y(D_x^m u(t)) - \partial_x(D_x^m u(t))\big|^2 \, d\mu \, dt \le C \quad \text{for all } T > 0. \tag{3.23}$$

*Proof.* Denote by $D^m_{x/x_k}$ any differential operator of order $m-1$ such that $D_x^m$ can be written under the form $\partial_{x_k} D^m_{x/x_k}$.

The guideline of the calculation below is as follows: the terms which include derivatives of order $m+2$ and $m+3$ of $u(t)$ are gathered to provide $L(\lambda\, \partial_{y_i}(D_x^m u(t)) - \partial_{x_i}(D_x^m u(t)))$; the terms which include derivatives of order $m+1$ in $x$ directions only are explicited and gathered in order to take advantage of the assumptions (1.22) and (1.23); all the other therms include, either derivatives of order $m+1$ with one derivative in a $y$ direction, or derivatives of order at most $m$: we respectively use Lemma 3.7 and the induction hypothesis (3.20).

For all $\lambda$ we have[2]

$$
\begin{aligned}
&\frac{d}{dt}\big|\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big|^2 \\
&\qquad\qquad = \quad 2\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad \times\big(\lambda\ \partial_{y_i}(D^m_x Lu(t)) - \partial_{x_i}(D^m_x Lu(t))\big) \\
&\qquad\qquad = \quad 2\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad \big\{L(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))) + \lambda\ \partial_{y_i y_j} H\ \partial_{x_j}(D^m_x u(t)) \\
&\qquad\qquad\qquad - \partial_{x_i y_j} H\ \partial_{x_j}(D^m_x u(t)) - \partial_{x_k y_j} H\ \partial_{x_i}(D^m_{x/x_k}(\partial_{x_j} u(t)))\big\} \\
&\qquad\qquad\qquad + 2\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad\ \times\ \text{terms including derivatives of order at most } m \text{ of } u(t) \\
&\qquad\qquad\qquad + 2\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad\ \times\ \text{terms including derivatives of order } m+1 \text{ of } u(t) \\
&\qquad\qquad\qquad\qquad \text{with one derivative in a } y \text{ direction.}
\end{aligned} \tag{3.24}
$$

Notice that

$$
\begin{aligned}
&\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\times\big(\lambda\ \partial_{y_i y_j} H\ \partial_{x_j}(D^m_x u(t)) - \partial_{x_i y_j} H\ \partial_{x_j}(D^m_x u(t))\big) \\
&\qquad\qquad = \quad -\lambda\ \big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad \times\partial_{y_i y_j} H\ (\lambda\ \partial_{y_j}(D^m_x u(t)) - \partial_{x_j}(D^m_x u(t))) \\
&\qquad\qquad\qquad + \lambda^2\ \big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big)\partial_{y_i y_j} H\ \partial_{y_j}(D^m_x u(t)) \\
&\qquad\qquad\qquad + \lambda\ \big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad \times\partial_{x_i y_j} H\ \big(\lambda\ \partial_{y_j}(D^m_x u(t)) - \partial_{x_j}(D^m_x u(t))\big) \\
&\qquad\qquad\qquad - \lambda\ \big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big)\partial_{x_i y_j} H\ \partial_{y_j}(D^m_x u(t)).
\end{aligned}
$$

For each $D^m_x$ sum over $i$, and use (1.22). In addition, proceed as in the proof of Lemma 3.4 to treat the second and last terms of the right hand side. It comes:

$$
\begin{aligned}
&\big(\lambda\ \partial_{y_i}(D^m_x u(t)) - \partial_{x_i}(D^m_x u(t))\big) \\
&\times\big(\lambda\ \partial_{y_i y_j} H\ \partial_{x_j}(D^m_x u(t)) - \partial_{x_i y_j} H\ \partial_{x_j}(D^m_x u(t))\big) \\
&\qquad \leq \quad (\varepsilon - (\lambda-1)\ \nu)\ \big|\lambda\ \partial_y(D^m_x u(t)) - \partial_x(D^m_x u(t))\big|^2 + \frac{1}{\varepsilon}\, P\, |\partial_y(D^m_x u(t))|^2
\end{aligned}
$$

for all $\varepsilon$ small enough and some positive polynomial function $P$. We now choose $\lambda$ positive and large enough to take advantage of (1.23).

[2] In the left and right hand sides of the next equalities we do not explicitly write the summation over the indices $i$, $j$ and $k$, over all the differential operators $D^m_x$ of the type $\partial_{x_{i_1}} \ldots \partial_{x_{i_m}}$, and over the families of operators $D^m_{x/x_k}$.

