CONSERVATIVE STOCHASTIC DIFFERENTIAL EQUATIONS:
MATHEMATICAL AND NUMERICAL ANALYSIS

ERWAN FAOU AND TONY LELIÈVRE

Abstract. We consider stochastic differential equations on the whole Euclidean space possessing a scalar invariant along their solutions. The stochastic dynamics therefore evolves on a hypersurface of the ambient space. Using orthogonal coordinate systems, we show the existence and uniqueness of smooth solutions of the Kolmogorov equation under some ellipticity conditions over the invariant hypersurfaces. If we assume, moreover, the existence of an invariant measure, we show the exponential convergence of the solution towards its average. In the second part, we consider numerical approximation of the stochastic differential equation, and show the convergence and numerical ergodicity of a class of projected schemes. In the context of molecular dynamics, this yields numerical schemes that are ergodic with respect to the microcanonical measure over isoenergy surfaces.

1. Introduction

We consider stochastic differential equations in \( \mathbb{R}^N \) in the Stratonovich sense, of the form

\[
dX(t) = f(X(t)) dt + \sigma(X(t)) \circ dW(t),
\]

written in coordinates

\[
dX_i(t) = f_i(X(t)) dt + \sum_{\ell=1}^{D} \sigma_{i\ell}(X(t)) \circ dW^{[\ell]}(t),
\]

where \( f(x) = (f^i(x))_{i=1}^N \) is an \( N \)-dimensional vector vector field, and \( \sigma(x) = (\sigma_{i\ell}(x)), \ i = 1, \ldots, N, \ \ell = 1, \ldots, D \) is a \( N \times D \) matrix. The vector \( W(t) \) with components \( W^{[\ell]}, \ \ell = 1, \ldots, D, \) is a standard Brownian motion in \( \mathbb{R}^D \). This equation corresponds to the Itô equation: For \( i = 1, \ldots, N, \)

\[
dX^i(t) = \left( f^i(X(t)) + \sum_{j=1}^{N} \sum_{\ell=1}^{D} \frac{1}{2} (\sigma_{i\ell}(t) \partial_j \sigma_{i\ell}(t))(X(t)) \right) dt + \sum_{\ell=1}^{D} \sigma_{i\ell}(X(t)) dW^{[\ell]}(t)
\]

where \( \partial_j \sigma_{i\ell} \) denotes the derivative with respect to \( x^j \) of the function \( \sigma_{i\ell}(t, x) \), \( x = (x^1, \ldots, x^N) \).
We assume in this work that the SDE (1.1) is \textit{conservative} in the following sense: there exists a function $H(x)$ (called the energy in the following) defined on $\mathbb{R}^N$ such that

\begin{equation}
\forall x \in \mathbb{R}^N, \forall \ell = 1, \ldots, D, \quad \langle \nabla H(x), f(x) \rangle = 0 \quad \text{and} \quad \langle \nabla H(x), \sigma_\ell(x) \rangle = 0
\end{equation}

where for all $\ell = 1, \ldots, D$, $\sigma_\ell(x)$ is the $N$-dimensional vector with components $\sigma_\ell^i(x)$, and where $(\cdot, \cdot)$ denotes the Euclidean scalar product on $\mathbb{R}^N$. The vector field $\nabla H(x)$ is the gradient vector of $H$ with respect to the coordinates $x = (x^1, \ldots, x^N)$.

Using Itô calculus, it is straightforward to show that if $H(x)$ is smooth enough on $\mathbb{R}^N$, we have for every solution $X(t)$ of (1.1)

\begin{equation}
\forall t \geq 0, \quad H(X(t)) = H(X(0)).
\end{equation}

This expresses that the function $H$ is an invariant of (1.1).

Under ellipticity assumptions on the vector fields $\sigma_\ell$, we show below that the SDE (1.1) exhibits a unique invariant measure over the manifolds

\begin{equation}
\Sigma_z = \{ x \in \mathbb{R}^N | H(x) = z \}.
\end{equation}

In the context of molecular dynamics, if the equation (1.1) is \textit{volume preserving} over the whole space $\mathbb{R}^N$, then the invariant measure turns out to be the \textit{microcanonical} probability measure (see for instance [11]),

\begin{equation}
d\nu(z) = \left( \int_{\Sigma_z} \frac{d\Sigma_z}{\|\nabla H\|} \right)^{-1} \frac{d\Sigma_z}{\|\nabla H\|}
\end{equation}

where $d\Sigma_z$ is the surface measure over the isosurface $\Sigma_z$ induced by the Euclidean metric of $\mathbb{R}^N$.

For a general function $H(x)$, it is very difficult in practice to exhibit an atlas of the manifolds $\Sigma_z = \{ x \in \mathbb{R}^N | H(x) = z \}$. The derivation of a numerical scheme to approximate the solution of (1.1) cannot thus rely on the choice of a “good” atlas as in [13], and one typically needs to use a projection step to remain on the manifold.

The numerical analysis then has to be made directly in $\mathbb{R}^N$ and the dynamics of (1.1) has to be understood in a vicinity of a given hypersurface $\Sigma_z$ rather than on the manifold itself.

The Kolmogorov equation associated with the SDE (1.1) shows, using the summation convention of covariant and contravariant indices, that

\begin{equation}
\partial_t v = Lv \quad \text{where} \quad Lv = f^i \partial_i v + \frac{1}{2} \sum_{\ell=1}^D \sigma_\ell^i \partial_i (\sigma_\ell^j \partial_j v),
\end{equation}

where $v(t, x)$ is a function depending on the time $t \geq 0$ and $x \in \mathbb{R}^N$. We recall that a solution of (1.7) is given by $v(t, x) = \mathbb{E}(v(0, X(t)) | X(0) = x)$.

After a possible translation of the function $H(x)$, we can always assume that the initial value of (1.1) lies on the manifold $\Sigma_0 = \{ x \in \mathbb{R}^N | H(x) = 0 \}$. In this work, we will show that we can define smooth solutions of (1.7) in a neighborhood of this manifold by considering an orthogonal parametrization of $\mathbb{R}^N$ by the hypersurface $\Sigma_0$ which is supposed to be an $N - 1$-dimensional submanifold of $\mathbb{R}^N$, and the parameter $z = H(x)$ for $x \in \mathbb{R}^N$. In such a way, we can parametrize a domain of $\mathbb{R}^N$ by a product $\Sigma_0 \times (-\varepsilon, \varepsilon)$ for small $\varepsilon$, provided that $\Sigma_0$ is smooth enough and compact (note that this last hypothesis could be weakened in the present analysis,
but it is in general fulfilled for the problem of computing averages in molecular
dynamics).

In this new coordinate system, the Kolmogorov operator can be seen as an
intrinsic operator on $\Sigma_0$ involving the covariant derivative and curvature terms
on $\Sigma_0$. The coordinate $z$ is then only a smooth parameter in the operator. In
particular, the derivative with respect to the global variable $z$ does not come into
play in the definition of the Kolmogorov operator in orthogonal coordinates. By
standard arguments, we can therefore define a smooth solution in the neighborhood
defined by $\Sigma_0 \times (-\varepsilon, \varepsilon)$, provided that the collection $\sigma_\ell(x)$ satisfies conditions so
that the operator $L$ can be viewed as an elliptic operator on $\Sigma_0$, uniformly in $z$.