We now have to consider the sum of all the terms of the type

$$\big(\lambda\, \partial_{y_i}(D_x^m u(t)) - \partial_{x_i}(D_x^m u(t))\big)\; \partial_{x_k y_j} H\; \partial_{x_i}(D^m_{x/x_k}(\partial_{x_j} u(t))).$$

We use

$$\begin{aligned}&\big|\lambda\, \partial_{y_i}(D_x^m u(t))\; \partial_{x_k y_j} H\; \partial_{x_i}(D^m_{x/x_k}(\partial_{x_j} u(t)))\big|\\ &\qquad\qquad \le \varepsilon \big|\partial_y(D_x^m u(t)) - \partial_x(D_x^m u(t))\big|^2 + C\big|\partial_y(D_x^m u(t))\big|^2,\end{aligned}$$

so that it remains to consider the sums of terms of the type

$$\partial_{x_i}(D_x^m u(t))\; \partial_{x_k y_j} H\; \partial_{x_i}(D^m_{x/x_k}(\partial_{x_j} u(t))).$$

One can show that this sum is equal to the sum over $i$ and all the differential operators $D_x^{m-1}$ of the type $\partial_{x_{i_1}} \dots \partial_{x_{i_{m-1}}}$ of quantities such as

$$\sum_{j,k=1}^{d} \partial_{x_k y_j} H\; \partial_{x_k}(\partial_{x_i}(D_x^{m-1} u(t)))\; \partial_{x_j}(\partial_{x_i}(D_x^{m-1} u(t))).$$

One then uses (1.22) and proceed as above.

We then proceed as at the end of the proof of Lemma 3.4, applying the induction hypothesis (3.20) and Lemma 3.7. □

We now are in a position to prove that the induction hypothesis is also true for the derivatives of order $(m+1)$. We first prove a statement analogous to Lemma 3.4.

**Lemma 3.9.** *Suppose that the induction hypothesis* (3.20) *holds. Let* $\partial^m u(t)$ *denote an arbitrary partial derivative of* $u(t)$ *ot order* $m$. *There exist strictly positive real numbers* $C$ *and* $\gamma$ *such that*

$$\int \big|\partial_y \partial^m u(T)\big|^2 d\mu \le C \exp(-\gamma T) \quad \textit{for all } T > 0. \tag{3.25}$$

*Proof.* We have

$$\begin{aligned}
&\frac{d}{dt}\big|\partial_{y_i}(\partial^m u(t))\big|^2\\
&\qquad = 2\big(\partial_{y_i}(\partial^m u(t))\big)\;\big(\partial_{y_i}(\partial^m L u(t))\big)\\
&\qquad = 2\partial_{y_i}(\partial^m u(t))\big\{L(\partial_{y_i}(\partial^m u(t))) + \partial_{y_i y_j} H\; \partial_{x_j}(\partial^m u(t))\big\}\\
&\qquad\quad + 2\partial_{y_i}(\partial^m u(t))\\
&\qquad\qquad \times \text{ terms including derivatives of order at most } m \text{ of } u(t)\\
&\qquad\quad + 2\partial_{y_i}(\partial^m u(t))\\
&\qquad\qquad \times \text{ terms including derivatives of order } m+1 \text{ of } u(t)\\
&\qquad\qquad\quad \text{with one derivative in a } y \text{ direction.}
\end{aligned}$$

We then proceed as in the proof of Lemma 3.5, applying the induction hypothesis (3.20), Lemma 3.7 and distinguishing two cases: if $\partial^m$ contains at least one derivation in a $y$ direction, we use the conclusion of the Lemma 3.7; if $\partial^m$ contains derivatives in $x$ directions only, we bound from above

$$\left|\partial_{y_i}(\partial^m u(t))\ \partial_{y_i y_j} H\ \partial_{x_j}(\partial^m u(t))\right|$$

by

$$C\ \left|\partial_{x_j}(\partial^m u(t)) - \lambda\ \partial_{y_j}(\partial^m u(t))\right|^2 + C\ \left|\partial_{y_j}(\partial^m u(t))\right|^2,$$

and then use (3.23). □

**Lemma 3.10.** *Suppose that the induction hypothesis* (3.20) *holds. Let* $\partial^m u(t)$ *denote an arbitrary partial derivative of* $u(t)$ *of order* $m$. *There exist strictly positive real numbers* $C$ *and* $\gamma$ *such that*

$$\int \left|\partial_x \partial^m u(T)\right|^2 d\mu \leq C \exp(-\gamma T) \quad \text{for all } T > 0. \tag{3.26}$$

*Proof.* If $\partial^m$ contains at least one derivative in a $y$ direction one can apply Lemma 3.9. If not, consider (3.24) with $\lambda = 0$. We thus explicitly get

$$\frac{d}{dt}\left|\partial_x \partial^m u(t)\right|^2.$$

The calculation then follows the same guidelines as in the proofs of the Lemmas 3.6 and 3.8. We omit the details. □

*Remark* 3.2. From the inequality (3.20) one could prove that, for all integer $m$, there exist strictly positive real numbers $C$ and $\gamma$ such that

$$|D^m u(t,x,y)| \leq \frac{C}{p(x,y)} \exp(-\gamma t)$$

for all $t > 0$ and $(x,y) \in \mathbf{R}^{2d}$. Such an estimate is too rough to analyze the convergence rate of the implicit Euler scheme: in Section 5 we will need that $|D^m u(t)|$ is bounded from above by a function of the type $\exp(-\gamma t)\ g(x,y)$ where $g$ is such that $\mathsf{E}\, g(X^h_{(p+1)h}, Y^h_{(p+1)h})$ is finite. The objective of Subsections 3.3 and 3.4 is to prove that $g$ can actually be chosen as a polynomial function.

### 3.3. Pointwise estimates on $u(t)$ in a ball

**Lemma 3.11.** *For any ball* $B$ *in* $\mathbf{R}^{2d}$, *there exist strictly positive real numbers* $C^B$ *and* $\gamma^B$ *such that*

$$|u(t,x,y)| \leq C^B \exp(-\gamma^B t) \tag{3.27}$$

*for all* $t > 0$ *and* $(x,y) \in B$.

*Proof.* Since the density $p(x,y)$ of $\mu$ is everywhere strictly positive (see Lemma 2.2) and continous (indeed, in view of (2.5) one can easily prove that the density transition $p(t,x,y,\alpha,\beta)$ of $(X_t,Y_t)$ issued from $(x,y)$ is a continuous function of $(x,y,\alpha,\beta)$, from which one can conclude that $p(x,y)$ is continuous since $p(x,y)$ is left invariant by the integration with respect to the transition density), there exists a positive real number $C$ such that, for all integer $m$, for all time $t$,

$$\begin{aligned}\int_B |D^m u(t)|^2 dx\,dy &\leq C \int_B |D^m u(t)|^2\; p(x,y)\,dx\,dy \\ &\leq C \int_{\mathrm{R}^{2d}} |D^m u(t)|^2\, d\mu.\end{aligned} \tag{3.28}$$

We apply estimate (3.20). It comes:

$$\int_B |D^m u(t)|^2 dx\,dy \leq C \exp(-\gamma_m t),$$

for some new positive real number $C$. We conclude by choosing $m$ large enough and using the Sobolev imbedding theorem. □

### 3.4. Estimates on $u(t)$ and its derivatives in $L^2(\pi_s)$

Let the weight function $\pi_s$ be defined as

$$\pi_s(x,y) = \frac{1}{\Gamma(x,y)^s}, \tag{3.29}$$

for some integer $s$ that we choose as follows. Easy computations lead to[3]

$$L^*(\pi_s) = -s\,\frac{L\Gamma}{\Gamma(x,y)^{s+1}} + \partial_{y_i}(F\;\partial_{y_i}H)\;\pi_s + \frac{s(s+1)}{2}\frac{(\partial_y\Gamma)^2}{\Gamma^{s+2}} - s\;\frac{\partial_{y_i y_j}\Gamma}{\Gamma^{s+1}}.$$

Therefore, in view of (1.20) one has

$$L^*(\pi_s) \leq -\frac{\delta}{2}\;s\;\pi_s + \phi_s\;\pi_s \tag{3.30}$$

for some integer $s$ chosen large enough and for some function $\phi_s$ which tends to 0 at infinity. For each integer $n$ we define an integer $s_n$ by possibly increasing the value of $s$ in order that

$$\int_{\mathrm{R}^{2d}} |D^k u(t)\pi_{s_n}|^2 dx\,dy < \infty \text{ for all } 0 \leq k \leq n, \tag{3.31}$$

which is possible in view of Lemmas 3.1 and 2.3.

[3] Again we do not explicitly write the summation over $i$ and $j$.

*Remark* 3.3. The following observation is intensively used in the calculations of the rest of this subsection. For all multi-index $J$ and integer $s$, there exists a smooth function $\psi_{J,s}(x,y)$ such that

$$\partial_J \pi_s(x,y) = \psi_{J,s}(x,y)\pi_s(x,y) \tag{3.32}$$

with

$$\psi_{J,s}(x,y) \xrightarrow[|(x,y)|\to\infty]{} 0. \tag{3.33}$$

Thus, inequality (3.31) implies that it is possible to choose a new integer still denoted by $s_n$ such that one also has

$$\int_{\mathrm{R}^{2d}} |D^k(u(t)\pi_s)|^2 \, dx\, dy < \infty \tag{3.34}$$

for all $t > 0$, $s \geq s_n$ and $k \leq n$.

We now prove a statement which relies on and improves Lemma 3.2.

**Lemma 3.12.** *There exist strictly positive real numbers $s$, $C$ and $\lambda$ such that*

$$\int |u(t)|^2 \pi_s(x,y)\, dx\, dy \leq C \exp(-\lambda t) \quad \textit{for all } t > 0. \tag{3.35}$$

*Proof.* Let $s$ be an integer larger than $s_1$. We thus have that $D(u(t)\pi_s)$ belongs to $L^2(\mathbf{R}^{2d})$. Moreover, observe that

$$\begin{aligned} L^*(u(t)\pi_s) \quad = \quad & L^*(\pi_s)\ u(t) + \frac{1}{2}\ \partial_{y_i y_i} u(t)\ \pi_s + \partial_{y_i} u(t)\ \partial_{y_i}\pi_s \\ & - \partial_{y_i} H\ \partial_{x_i} u(t)\ \pi_s + \partial_{x_i y_i} H\ u(t)\ \pi_s + \partial_{x_i} H\ \partial_{y_i} u(t)\ \pi_s \\ & + \partial_{y_i}(F \partial_{y_i} H)\ u(t)\ \pi_s + F\ \partial_{y_i} H\ \partial_{y_i} u(t)\ \pi_s. \end{aligned}$$