We then consider the case where (1.1) possesses a family of invariant measures
on $\Sigma_z$, depending smoothly on $z$. An important case where this situation occurs is
when (1.1) is volume preserving, which means that the constants are in the kernel
of the adjoint of the operator (1.7) which is written as

\begin{equation}
L^T v = -\partial_i (f^i v) + \frac{1}{2} \sum_{\ell=1}^{D} \partial_i (\sigma^i_\ell \partial_j (\sigma^j_\ell v)).
\end{equation}

We will refer to this operator as the Fokker-Planck operator associated with (1.1).
Note that (1.1) will be volume preserving in the case where

\begin{equation}
\partial_i f^i(x) = 0 \quad \text{and} \quad \forall \ell = 1, \ldots, D, \quad \partial_i \sigma^i_\ell(x) = 0,
\end{equation}

but this condition is not necessary (see the examples in Section 4).

The motivation for considering such volume and energy preserving systems arises
from computational problems of NVE averages in molecular dynamics (namely,
averages with respect to the microcanonical measure (1.6), see for instance [1, 11]).
In this situation, for $d$ interacting particles, the phase space variable decomposes
into $x = (p_1, \ldots, p_d, q_1, \ldots, q_d)$ where each $q_i \in \mathbb{R}^3$ denotes the position of a particle,
and $p_i$ its momentum. The energy Hamiltonian $H(p, q)$ is then written as

\begin{equation}
H(p, q) = \sum_{i=1}^{d} \frac{p_i^T p_i}{2m_i} + U(q_1, \ldots, q_d)
\end{equation}

where $m_i$ is the mass of the $i$-th particle and $U$ is the potential function.

The principle of molecular dynamics is to simulate numerically the solution of the corresponding hamiltonian equations

\begin{equation}
\frac{dq}{dt} = \frac{\partial H}{\partial p}(p, q) \quad \text{and} \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}(p, q),
\end{equation}

which define a volume and energy preserving (deterministic) system. The ergodic hypothesis states that, if $(p(t), q(t))$ denotes a solution of (1.11) evolving on a hy-

\begin{equation}
limit_{T \to \infty} \frac{1}{T} \int_0^T g(p(t), q(t)) \, dt = \int_{\Sigma_z} g(y) \, d\nu(z)
\end{equation}

where the measure $d\nu(z)$ in the right-hand side is the microcanonical measure (1.6).

However, it is well known that this hypothesis failed in general; in particular,
for integrable or close to integrable systems exhibiting hidden stable invariants
preserved by symplectic numerical methods (see [14, 5] and the references therein).
At variance the numerical schemes we present below are shown to be ergodic with respect to the microcanonical measure.

Another context where such families of measures appear is free energy computations by thermodynamic integration [7]. In this case, one needs to sample a Boltzmann-Gibbs probability measure conditioned to a fixed value of a reaction coordinate:

$$\exp(-\beta V) |\nabla \xi|^{-1} dS_z$$

where $V : \mathbb{R}^N \to \mathbb{R}$ is the energy, $\xi : \mathbb{R}^N \to \mathbb{R}$ is the reaction coordinate and $S_z = \{x \in \mathbb{R}^N | \xi(x) = z\}$ (here $\xi$ plays the role of the above function $H$).

In the first part of this paper, under an ellipticity assumption on the vector fields $\sigma[\ell]$ and in the case where $(1.1)$ admits an invariant measure, we show the exponential convergence of the solution of $(1.7)$ towards its average over the microcanonical measure, uniformly in $z$ for $z$ in a small interval. Moreover, we show that all the tangent space derivatives of $v(t,x)$ decay exponentially in time. This implies, in particular, the ergodicity of the exact solution of $(1.1)$ and gives the equivalent of $(1.12)$ for the stochastic flow of $(1.1)$ in the volume preserving case.

In the second part of this work, we consider numerical schemes satisfying in essence two conditions: consistency with $(1.1)$ and preservation of the energy $H(x)$. Using the results obtained in Section 2, we show the weak convergence of such numerical schemes, which gives a new way of computing NVE averages in the context of molecular dynamics. We conclude by giving examples of schemes satisfying the above conditions.

The method and analysis we use are closely related to previous works by Talay et al. [22, 13, 20, 21]. The particularity here lies in a systematic use of differential geometry to understand the properties of the Kolmogorov operator on a hypersurface of the ambient space. Let us also mention the analysis made in [17, 16] where constrained symplectic SDEs and appropriate discretization schemes are introduced.

In the following, for a given $\varepsilon > 0$, we define the domain

$$\Omega^\varepsilon = \{x \in \mathbb{R}^N | H(x) < \varepsilon\}.$$

It defines a neighborhood of the hypersurface $\Sigma_0$.

2. Analysis of the Kolmogorov Operators

The goal of this section is to derive expressions of the operators $L$ and $L^T$ in terms of intrinsic objects defined on the hypersurfaces $\Sigma_z$.

2.1. Orthogonal coordinate system. We can write the Kolmogorov and Fokker-Planck operators as

$$L v = f^i \nabla_i v + \sum_{\ell=1}^D \frac{1}{2} \sigma_{[\ell]}^i \nabla_i (\sigma_{[\ell]}^j \nabla_j v),$$

$$L^T v = -\nabla_i (f^i v) + \sum_{\ell=1}^D \frac{1}{2} \nabla_i (\sigma_{[\ell]}^i \nabla_j (\sigma_{[\ell]}^j v)),$$

respectively, where $\nabla$ is the Riemannian connection of $\mathbb{R}^N$ associated with the Euclidean product $\langle \cdot, \cdot \rangle$. The advantage of these expressions is that they are expressed only in terms of globally defined objects: the covariant derivative and the metric tensor. In the Euclidean coordinate system $\{x^1, \ldots, x^N\}$, the metric tensor is the identity tensor, and this expression reduces to (1.7). For a general
local coordinate system \(\{y^1, \ldots, y^N\}\), the metric is represented by the matrix \(g_{ij}(y)\), \(i, j = 1, \ldots, N\) which is the component of the metric tensor in the coordinate system \(\{y^i\}\). With this metric tensor are associated the Christoffel symbols (see for instance \[4\])

\[
(2.2) \quad \Gamma^k_{ij} = \frac{1}{2} g^{km} (\partial_i g_{jm} + \partial_j g_{im} - \partial_m g_{ij}),
\]

where \(\partial_i\) denotes the derivative with respect to \(y^i\), where \(g^{km}\) are the components of the inverse of the metric tensor defined by \(g^{km}g_{ml} = \delta^k_l\), the Kronecker tensor.

The coordinate system \(\{y^i\}\) induces a local basis \(\frac{\partial}{\partial y^i}\) in the corresponding tangent bundle. For a vector field \(T = T^i \frac{\partial}{\partial y^i}\), the covariant derivative is the \((1, 1)\) tensor fields with components

\[
\nabla_i T^j = \partial_i T^j + \Gamma^j_{im} T^m.
\]

Now an expression like \(\nabla_i T^i\) is the contraction of the previous tensor, and is thus a function equal to

\[
\nabla_i T^i = \partial_i T^i + \Gamma^i_{im} T^m.
\]

It is worth noting that the right-hand side in this expression \textit{a priori} depends on the local coordinate system \(\{y^i\}\) while the left-hand side denotes a function defined globally on \(\mathbb{R}^N\). For a function \(v\), the covariant derivative reduces to \(\nabla_i v = \partial_i v\).