Therefore, in view of (3.30) and (3.32), as the derivatives of $F$ and $\partial_y H$ are bounded, after having integrated by parts $\int \partial_{y_i} H\ \partial_{x_i} u(t)\ \pi_s\, dx\, dy$ one has

$$\begin{aligned} & \int u(t)\ Lu(t)\ \pi_s(x,y)\, dx\, dy \\ & \qquad \leq -\int |u(t)|^2 \pi_s(x,y)\, dx\, dy + \int |u(t)|^2\ |\phi_s(x,y)|\ \pi_s(x,y)\, dx\, dy \end{aligned}$$

for $s$ large enough and some function $\phi_s(x,y)$ which tends to 0 at infinity in view of (3.33). We now choose a ball $B$ in $\mathbf{R}^{2d}$ large enough to have

$$|\phi_s(x,y)| \leq \frac{1}{2} \quad \text{for all } (x,y) \in \mathbf{R}^{2d} - B.$$

Then

$$\int u(t)Lu(t)\pi_s(x,y)\,dx\,dy$$
$$\leq -\frac{1}{2}\int |u(t)|^2\pi_s(x,y)\,dx\,dy + \int_B |u(t)|^2\ |\phi_s(x,y)|\ \pi_s(x,y)\,dx\,dy.$$

From Lemma 3.11 we then deduce

$$\frac{d}{dt}\int |u(t)|^2\pi_s(x,y)\,dx\,dy \leq -\frac{1}{2}\int |u(t)|^2\pi_s(x,y)\,dx\,dy + C\exp(-\gamma^B t),$$

from which we readily get (3.35) by differentiating in time

$$\exp(\lambda t)\int |u(t)|^2\pi_s(x,y)\,dx\,dy$$

and then choosing $\lambda$ small enough. □

**Lemma 3.13.** *For all integer $m$, there exist strictly positive real numbers $C_m$ and $\lambda_m$ such that*

$$\int |D^m u(t)|^2\pi_s(x,y)\,dx\,dy \leq C_m\exp(-\lambda_m t) \quad \text{for all } t>0. \tag{3.36}$$

*Proof.* The lemma is proven by induction on $m$ and by combining arguments used in the proofs of the lemmas in Subsection 3.2 and of Lemma 3.12. We omit the details. □

### 3.5. Proof of Lemma 3.1

We are now in a position to prove the pointwise estimate (3.2) on $u(t)$ and its derivatives. They are obtained owing to the Sobolev imbedding theorem and the preceding estimates in $L^2(\pi_s)$ which imply: there exist strictly positive real numbers $C_m$ and $\lambda_m$ such that

$$\int \left|D^m\big(u(t)\pi_{s_m}\big)\right|^2 dx\,dy \leq C_m\exp(-\lambda_m t) \quad \text{for all } t>0. \tag{3.37}$$

## 4. Moments and ergodicity of the implicit Euler scheme

The explicit Euler scheme (1.7) is unsatisfying when applied to systems with non globally Lipschitz coefficients. For example, consider the one dimensional equation

$$\xi_t = -\int_0^t \xi_s^3 ds + W_t. \tag{4.1}$$

In Talay [23] and Mattingly, Stuart and Higham [10] one proves that

$$\mathsf{E}\,|\xi_t|^2 \leq \frac{3}{2} \quad \text{for all } t \geq 0, \tag{4.2}$$

whereas any moment of the ordinary Euler scheme

$$\tilde{\xi}^h_{(p+1)h} = \tilde{\xi}^h_{ph} - (\tilde{\xi}^h_{ph})^3 h + \Delta^h_{p+1} W \tag{4.3}$$

tends to infinity with $p$ for all step-size $h$ large enough, and any moment of the implicit Euler scheme

$$\xi^h_{(p+1)h} = \xi^h_{ph} - (\xi^h_{(p+1)h})^3 h + \Delta^h_{p+1} W \tag{4.4}$$

is uniformly bounded with respect to $p$ for all step-size $h$.

The preceding consideration suggests to use the implicit Euler scheme to discretize systems of the type (1.1) when the coefficients are not globally Lipschitz:

$$\left\{\begin{array}{lcl} X^h_{(p+1)h} & = & X^h_{ph} + \partial_y H(X_{(p+1)h}, Y_{(p+1)h})\ h, \\ Y^h_{(p+1)h} & = & Y^h_{ph} - \partial_x H(X_{(p+1)h}, Y_{(p+1)h})\ h \\ & & \quad - F(X_{(p+1)h}, Y_{(p+1)h})\ \partial_y H(X_{(p+1)h}, Y_{(p+1)h})\ h + \Delta^h_{p+1} W. \end{array}\right. \tag{4.5}$$

Our error analysis of the implicit Euler scheme requires the set of hypotheses (1.1) and the supplementary following assumptions:

**Hypothesis 4.1.** The Hessian matrix of $\Gamma$ is positive semidefinite.
In addition, for all $h$ small enough, the determinant of the matrix

$$\begin{pmatrix} Id_{\mathrm{R}^d} & h\ \partial_{yy} H(x,y) \\ -h\ \partial_{xx} H(x,y) & Id_{\mathrm{R}^d} \end{pmatrix} \tag{4.6}$$

is bounded from below by a strictly positive constant uniform with respect to $h$.

*Remark* 4.1. In view of Hypothesis 4.1 it is easy to check by induction that, for all step-size $h$, $(X^h_{ph}, Y^h_{ph})$ is well defined for all $p$.

**Example 4.1.** In the case of Example 1.2, the Hypothesis 4.1 is satisfied when the second derivatives of $V$ are bounded functions.

**Lemma 4.1.** *Suppose that the sets of Hypotheses* 1.1 *and* 4.1 *hold. The implicit Euler scheme* (4.5) *with all step-size $h$ small enough satisfies*:

(i) *For all integer $m$ there exist integers $K_m$ and $k_m$ such that*

$$\mathsf{E}_{x,y}\left\{|X^h_{ph}|^m + |Y^h_{ph}|^m\right\} \leq K_m(1 + |x|^{k_m} + |y|^{k_m}) \quad \textit{for all } p \in \mathbf{N}. \tag{4.7}$$

(ii) *This implicit Euler scheme has a unique invariant probability measure $\mu^h$. In addition, for all integer $m$ there exists a positive real number $C_m$ such that*
$$\int_{\mathrm{R}^{2d}} (|x|^m + |y|^m)\mu^h(dx, dy) < C_m. \tag{4.8}$$

*Proof.* Notice that we can derive estimate (4.8) from (4.7) by using the Markov property of the chain $(X^h_{ph}, Y^h_{ph})$ and the invariance of $\mu^h$.

We start by proving estimate (4.7). One obviously has

$$\begin{aligned} X^h_{ph} &= X^h_{(p+1)h} - \partial_y H(X_{(p+1)h}, Y_{(p+1)h})\ h, \\ Y^h_{ph} + \Delta^h_{p+1} W &= Y^h_{(p+1)h} + \partial_x H(X_{(p+1)h}, Y_{(p+1)h})\ h \\ &\quad + F(X_{(p+1)h}, Y_{(p+1)h})\ \partial_y H(X_{(p+1)h}, Y_{(p+1)h})\ h. \end{aligned}$$

Apply the function $\Gamma$ to the left and right hand sides of the preceding equalities. A Taylor expansion up to the second order and our hypothesis on the Hessian matrix of $\Gamma$ imply that

$$(1 + \lambda h)\ \Gamma(X^h_{(p+1)h}, Y^h_{(p+1)h}) \le \Gamma(X^h_{ph}, Y^h_{ph} + \Delta^h_{p+1} W) + C\, h$$

for some deterministic positive real numbers $C$ and $\lambda$ uniform with respect to $h$ and $p$. The estimate (4.7) readily follows.

We now prove the existence and uniqueness of the invaraint measure $\mu^h$. The inequality (4.7) implies the existence of an invariant probability measure. To prove the uniqueness, we will use standard techniques for Markov chains (see, e.g., Meyn and Tweedie [11], Shardlow and Stuart [18], Mattingly, Stuart and Higham [10]).

Our objective is to apply the following result stated in Meyn and Tweedie [11, Theorem 13.0.1]: we get the uniqueness of the invariant probability $\mu^h$ if we can prove that the chain $(X^h_{ph}, Y^h_{ph})$ is positive Harris recurrent, which means that

(a) The chain is $\psi$-irreducible for some measure $\psi$, in the sense that there exists a measure $\psi$ such that, for all Borel subset $A$ of $\mathbf{R}^{2d}$,

$$\begin{aligned} &\psi(A) > 0 \\ &\Longrightarrow \mathsf{P}\ \left\{ \min\{p \ge 1 : (X^h_{ph}, Y^h_{ph}) \in A\} < \infty \ \middle|\ (X^h_0, Y^h_0) = (x, y) \right\} > 0 \end{aligned}$$

for all $(x, y) \in \mathbf{R}^{2d}$.

(b) Every Borel set $A$ of strictly positive $\psi$-measure is Harris recurrent, i.e,

$$\mathsf{P}\left\{ (X^h_{ph}, Y^h_{ph}) \in A \text{ for an infinite number of } p \ \middle|\ (X^h_0, Y^h_0) = (x, y) \right\} > 0$$

for all $(x, y) \in \mathbf{R}^2$.

(c) The chain is positive, i.e., it admits an invariant probability measure.

For this definition, we refer to [11, p. 200, 230–231].