In a local coordinate system, we thus see that (2.1) can be written as

\[
Lv = f^i \partial_i v + \sum_{\ell=1}^{D} \frac{1}{2} \sigma^i_{[\ell]} \partial_i (\sigma^j_{[\ell]} \partial_j v),
\]

and

\[
(2.3) \quad L_T v = -\partial_i (f^i v) - \Gamma^i_{im} f^m v + \frac{1}{2} \sum_{\ell=1}^{D} \partial_i (\sigma^i_{[\ell]} \partial_j (\sigma^j_{[\ell]} v)) + \frac{1}{2} \sum_{\ell=1}^{D} \partial_i (\sigma^i_{[\ell]} \Gamma^j_{km} \sigma^k_{[\ell]} v)
\]

\[
+ \frac{1}{2} \sum_{\ell=1}^{D} \Gamma^i_{im} (\sigma^i_{[\ell]} \partial_j (\sigma^j_{[\ell]} v)) + \frac{1}{2} \sum_{\ell=1}^{D} \Gamma^i_{im} (\sigma^i_{[\ell]} \Gamma^j_{km} \sigma^k_{[\ell]} v).
\]

(Recall that for Euclidean coordinates, all the Christoffel symbols vanish).

We introduce now an orthogonal coordinate system that will reflect the foliation of \(\Omega^c\) by the isosurfaces \(\Sigma_z\) defined by \([15]\).

Next, we make the following assumption on \(H\):  

\textbf{Hypothesis 2.1.} The function \(H\) is smooth, and there exists \(\varepsilon > 0\) such that for all \(z \in (-\varepsilon, \varepsilon)\), \(\Sigma_z\) is compact, and for all \(x \in \Sigma_z\), \(\nabla H(x) \neq 0\). This implies that for \(z \in (-\varepsilon, \varepsilon)\), \(\Sigma_z\) is a smooth submanifold of \(\mathbb{R}^N\) of dimension \(N - 1\).

Note that \(\Omega^c := \bigcup_{z \in (-\varepsilon, \varepsilon)} \Sigma_z\). We then denote by \(F_s(x)\) the flow of the ordinary differential equation

\[
(2.4) \quad \frac{d}{ds} F_s(x) = \frac{\nabla H(F_s(x))}{\|\nabla H(F_s(x))\|^2}, \quad F_0(x) = x \in \mathbb{R}^N,
\]

where \(\|\cdot\|\) denotes the Euclidean norm in \(\mathbb{R}^N\). Up to a possible reduction of \(\varepsilon\), Hypothesis \([2,3]\) implies that for all \(x \in \Omega^c\), \(F_s(x)\) is well defined for \(|s| < \varepsilon\).
Let \((U, \varphi)\) be a local chart on \(\Sigma_0\) with \(U \subset \Sigma_0\). The application \(\varphi : \Sigma_0 \supset U \to \varphi(U) \subset \mathbb{R}^{N-1}\) induces local coordinates \((y^\alpha)_{\alpha=1}^{N-1} \in \varphi(U) \subset \mathbb{R}^{N-1}\). In the following, Greek indices will refer to \(N - 1\)-dimensional indices running from 1 to \(N - 1\). We define the application
\[
(2.5) \quad \varphi(U) \times (-\varepsilon, \varepsilon) \ni (y^\alpha, z) \mapsto \Phi(y^\alpha, z) := F_z \circ \varphi^{-1}(y^\alpha).
\]
This is a local coordinate system over the domain \(\Omega^f\). By construction, we have
\[
\partial_z[H(\Phi(y^\alpha, z))] = \langle \nabla H(\Phi(y^\alpha, z)), \partial_z \Phi(y^\alpha, z) \rangle
= \langle \nabla H(\Phi(y^\alpha, z)), \nabla H(\Phi(y^\alpha, z)) \rangle / \|\nabla H(\Phi(y^\alpha, z))\|^2
= 1,
\]
whence for all \(z \in (-\varepsilon, \varepsilon)\) and all \(y^\alpha \in U\),
\[
(2.6) \quad H(\Phi(y^\alpha, z)) = z.
\]
Hence, the inverse application of (2.5) is the map
\[
(2.7) \quad \Omega^f \ni x \mapsto (\varphi(F_{-H(x)}(x)), H(x)) \in \varphi(U) \times (-\varepsilon, \varepsilon)
\]
and it is clear that an atlas on \(\Sigma_0\) corresponds to an atlas on \(\Omega^f\) made up of the charts in (2.7).

For given \((y^\alpha, z) \in \varphi(U) \times (-\varepsilon, \varepsilon)\), we denote for simplicity by \((Y_\alpha, Z_z)\) the local basis of \(\mathbb{R}^N\) corresponding to the local basis \((\partial_{y^\alpha}, \partial_z)\) and we identify the vector fields \(Y_i, i = a, z\) with the vectors of \(\mathbb{R}^N\), \(Y_\alpha = \partial_\alpha \Phi\) and \(Z_z = \partial_z \Phi\).

The metric tensor \(g_{ij}(y^\alpha, z)\) splits into the “surfacic” components \(g_{\alpha \beta}\) and the “normal” components \(g_{zz}\) and \(g_{z\alpha}\). By definition we have
\[
(2.8) \quad g_{zz}(y^\alpha, z) = \langle Y_z(y^\alpha, z), Y_z(y^\alpha, z) \rangle
= \langle \partial_z \Phi(y^\alpha, z), \partial_z \Phi(y^\alpha, z) \rangle
= 1 / \|\nabla \Phi(y^\alpha, z)\|^2.
\]
From (2.6), we deduce that for all \(\beta\),
\[
\langle \nabla \Phi(y^\alpha, z), \partial_\beta \Phi(y^\alpha, z) \rangle = 0.
\]
Dividing this expression by \(\|\nabla \Phi(y^\alpha, z)\|^2\), we obtain
\[
\langle \partial_z \Phi(y^\alpha, z), \partial_\beta \Phi(y^\alpha, z) \rangle = 0,
\]
which means that
\[
g_{z\beta} = \langle Y_\beta, Z_z \rangle = 0.
\]
This expression justifies the name “orthogonal coordinate system” on \(\Omega^f\).

Finally, we see that
\[
g_{\alpha \beta}(y^\alpha, z) = \langle Y_\alpha, Y_\beta \rangle = \langle \partial_\alpha \Phi, \partial_\beta \Phi \rangle
\]
expands as
\[
(2.9) \quad g_{\alpha \beta}(y^\alpha, z) = a_{\alpha \beta} + \sum_{k \geq 1} \frac{z^k}{k!} \partial_z^k g_{\alpha \beta}(y^\alpha, 0)
\]
where \(a_{\alpha \beta} = \langle \partial_\alpha \varphi^{-1}, \partial_\beta \varphi^{-1} \rangle\) is the metric tensor on \(\Sigma_0\) (the expansion in \(z\) comes from the analyticity in \(z\) of the solution \(F_z(x)\) of (2.4)).

Using this orthogonal coordinate system, it is easy to show that each tensor field on \(\Omega^f\) can be decomposed into several tensor fields on \(\Sigma_0\) depending smoothly
on \( z \in (\varepsilon, \varepsilon) \). For instance, if \( u^i \) is a vector field on \( \Omega^\varepsilon \), it can be represented by a couple \((u^\alpha(y, z), u^\varepsilon(y, z)) \in C^\infty((\varepsilon, \varepsilon), \Gamma(T^1\Sigma_0) \times C^\infty(\Sigma_0))\) where \( \Gamma(T^1\Sigma_0) \) denotes the space of vector fields on \( U \) and where \( y = (y^\alpha) \) denotes a point in \( \Sigma_0 \). The proof of this result relies on the special choice of the atlas defined by \((2.7)\) and this situation is very similar to the case of shells (see [9, 8]).

In the orthogonal coordinate system defined above, the local charts \((2.7)\) define local coordinates on open subsets of \( \Omega^\varepsilon \). We do not give a proof here, as it is very similar to the situation of shells (see [9, 8] for similar statements). Roughly speaking, it states that freezing one or two coefficients of the Christoffel symbol to \( \varepsilon \) for similar statements). Roughly speaking, it states that freezing one or two indices to \( \varepsilon \) in orthogonal coordinates, we can easily show the following result:

**Lemma 2.2.** Let \((y, z) = (y^\alpha, z)\) be a local orthogonal coordinate system induced by a local chart \((U, \varphi)\) on \( \Sigma_0 \) and let \( \Gamma^k_{ij} \) be the Christoffel symbols \((2.2)\). Then the expressions

\[
\begin{align*}
\Gamma^3_{\alpha\alpha}(y, z)Y_0(0) & \otimes dy^\alpha(y, 0), \\
\Gamma^z_{\alpha\beta}(y, z)dy^\beta(y, 0) & \otimes dy^\alpha(y, 0), \\
\Gamma^z_{\alpha\varepsilon}(y, z)dy^\alpha(y, 0), \\
\Gamma^z_{\varepsilon\varepsilon}(y, z)Y_0(0),
\end{align*}
\]

all define tensor fields on \( \Sigma_0 \) depending smoothly on \( z \).

We do not give a proof here, as it is very similar to the situation of shells (see [9, 8] for similar statements). Roughly speaking, it states that freezing one or two coefficients of the Christoffel symbol to \( z \) in orthogonal coordinates yields tensor fields.

Let \( u^i \) be a smooth vector field on \( \Omega^\varepsilon \). The covariant derivative of \( u^i \) in a local coordinate system is given by

\[
\nabla_i u^j = \partial_i u^j + \Gamma^j_{ik} u^k
\]

and defines a tensor field on \( \Omega^\varepsilon \).

In a local orthogonal basis, this tensor field can be decomposed into four parts by freezing two, one or zero indices to \( z \): the function \( \nabla_z u^z \), the vector field \( \nabla_z u^\alpha \), the one-form field \( \nabla_\alpha u^z \) and the order 2 tensor field \( \nabla_\alpha u^\beta \). More precisely, we have

\[
\nabla_z u^z = \partial_z u^z + \Gamma^z_{\varepsilon\varepsilon}(y, z)u^z
\]

defining a function in \( C^\infty((\varepsilon, \varepsilon), C^\infty(\Sigma_0)) \), and similarly

\[
\nabla_z u^\alpha = \partial_z u^\alpha + \Gamma^\alpha_{z\alpha}(y, z)u^\varepsilon + \Gamma^\alpha_{\alpha\varepsilon}(y, z)u^z,
\]

which defines an element of \( C^\infty((\varepsilon, \varepsilon), \Gamma(T^1\Sigma_0)) \) after using the previous lemma.

In the following, we denote by \( D_\alpha \) the covariant derivative on \( \Sigma_0 \), associated with the metric \( a_{\alpha\beta} \). As \( \partial_\alpha u^z = D_\alpha u^z \) because \( u^z \) is a scalar function, we obtain that the expression

\[
\nabla_\alpha u^z = D_\alpha u^z + \Gamma^z_{\varepsilon\varepsilon}(y, z)u^\varepsilon + \Gamma^z_{\alpha\varepsilon}(y, z)u^z
\]
defines an element of $C^\infty((-\varepsilon, \varepsilon), \Gamma(T_1\Sigma_0))$, where $\Gamma(T_1\Sigma_0)$ denotes the space of \allowbreak 1-form fields on $\Sigma_0$.

Finally, we write
\begin{equation}
(2.13) \quad \nabla_\alpha u^\beta = \partial_\alpha u^\beta + \Gamma^\beta_{\alpha\sigma}(y, z)u^\sigma + \Gamma^\beta_{\alpha z}(y, z) u^z.
\end{equation}

But using the expansion (2.9) it can be shown (see [9]) that the expression
\[ \Lambda^\alpha_{\alpha\beta}(y, z) = \Gamma^\alpha_{\alpha\beta}(y, z) - \Gamma^\alpha_{\alpha\beta}(y, 0) \]
defines a tensor field on $\Sigma_0$. Note that $\Gamma^\alpha_{\alpha\beta}(y, 0)$ are the Christoffel symbols on $\Sigma_0$ associated with the metric $a_{\alpha\beta}$. Consequently, we can write (2.13) as
\begin{equation}
(2.14) \quad \nabla_\alpha u^\beta = D_\alpha u^\beta + \Lambda^\beta_{\alpha\sigma}(y, z)u^\sigma + \Gamma^\beta_{\alpha z}(y, z) u^z,
\end{equation}
which now makes sense in $C^\infty((-\varepsilon, \varepsilon), \Gamma(T_1\Sigma_0))$ where $\Gamma(T_1\Sigma_0)$ denotes the space of tensor fields of type $(1, 1)$ on $\Sigma_0$. Note that $\Lambda^\alpha_{\alpha\beta}(y, 0) = 0$ and that $\Lambda^\alpha_{\alpha\beta}(y, z)$ is analytic in $z$.

2.2. Existence of solutions. We express now the operators $L$ and $L^T$ in orthogonal coordinate system. Let $f^a$ be a vector field on $\Omega^z$ such that $(\nabla H, f) = 0$. Written in an orthogonal coordinate system, this implies that $f^z = 0$. In the following, we write $f^\beta(y, z)$ the tangential part of the vector field $f$ viewed as a vector field on the manifold $\Sigma_0$ depending smoothly on $z$. We hence have
\[ f^i \nabla_i v = f^a(y, z) D_\alpha v(y, z). \]

Using the equations (2.10) and (2.14), we find
\begin{equation}
(2.15) \quad \nabla_j (f^j v) = \nabla_\alpha (f^a v) + \nabla_z (f^z v) = D_\alpha (f^a v) + \Lambda^\alpha_{\alpha\sigma} (f^\sigma v)
\end{equation}
where all the tensors are evaluated at the point $(y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon)$. Similarly, we can write
\[ \nabla_i (f^j \nabla_j (f^i v)) = D_\alpha (f^a \nabla_j (f^j v)) + \Lambda^\alpha_{\alpha\sigma} (f^\sigma \nabla_j (f^i v)) \]
whence using (2.15)
\[ \nabla_i (f^i \nabla_j (f^j v)) = D_\alpha (f^a D_\beta (f^\beta v)) + D_\alpha (f^a \Lambda^\beta_{\alpha\sigma} (f^\sigma v)) + \Lambda^\alpha_{\alpha\sigma} (f^\sigma D_\beta (f^\beta v)) + \Lambda^\alpha_{\alpha\sigma} (f^\sigma \Lambda^\beta_{\beta\sigma} (f^\sigma v)). \]