To prove (a), we apply Theorems 7.2.5 and 7.2.6 in [11]: it is enough to prove that the chain is forward accessible and a globally attracting state exists. We do not rewrite here the definition of the forward accessibility (see [11, p. 151]) since we use a sufficient condition. As shown in [11, Proposition 7.1.4], the forward accessibility is a consequence of the following fact: $G_u$ denoting the function

$$(x,y) \longrightarrow G_u(x,y) := \begin{pmatrix} x + h\ \partial_y H(x,y) \\ y - h\ (\partial_x H(x,y) + F(x,y)\ \partial_y H(x,y)) + u \end{pmatrix}$$

which, for all $h$ small enough, is one–to–one in view of our assumption on the matrix (4.6), the matrix

$$\left[ \nabla_{x,y} G_u^{-1} \cdot \frac{\partial G_u^{-1}}{\partial u}, \frac{\partial G_u^{-1}}{\partial u} \right]$$

has full rank for all $(x,y)$. Moreover, the existence of a globally attracting state (in the sense of [11, p. 160]) is also obvious: using our Lyapunov function $\Gamma$ it is easy to check that 0 is attracting for the chain defined by substituting 0 to the increments of $(W_t)$.

We now prove (b) and (c). In view of [11, Theorem 11.3.4], a sufficient condition is as follows (it involves the notion of petite sets as defined in [11, p. 121] which we do not recall here since, as we will show, we are here allowed to substitute 'compact set' to 'petite set' in the statement): there exists a petite set $C$ in $\mathbf{R}^{2d}$, a real number $b > 0$ and a real valued function $V$ such that

$$\mathsf{E}_{x,y}\left[V(X_h^h, Y_h^h)\right] - V(x,y) \leq -1 + b\ \mathbf{I}_C(x,y) \quad \text{for all } (x,y) \in \mathbf{R}^{2d}. \tag{4.9}$$

Admit for a while that the closure of every ball $B(0,R)$, $R > 0$, is petite. Then it is clear that the function $V(x,y) := C\,\Gamma(x,y)$ with $C$ large enough is a good candidate. It thus remains to prove that every compact set is petite. The forward accessibility and the fact that the law of $\Delta_{p+1}^h W$ is supported by the whole space imply that the chain $(X_{ph}^h, Y_{ph}^h)$ is a T-chain (see [11, p. 127] for the definition of T-chains, and [11, Proposition 7.1.5] for the claim). For a $\psi$-irreducible T-chain every compact set is petite (see [11, Proposition 6.2.5]). That ends the proof. □

*Remark* 4.2. The conclusions of Lemma 4.1 imply that

$$\frac{1}{N}\sum_{p=1}^{N} f(X_{ph}, Y_{ph}) \xrightarrow[N\to\infty]{} \int f(x,y)\mu^h(dx,dy) \quad \mathsf{P}-a.s., \tag{4.10}$$

and

$$\frac{1}{N}\sum_{p=1}^{N} \mathsf{E}\, f(X_{ph}, Y_{ph}) \xrightarrow[N\to\infty]{} \int f(x,y)\mu^h(dx,dy), \tag{4.11}$$

for all function $f$ with polynomial growth at infinity.

## 5. Expansion of the discretization error of the implicit Euler scheme

We aim to prove

**Theorem 5.1.** *Suppose that the sets of hypotheses* (1.1) *and* (4.1) *hold. Suppose that $f$ is a smooth function, and that all its derivatives have a polynomial growth at infinity. The discretization error of the implicit Euler scheme applied to equation* (1.1) *with any step-size small enough satisfies*

$$\int f\,d\mu - \int f\,d\mu^h = C_1 h + \cdots + C_K h^K + \mathcal{O}\left(h^{K+1}\right) \tag{5.1}$$

*for all $K$ in $\mathbf{N} - \{0\}$, where the real numbers $C_k$ are uniform in $h$.*

For the sake of simplicity we limit ourselves to $K = 1$. The proof is the same as for the explicit Euler scheme in Talay and Tubaro [22]. We write it here for the reader's convenience. We first introduce the following notation: $X \overset{\mathsf{E}}{=} Y$ stands for $\mathsf{E}\,X = \mathsf{E}\,Y$.

Set

$$Z^h_{ph} := (X^h_{(p+1)h}, Y^h_{(p+1)h})$$

A Taylor expansion and tedious computations lead to

$$u(jh, Z^h_{(p+1)h}) \overset{\mathsf{E}}{=} u(jh, Z^h_{ph}) + Lu(jh, Z^h_{ph})h + \mathcal{C}_0(jh, Z^h_{ph})h^2 + r^h_{j,p+1}h^3, \tag{5.2}$$

for all integers $j$ and $p$. The function $\mathcal{C}_0(t, y)$ is a sum of terms of the type $\phi(x, y)\partial_J u(t, x, y)$, where $\phi$ is a function with polynomial growth at infinity and $J$ a multi-index. The remainder term $r^h_{j,p+1}$ is a sum of terms of the type

$$\mathsf{E}\left[P(Z^h_{ph})\partial_J u(jh, Z^h_{ph} + \theta(Z^h_{(p+1)h} - Z^h_{ph}))\right],$$

where $P(y)$ is a a function with polynomial growth at infinity and $\theta$ is a random variable taking values in $(0, 1)$. In view of the estimates of Proposition 3.1 and of Lemma 4.1, it holds that

$$\sum_{j=0}^{+\infty} |r^h_{j,p+1}| \leq \frac{C_0}{1 - \exp(-\gamma h)}\,\mathsf{E}(1 + |Z^h_{ph}|^s + |Z^h_{(p+1)h}|^s),$$

for some integer $s$ and some strictly positive real numbers $C_0$, $\gamma$. Therefore,

$$\sum_{j=0}^{+\infty} |r^h_{j,p+1}| \leq \frac{C}{h}(1 + \mathsf{E}\,|Z^h_0|^s),$$

for some strictly positive real number $C$.

Observe that, from equation (3.9), one has

$$u((j+1)h, Z^h_{ph}) \overset{\mathsf{E}}{=} u(jh, Z^h_{ph}) + Lu(jh, Z^h_{ph})h + \frac{1}{2}L^2 u(jh, Z^h_{ph})h^2 + \tilde{r}^h_{j,p+1}h^3 \quad (5.3)$$

for some remainder $\tilde{r}^h_{j,p+1}$ of the same type as $r^h_{j,p+1}$.

Therefore, if we set

$$R^h_{j,p+1} := r^h_{j,p+1} - \tilde{r}^h_{j,p+1},$$

Equalities (5.2) and (5.3) lead to

$$u(jh, Z^h_{(p+1)h}) \overset{\mathsf{E}}{=} u((j+1)h, Z^h_{ph}) + \mathcal{C}(jh, Z^h_{ph})h^2 + R^h_{j,p+1}h^3 \quad (5.4)$$

for some function $\mathcal{C}$ of the same type as $\mathcal{C}_0$. In addition, one has

$$\sum_{j=0}^{+\infty} |R^h_{j,p+1}| \leq \frac{C}{h}(1 + \mathsf{E}\,|Z^h_0|^s) \quad (5.5)$$

for some real number $C$ independent of $h$. Observe that

$$\frac{1}{N}\sum_{p=1}^{N} f(Z^h_{ph}) = \frac{1}{N}\sum_{p=1}^{N} u(0, Z^h_{ph}) + \int_{R^2} f\, d\mu.$$

With successive uses of (5.4), one obtains

$$\begin{aligned} \frac{1}{N}\sum_{p=1}^{N} f(Z^h_{ph}) \quad \overset{\mathsf{E}}{=} \quad & \int_{R^2} f\, d\mu + \frac{1}{N}\sum_{p=1}^{N} u(ph, Z_0) + \frac{1}{N}\sum_{p=1}^{N}\sum_{j=0}^{p-1} \mathcal{C}(jh, Z^h_{ph})h^2 \\ & + \frac{1}{N}\sum_{p=1}^{N}\sum_{j=0}^{p-1} R^h_{j,p}h^3. \end{aligned} \quad (5.6)$$

We now make $N$ tend to infinity in both sides of the preceding equality. As $(X_t)$ is an ergodic process, in view of equality (3.1), one has

$$\lim_{N\to\infty} \frac{1}{N}\sum_{p=1}^{N} \mathsf{E}\, u(ph, Z_0) = 0.$$

In addition, in view of Remark 4.2, we have

$$\lim_{N\to\infty} \mathsf{E}\, f(Z^h_{ph}) = \int_{R^2} f\, d\mu^h.$$

Moreover, from the construction of $\mathcal{C}(t,y)$ it results that

$$h \lim_{N\to\infty} \frac{1}{N} \sum_{p=1}^{N} \sum_{j=0}^{p-1} \mathsf{E}\,\mathcal{C}(jh, Z_{ph}^h) = \int_0^\infty \int_{R^2} \mathcal{C}(t,x,y)\mu(dx,dy)\,dt + \mathcal{O}\left(h\right).$$

We finally use inequality (5.5). That ends the proof.

## 6. Conclusion

We have proven that the function $\mathsf{E}_{x,y}\, f(X_t,Y_t) - \int f\,d\mu$ and all its derivatives tend to 0 exponentially fast when $t$ goes to infinity. This has allowed us to get an optimal estimate on the convergence rate of the implicit Euler scheme for the approximation of $\int f\,d\mu$.