In the following, we denote by $D$ the intrinsic operator with components $D_\alpha$ in a local coordinate system. Using the previous expression, we can easily show the following result:

**Theorem 2.3.** In orthogonal coordinates, the Kolmogorov operator $L$ viewed as an operator on the manifold $\Sigma_0 \times (-\varepsilon, \varepsilon)$ is written as
\[ L(y, z; D)v = f^a(y, z) D_\alpha v + \frac{1}{2} \sum_{\ell=1}^{D} \sigma^\alpha_{[\ell]}(y, z) D_\alpha (\sigma^\beta_{[\ell]}(y, z) D_\beta v) \]
where $\sigma^\alpha_{[\ell]}(y, z)$ and $f^\alpha(y, z)$ are the components of the vector fields $\sigma_{[\ell]}$ and $f$ in orthogonal coordinates. Similarly, the Fokker-Planck operator $L^T$, shows in orthogonal coordinates that
\[ L^T(y, z; D)v = \frac{1}{2} \sum_{\ell=1}^{D} D_\alpha (\sigma^\alpha_{[\ell]}(y, z) D_\beta (\sigma^\beta_{[\ell]}(y, z) v)) - D_\alpha (f^\alpha(y, z) v) + P(y, z; D)v, \]
$P(y, z; D)$ is an intrinsic operator of order 1 on $\Sigma_0$ depending analytically on $z$ and such that

$$P(y, 0; D) = 0.$$ 

In the rest of this work, we make the following hypothesis:

**Hypothesis 2.4.** For all one-form fields $\xi_\alpha \in \Gamma(T_1 \Sigma_0)$ and all $(y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon)$,

$$\sum_{\ell=1}^{D} \sigma_\ell^\alpha(y, z) \sigma_\ell^\beta(y, z) \xi_\alpha \xi_\beta \geq \Lambda \xi_\alpha \xi_\alpha$$

for some $\Lambda$ independent on $z$.

This means that $L(y, z; D)$ is elliptic on $\Sigma_0$, uniformly in $z \in (-\varepsilon, \varepsilon)$.

For a function $\varphi$ on $\Sigma_0$, and $\ell \geq 0$, we denote by

$$|D^\ell \varphi|^2 = D^{\alpha_1}D^{\alpha_2} \cdots D^{\alpha_{\ell}} \varphi D_{\alpha_1}D_{\alpha_2} \cdots D_{\alpha_{\ell}} \varphi$$

the semi-norm of order $\ell$ of $\varphi$. Note that the contravariant indices denote a multiplication by the inverse of the metric tensor $a^{\alpha \beta}$ on $\Sigma_0$: $D^\alpha = a^{\alpha \beta} D_\beta$.

We recall that for $k \geq 0$ the Sobolev space $H^k(\Sigma_0)$ on $\Sigma_0$ is defined as the space of functions $\varphi : \Sigma_0 \to \mathbb{R}$ such that for all $\ell$, $0 \leq \ell \leq k$,

$$\|D^\ell \varphi\|_{L^2} = \int_{\Sigma_0} |D^\ell \varphi| d\Sigma_0 < \infty.$$ 

The corresponding norm is written as

$$\|\varphi\|_{H^k} = \sum_{\ell=0}^{k} \|D^\ell \varphi\|_{L^2}.$$ 

On the manifold $\Sigma_0$, the covariant derivative operator commutes with the contraction by the metric tensor: $D^\alpha \varphi = a^{\alpha \beta} D_\beta \varphi = D_\beta a^{\alpha \beta} \varphi$. However, the covariant derivative does not commute with itself. For a given one-form field $u_\alpha$ on $\Sigma_0$, we have, for instance,

$$D_\alpha D_\beta u_\nu - D_\beta D_\alpha u_\nu = R_{\alpha \beta \delta \nu} u^\delta$$

where $u^\delta = a^{\delta \sigma} u_\sigma$ is the contravariant vector field corresponding to $u_\alpha$, and where $R_{\alpha \beta \delta \nu}$ is the curvature tensor on $\Sigma_0$ depending only on the metric tensor $a_{\alpha \beta}$ (see for instance [4]).

**Theorem 2.5.** Let $v_0 \in C^\infty(\Sigma_0 \times (-\varepsilon, \varepsilon))$. Under Hypothesis 2.4, the equation

$$\partial_t v(t, y, z) = L(y, z; D)v(t, y, z), \quad v(0) = v_0$$

possesses a unique solution $v(t, y, z) \in C^\infty((0, +\infty) \times \Sigma_0 \times (-\varepsilon, \varepsilon))$.

**Proof.** The result is a consequence of a priori bounds for the solution of (2.18). Let $(\cdot, \cdot)_{L^2}$ denote the $L^2$ product on $\Sigma_0$. If $v$ is a solution, we can write

$$\frac{1}{2} \partial_t \|v\|_{L^2}^2 = (L(y, z; D)v, v)_{L^2}$$

$$\leq -\frac{1}{2} \sum_{\ell=1}^{D} \sigma_\ell^\alpha(y, z) \sigma_\ell^\beta(y, z) D_\alpha v D_\beta v + (Q(y, z; D)v, v)_{L^2}$$

where $Q(y, z; D)$ is an operator of order 1 on $\Sigma_0$ depending smoothly on $z$. This shows that for constants $C$, $c$ and $b$ depending on $\varepsilon$, we have

$$\frac{1}{2} \partial_t \|v\|_{L^2}^2 \leq -C\|Dv\|_{L^2}^2 + c\|Dv\|_{L^2} \|v\|_{L^2} + b\|v\|_{L^2}^2,$$
whence
\[ \partial_t \|v\|_{L^2}^2 \leq M \|v\|_{L^2}^2 , \]
for a constant \( M \) depending on \( \varepsilon \). This shows the a priori estimate
\[ \|v\|_{L^2} \leq e^{tM} \|v_0\|_{L^2} . \]
Using (2.17), we obtain similar estimates for the derivatives \( \partial_x^k \mathbf{D}^\ell v \) which satisfy equations of the form
\[ \partial_t \partial_x^k \mathbf{D}^\ell v = L(y, z; \mathbf{D}) \partial_x^k \mathbf{D}^\ell v + R(y, z; \mathbf{D}, \partial_x) v \]
where \( R \) is an operator of order \( k + \ell - 1 \) in \( \partial_x \) and \( D_\alpha \). We hence easily obtain a priori estimates for these derivatives by induction. The uniqueness is clear, as the equation (2.18) is linear. \( \Box \)

2.3. **Invariant measure and exponential convergence.** Throughout the rest of this section, we assume, moreover, that there exists a smooth function \( \rho(x) > 0 \) defined on \( \Omega^+ \) such that for all \( x \in \Omega^+ \), we have
\[ (2.19) \quad L^T \rho = 0. \]
Note that in the case where \( \rho \equiv 1 \) the constant function is equal to 1 on \( \Omega^+ \), then (2.19) expresses that the equation (1.1) preserves the volume.

For a given function \( \rho > 0 \), many systems can be constructed to guarantee condition (2.19). For example, as in (1.9), this will be the case when
\[ \partial_t (\rho f^i) = 0 \quad \text{and} \quad \forall \ell = 1, \ldots, D, \quad \partial_t (\rho \sigma^i_{\ell |q}) = 0. \]
In Section 4.1 we give examples of volume preserving systems satisfying (1.9) (stochastic shakers) from which we easily construct systems satisfying the previous equation for the arbitrary function \( \rho > 0 \). Other examples of volume preserving dynamics are given by the projected gradient dynamics detailed in Section 4.1.

The hypothesis (2.19) states that \( \rho \) is an invariant measure for (1.1). Indeed, for \( x \in \Omega^+ \), let \( X(t, x) \) denote the solution of (1.1) starting at \( x \) for \( t = 0 \). Using (1.4), we have \( X(t, x) \in \Omega^+ \) for all \( t \geq 0 \). Then for functions \( \varphi(x) \) with compact support in \( \Omega^+ \), using the Kolmogorov equation we have
\[ \frac{d}{dt} \int_{\Omega^+} \mathbb{E}(\varphi(X(t, x)) \rho(x)) \, dx = \langle L(x^1; \partial_t) \mathbb{E}(\varphi(X(t, x)) \rho(x))_{L^2(\Omega^+)} \rangle = \langle \mathbb{E}(\varphi(X(t, x)) \rho(x))_{L^2(\Omega^+)} \rangle = 0 \]
where \( \langle \cdot, \cdot \rangle_{L^2(\Omega^+)} \) denotes the \( L^2 \) product in Cartesian coordinates on \( \Omega^+ \), and where \( L^T(x^1; \partial_t) \) denotes the adjoint of the operator \( L(x^1; \partial_t) \) given by (1.7) in Cartesian coordinates. In a normal orthogonal coordinate system \( (y^\alpha, z) \), the volume form \( dx \) is transformed into the measure \( \sqrt{|g_{ij}(y, z)|} \, dy^\alpha \wedge dz \) where \( |g_{ij}(z)| \) denotes the determinant of the metric tensor \( g_{ij}(y, z) \). Using the equations (2.8) and (2.9), this measure can be written as
\[ (2.20) \quad \frac{1}{\|\nabla H\|} \, d\Sigma_z \wedge dz \]
where \( d\Sigma_z \) is the volume form \( \sqrt{|g_{ij}(y, z)|} \, dy^\alpha \) induced by the Euclidean measure in \( \mathbb{R}^N \) on the surface \( \Sigma_z \). Hence, we can write for all \( x = (y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon) \),
\[ \rho(x) \, dx = \mu(y, z) \, d\Sigma_0 \wedge dz \]
where
\begin{equation}
\mu(y, z) = \frac{\rho(y, z)}{\|\nabla H(y, z)\|} \sqrt{\frac{|g_{\alpha \beta}(y, z)|}{|g_{\alpha \beta}(y, 0)|}}
\end{equation}
defines a density function on \(\Sigma_0\) depending smoothly on \(z\). Later, we will often ignore the dependency in \(y \in \Sigma_0\), and write \(\mu(z)\) for \(\mu(y, z)\). Shortly, we can write
\[\rho(x) \, dx = d\mu(z) \wedge dz\]
where \(d\mu(z) = \mu(z) d\Sigma_0\). Note that in the case where \(\rho \equiv 1\), then up to a multiplication by a constant, the measure \(d\mu(z)\) is the microcanonical measure on \(\Sigma_z\) (see \(\text{[1.12]}\)).

**Proposition 2.6.** For all \((y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon)\), let \(L(y, z; D)\) denote the adjoint of \(L(y, z; D)\) with respect to the \(L^2\) product on the hypersurface \(\Sigma_0\). Let \(\mu(y, z)\) be the function defined by \(\text{(2.21)}\). Then for all \((y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon)\), we have
\begin{equation}
L(y, z; D)^* \mu(y, z) = 0.
\end{equation}

**Remark 2.7.** Before giving the proof, let us note that the operators \(L^*\) (the adjoint in \(L^2(\Sigma_0)\)) and \(L^T\) (the adjoint in \(L^2(\Omega^*)\)) do differ in general.

**Proof.** Let \(\varphi\) be a function defined on \(\Omega^* \simeq \Sigma_0 \times (-\varepsilon, \varepsilon)\). Let \(\rho\) be the function satisfying \(\text{[2.19]}\). In Cartesian coordinates we have
\[
\langle \rho, L(x^i; \partial_i) \varphi \rangle_{L^2(\Omega^*)} = \langle L^T(x^i; \partial_i) \rho, \varphi \rangle_{L^2(\Omega^*)} = 0.
\]

Written in orthogonal coordinates \((y, z) = (x^i)\), this equation reads
\[
\int_{\Sigma_0 \times (-\varepsilon, \varepsilon)} (L(y, z; D) \varphi(y, z)) \mu(y, z) d\Sigma_0 dz = 0.
\]

We now take \(\psi(y, z) = v(y)\chi(z)\) where \(v\) is a function on \(\Sigma_0\) and \(\chi(z)\) a function on \((-\varepsilon, \varepsilon)\). Using Theorem \(\text{[2.3]}\) we then have
\[L(y, z; D) \varphi(y, z) = \chi(z) L(y, z; D) v(y)\]

Ignoring the argument \(y\) in the equation, we get
\[
\int_{\Sigma_0} \chi(z) dz \int_{\Sigma_0} \mu(z) (L(z; D) v) d\Sigma_0 = 0.
\]

As \(\chi(z)\) is an arbitrary function of \(z \in (-\varepsilon, \varepsilon)\), this implies that for all \(z \in (-\varepsilon, \varepsilon)\) and all \(v \in C^\infty(\Sigma_0)\),
\[
\int_{\Sigma_0} \mu(z) (L(z; D) v) d\Sigma_0 = \int_{\Sigma_0} v(L(z; D)^* \mu(z)) d\Sigma_0 = 0.
\]

This yields the result. \(\square\)

**Lemma 2.8.** Let \(v(t, y, z)\) be a solution of the equation \(\partial_t v = L(y, z; D) v\). For all \(z \in (-\varepsilon, \varepsilon)\), and all \(t \geq 0\), we have that
\begin{equation}
\partial_t \int_{\Sigma_0} v(t, y, z)^2 \, d\mu(y, z) = \sum_{\ell=1}^{D} \int_{\Sigma_0} \left( \sigma^\alpha_{[\ell]}(y, z) D_\alpha v(t, y, z) \right)^2 \, d\mu(y, z).
\end{equation}
Proof. For all smooth functions $v$ defined on $\Sigma_0 \times (-\varepsilon, \varepsilon)$, we have on the manifold $\Sigma_0 \times (-\varepsilon, \varepsilon)$ (ignoring again the argument $y \in \Sigma_0$),

$$L(z; D)v^2 = 2vf^\alpha(z)D_\alpha v + \sum_{\ell=1}^D \sigma^\alpha_{[\ell]}(z)D_\alpha (v\sigma^\beta_{[\ell]}(z)D_\beta v),$$

which implies that

$$L(z; D)v^2 = 2vL(z; D)v + \sum_{\ell=1}^D \left( \sigma^\alpha_{[\ell]}(z)D_\alpha v \right)^2.$$