In this paper we limited ourselves to systems (1.1) with a constant diffusion matrix. Our analysis extends to systems of the type

$$\begin{cases} X_t &= X_0 + \displaystyle\int_0^t \partial_y H(X_s,Y_s)\,ds, \\ Y_t &= Y_0 - \displaystyle\int_0^t \partial_x H(X_s,Y_s)\,ds - \int_0^t F(X_s,Y_s)\;\partial_y H(X_s,Y_s)\,ds \\ & \quad + \displaystyle\int_0^t \sigma(X_s,Y_s)\,dW_s \end{cases}$$

where $\sigma$ satisfies smoothness and boundedness conditions as well as the uniform ellipticity condition

$$0 < \nu\;|\zeta|^2 \le \sum_{i,j=1}^{d} \left(\sigma(x,y)\;\sigma(x,y)^*\right)_j^i \;\zeta_i\;\zeta_j \quad \text{for all } x,y,\zeta \in \mathbf{R}^d.$$

To deal with such systems one needs to increase the complexity of several proofs of the present paper. The two main changes are: first, the smoothness and strict positivity of the transition density require Malliavin calculus techniques involving clever localization arguments in order to handle with the possible unboundness of the first derivatives of $\partial_x H(x,y)$; second, the calculations of Section 3 involve additional terms depending on the derivatives of $\sigma$; these terms make expressions such as those of Section 3 still more lengthy (and almost unreadable), and need to be controlled carefully in order to obtain estimates such as (3.17), (3.26), etc. Of course, the set of our assumptions needs to be properly modified.

Our last comment concerns the Hypothesis 1.1. It does not seem to us too stringent. However it is not necessary at all. It seems possible to construct particular functions $H$ which do not satisfy all our requirements and for which all the statements of Section 3 nevertheless hold true. This means that our methodology can be applied to various families of Hamiltonian systems.

### Acknowledgement

This paper (hopefully) answers to a question asked by C. Soize to the author.

### References

[1] D. Bakry (1994) L'hypercontractivité et son utilisation en théorie des semi-groupes. Cours 'a l'École d'été de Saint-Flour 1992. *Lect. Notes Math.* **1581**, 1–114.

[2] P. Billingsley (1968) *Convergence of Probability Measures*. Wiley, New-York.

[3] F. Campillo (1995) Optimal ergodic control of nonlinear stochastic systems. In: *Probabilistic Methods in Applied Physics*, P. Krée, W. Wedig (eds.), Lect. Notes Phys. **451**, Springer-Verlag, 239–269.

[4] S.N. Ethier and T.G. Kurtz (1986) *Markov Processes*: *Characterization and Convergence*. John Wiley & Sons.

[5] H. Ganidis, B. Roynette and F. Simonot (1999) Convergence rate of some semi-groups to their invariant probability. *Stoch. Process. Appl.* **72** (2), 243–264.

[6] R.Z. Has'minskii (1980) *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff.

[7] I. Karatzas and S.E. Shreve (1988) *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

[8] D. Lamberton and G. Pagès Recursive computation of the invariant measure of a diffusion. To appear in *Bernoulli*.

[9] F. Malrieu (2001) Inégalités de Sobolev logarithmiques pour des problèmes d'évolution non linéaires. Ph.D. Thesis, Université Paul Sabatier.

[10] J. Mattingly, A.M. Stuart and D. Higham (2001) Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. University of Strathclyde Mathematics Research Report 7. To appear in *Stoch. Proc. Appl.*.

[11] S.P. Meyn and R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag.

[12] D. Michel and E. Pardoux (1990) An introduction to Malliavin calculus and some of its applications. In: *Recent Advances in Stochastic Calculus*, Progress in Automation and Information Systems, J-S. Baras, V. Mirelli (eds.), Springer-Verlag, 65–104.

[13] D. Nualart (1998) Analysis on Wiener space and anticipating stochastic calculus. In: *École d'été de Probabilités de Saint-Flour* XXV, P. Bernard (ed.), Lect. Notes Math. **1690**, Springer-Verlag.

[14] E. Nummelin (1984) *General Irreducible Markov Chains and Non-negative Operators.* Cambridge Tracts in Mathematics, Cambridge University Press.

[15] G. Pagès (2001) Sur quelques algorithmes rcursifs pour les probabilits numriques. *ESAIM: Probability and Statistics* **5**, 141–170. http://www.edpsciences.com/articles/ps/abs/2001/01/contents/contents.html

[16] P. Protter (1990) *Stochastic Integration and Differential Equations.* Springer-Verlag, Berlin.

[17] D. Revuz (1975) *Markov Chains.* North Holland.

[18] T. Shardlow and A.M. Stuart (2000) A perturbation theory for ergodic Markov chains with application to numerical approximation. *Siam J. Numer. Anal.* **37**, 1120–1137.

[19] A.V. Skorokhod (1989) *Asymptotic Methods in the Theory of Stochastic Differential Equations.* Transl. Math. Momographs **78**, A.M.S.

[20] C. Soize (1994) *The Fokker – Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions.* Series on Advances in Mathematics for Applied Sciences **17**, World Scientific, Singapore.

[21] D. Talay (1990) Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports* **29** (1), 13–36.

[22] D. Talay and L. Tubaro (1990) Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** (4), 94–120.

[23] D. Talay (1999) Approximation of invariant measures of nonlinear Hamiltonian and dissipative stochastic differential equations. In: *Progress in Stochastic Structural Dynamics*, R. Bouc, C. Soize (eds.), Publication du L.M.A.-C.N.R.S. **152**, 139–169.

[24] A.S. Üslünel and M. Zakai (2000) *Transformation of Measure on Wiener Space.* Springer.