Now if $v$ is a solution of $\partial_tv = Lv$, we have

$$\partial_t \int_{\Sigma_0} v^2 d\mu(z) = 2\langle vL(z; D)v, \mu(z) \rangle = \langle L(z; D)v^2, \mu(z) \rangle - \sum_{\ell=1}^D \int_{\Sigma_0} \left( \sigma^\alpha_{[\ell]}(z)D_\alpha v \right)^2 d\mu(z),$$

and the result follows from $L(z; D)^*\mu(z) = 0$. \qed

In the following, we denote by $\|v\|_{\mu(z)}$ the weighted norm

$$\|v\|^2_{\mu(z)} = \int_{\Sigma_0} v^2 d\mu(z).$$

Note that the norms $\|\cdot\|_{L^2(\Sigma_0)}$ and $\|\cdot\|_{\mu(0)}$ are not the same. However, as $\rho$ is a positive function, $\Sigma_0$ is compact and $z \in (-\varepsilon, \varepsilon)$, there exist two positive constants $c$ and $C$ such that for all $z \in (-\varepsilon, \varepsilon)$,

$$c \leq \mu(z) \leq C,$$

on $\Sigma_0$ and hence the norms $\|\cdot\|_{L^2(\Sigma_0)}$ and $\|\cdot\|_{\mu(z)}$ are in fact equivalent uniformly in $z$.

**Lemma 2.9.** Under Hypothesis 2.4, there exists a constant $\lambda_0$ depending on $\Sigma_0$ and $\varepsilon$ such that for all functions $v(z) \in C^0((-\varepsilon, \varepsilon), H^1(\Sigma_0))$ satisfying

$$\forall z \in (-\varepsilon, \varepsilon), \int_{\Sigma_0} v(z) d\mu(z) = 0,$$

then we have

$$\forall z \in (-\varepsilon, \varepsilon), \quad \lambda_0 \|v(z)\|^2_{\mu(z)} \leq \sum_{\ell=1}^D \|\sigma^\alpha_{[\ell]}(z)D_\alpha v(z)\|^2_{\mu(z)}.$$

Proof. As $\Sigma_0$ is a compact manifold, there exists a constant $a > 0$ such that for all functions $\varphi \in H^1(\Sigma_0)$ satisfying

$$\langle \varphi \rangle_0 := \int_{\Sigma_0} \varphi d\Sigma_0 = 0,$$

we have the Poincaré-Wirtinger inequality (see [4])

$$a\|\varphi\|^2_{L^2(\Sigma_0)} \leq \|D\varphi\|^2_{L^2(\Sigma_0)}.$$
Using (2.10) and (2.25), we have for all such functions \( \varphi \),
\begin{equation}
\sum_{\ell=1}^{D} \left\| \sigma_{\ell}^{0}(z) D_{\alpha} \varphi \right\|_{\mu(z)}^{2} \geq \Lambda \left\| D\varphi \right\|_{\mu(z)}^{2} \geq c\Lambda \left\| D\varphi \right\|_{L^{2}(\Sigma^{0})}^{2}
\end{equation}
and hence using again (2.25), we see that there exists a constant \( b > 0 \) such that for all \( z \in (-\varepsilon, \varepsilon) \), and all \( \varphi \) satisfying (2.28),
\begin{equation}
|b\|\varphi\|_{\mu(z)}^{2} \leq \sum_{\ell=1}^{D} \left\| \sigma_{\ell}^{0}(z) D_{\alpha} \varphi \right\|_{\mu(z)}^{2}.
\end{equation}
Now if \( v(z) \in C^{0}((-\varepsilon, \varepsilon), H^{1}(\Sigma_{0})) \) satisfies (2.28), we can write for all \( z \in (-\varepsilon, \varepsilon) \),
\begin{equation}
b\|v(z) - \langle v(z)\rangle_{0}\|_{\mu(z)}^{2} \leq \sum_{\ell=1}^{D} \left\| \sigma_{\ell}^{0}(z) D_{\alpha} v(z) \right\|_{\mu(z)}^{2}
\end{equation}
where \( \langle v(z)\rangle_{0} \) denotes the average (2.28) of \( v(z) \). Let \( z \in (-\varepsilon, \varepsilon) \), and assume that \( \psi \) satisfies \( \int_{\Sigma_{0}} \psi d\mu(z) = 0 \). We have that
\begin{equation}
\|\psi\|_{\mu(z)}^{2} = \|\psi - \langle \psi\rangle_{0} + \langle \psi\rangle_{0}\|_{\mu(z)}^{2} = \|\psi - \langle \psi\rangle_{0}\|_{\mu(z)}^{2} + 2 \int_{\Sigma_{0}} (\psi - \langle \psi\rangle_{0})\langle \psi\rangle_{0} d\mu(z) + \langle \psi\rangle_{0}^{2} \int_{\Sigma_{0}} d\mu(z) = \|\psi - \langle \psi\rangle_{0}\|_{\mu(z)}^{2} - \langle \psi\rangle_{0}^{2} \int_{\Sigma_{0}} d\mu(z) \leq \|\psi - \langle \psi\rangle_{0}\|_{\mu(z)}^{2},
\end{equation}
Applying this formula to \( \psi = v(z) \) and combining with (2.30) then yields the result with \( \lambda_{0} = b \).

**Theorem 2.10.** Let \( v(t, y, z) \) be a solution of the equation \( \partial_{t}v = Lv \) with initial value \( v_{0}(z) \in C^{\infty}((-\varepsilon, \varepsilon) \times \Sigma_{0}) \). Let \( \mu(z) \) the function defined in (2.21) and \( \|\cdot\|_{\mu(z)} \) the corresponding weighted \( L^{2} \) norm. Under the assumption (2.16), for all \( z \in (-\varepsilon, \varepsilon) \), and all \( t \geq 0 \), we have that
\begin{equation}
\|v(t, z) - \int v(0, z) d\mu(z)\|_{\mu(z)} \leq \|v(0, z) - \int v(0, z) d\mu(z)\|_{\mu(z)} \exp(-\gamma_{0}t)
\end{equation}
where \( \gamma_{0} = \lambda_{0}/2 \), \( \lambda_{0} \) being the constant appearing in (2.27). Moreover, for all \( J \in \mathbb{N} \), there exist constants \( \gamma_{J} \) and \( C_{J} \) such that for all \( z \in (-\varepsilon, \varepsilon) \),
\begin{equation}
\|D^{J}v(t, z)\|_{\mu(z)} \leq C_{J} \exp(-\gamma_{J}t).
\end{equation}
**Proof.** We have using (2.22),
\begin{equation}
\partial_{t} \int_{\Sigma_{0}} v(t, z) d\mu(z) = (L(z; D)v(t, z), \mu(z))_{L^{2}} = (v(t, z), L(z; D)^{k} \mu(z))_{L^{2}} = 0.
\end{equation}
This shows that the average of \( v(t, z) \) with respect to the measure \( d\mu(z) \) is constant with respect to \( t \). Now we easily see that \( L(z; D)1 = 0 \). Hence the function \( w(t, z) = v(t, z) - \int v(t, z) d\mu(z) \) satisfies the equation \( \partial_{t}w = L(z; D)w \). The equation (2.31) is then an easy consequence of the two previous lemmas.
To prove (2.32), let us consider first the case where \( J = 1 \).
From Lemma (2.28) and (2.16), we have
\begin{equation}
\partial_{t} \|w(t, z)\|_{\mu(z)}^{2} \leq -\Lambda \|Dw(t, z)\|_{\mu(z)}^{2}.
\end{equation}
Multiplying this equation by $e^{\delta t}$ with $\delta > 0$ and integrating from 0 to $T$, we obtain

$$
\int_0^T e^{\delta t} \partial_t \| w(t, z) \|_{\mu(z)}^2 \, dt + \Lambda \int_0^T e^{\delta t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt \leq 0.
$$

After integration by parts, this shows that

$$
e^{\delta T} \| w(T, z) \|_{\mu(z)}^2 + \Lambda \int_0^T e^{\delta t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt \leq \| w(0, z) \|_{\mu(z)}^2 + \delta \int_0^T e^{\delta t} \| w(t, z) \|_{\mu(z)}^2 \, dt.
$$

Using (2.31) and taking $\delta < \gamma_0$ shows that there exists a constant $M$ such that for all $z \in (-\varepsilon, \varepsilon)$,

$$
\int_0^\infty e^{\delta t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt \leq M.
$$

It is clear that the operator $L(z; D)$ acts on tensor fields, and we have

$$
D\alpha L(z; D)w = L(z; D)(D\alpha w) + [D\alpha, L(z; D)]w
$$

where $[D\alpha, L(z; D)]$ is an operator of order 1 on $\Sigma_0$ depending smoothly on $z$, and involving the curvature tensor on $\Sigma_0$ (see (2.17)). Hence $D\alpha w(t, v)$ satisfies the equation

$$
\partial_t D\alpha w = L(z; D)(D\alpha w) + [D\alpha, L(z; D)]w.
$$

Computations similar to those made in Lemma 2.8 show that there exist positive constants $c_1$, $c_2$ and $c_3$ independent on $z$ such that

$$
\partial_t \| Dw(t, z) \|_{\mu(z)}^2 \leq -c_1 \| Dw(t, z) \|_{\mu(z)}^2 + c_2 \| Dw(t, z) \|_{\mu(z)}^2 + c_3 \| w(t, z) \|_{\mu(z)}^2.
$$

Multiplying by $e^{\gamma t}$ with $0 < \gamma_1 < \delta$ and integrating from 0 to $T$ yields

$$
\int_0^T e^{\gamma t} \partial_t \| Dw(t, z) \|_{\mu(z)}^2 \, dt + c_1 \int_0^T e^{\gamma t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt \leq c_2 \int_0^T e^{\gamma t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt + c_3 \int_0^T e^{\gamma t} \| w(t, z) \|_{\mu(z)}^2 \, dt \leq C_0
$$

for some constant $C_0$ independent of $T$ and $z$, where we used (2.32), (2.34) and the fact that $\gamma_1 < \delta < \gamma_0$. We thus have

$$
e^{\gamma T} \| Dw(T, z) \|_{\mu(z)}^2 \leq C_0 + \| Dw(0, z) \|_{\mu(z)}^2 + \gamma_1 \int_0^T e^{\gamma t} \| Dw(t, z) \|_{\mu(z)}^2 \, dt
$$

and using again (2.33) we conclude that

$$
\| Dw(T, z) \|_{\mu(z)}^2 \leq C_1 \exp(-\gamma_1 T)
$$

for a positive constant $C_1$ independent of $T$ and $z$. This shows (2.32) for $J = 1$, as $Dv = Dw$. Moreover, we derive from (2.34) that

$$
\int_0^\infty e^{\gamma t} \| D^2 w(t, z) \|_{\mu(z)}^2 \, dt \leq C_0.
$$

The result is then easily shown by induction using computations similar to the previous ones, with the fact that for all $J \in \mathbb{N}$,

$$
D^J L(z; D) = L(z; D)D^J + [D^J, L(z; D)].
$$
where \(|D^J, L(z; D)|\) is an operator of order \(J - 1\) on \(\Sigma_0\) and depending smoothly on \(z\).

As \(\Sigma_0\) is a compact manifold, using (2.28) and Sobolev embedding theorems (see [4]) we immediately get the following result:

**Corollary 2.11.** Let \(v(t, y, z)\) be a solution of the equation \(\partial_t v = L v\) with initial value \(v_0(y, z) \in C^\infty(\Sigma_0 \times (-\varepsilon, \varepsilon))\). Under the assumption (2.16), for all \(J \in \mathbb{N}\), there exists constants \(\nu_J\) and \(M_J\) depending only on \(\Sigma_0\) and \(\varepsilon\) such that for all \((y, z) \in \Sigma_0 \times (-\varepsilon, \varepsilon)\), and all \(t \geq 0\),

\[
|D^J v(t, y, z)| \leq M_J \exp(-\nu_J t).
\]

**Remark 2.12.** The previous result gives uniform bounds with respect to \(z \in (-\varepsilon, \varepsilon)\). This fact is not necessary for the analysis of the properties of the projected numerical schemes defined below, which require only the previous estimates on the manifold \(\Sigma_0\). However, we believe that these uniform bounds are necessary to understand the good behaviour of nonprojected schemes; see for instance [16].

### 3. Numerical analysis

We now consider the discretisation of the stochastic differential equation (1.1). Following [20], we define the sequence

\[
\{U^\ell_p, \quad \ell = 1, \ldots, D, \quad p \in \mathbb{N}\}
\]

of i.i.d. random variables satisfying the conditions

\[
\forall p \in \mathbb{N}, \quad \forall \ell = 1, \ldots, D, \quad \mathbb{E}(U^\ell_p)^m = \begin{cases} 0 & \text{if } m = 1, 3 \text{ or } 5, \\ 1 & \text{if } m = 2, \end{cases}
\]

and the conditions

\[
\mathbb{E}(U^\ell_p)^4 < \infty \quad \text{and} \quad \mathbb{E}(U^\ell_p)^6 < \infty.
\]

Notice that these last two conditions are, in particular, satisfied if \(|U^\ell_p| \leq C\) a.s. for a constant \(C\) independent of \(p\) and \(\ell\). Examples of such families of random variables can be found in [20].

Let \(X_0 \in \Sigma_0\) and a time step \(h \leq h_0\). We assume that we can define a family \(\{X_p\}\) of random variables satisfying the following conditions:

1. **(C1)** \(\forall p \in \mathbb{N}, \quad X_p \in \Sigma_0\).
2. **(C2)** Let \(X_p(\theta) = X_p + \theta(X_{p+1} - X_p), \quad \theta \in (0, 1)\). Then
   \[
   \forall p \in \mathbb{N}, \quad \forall \theta \in (0, 1), \quad |H(X_p(\theta))| < \varepsilon.
   \]
3. **(C3)** In Cartesian coordinates we have
   \[
   X^i_{p+1} = X^i_p + \sqrt{h} \sum_{\ell=1}^D \sigma^i_{\ell p}(X_p) U^\ell_p
   + h \left( f^i(X_p) + \frac{1}{2} \sum_{\ell=1}^D \sigma^i_{\ell p}(X_p) \partial_j \sigma^j_{\ell p}(X_p) \right)
   + h w^i_p + h^{3/2} r^i_p + h^2 s^i_p
   \]

where for all \(p\), if \(\mathcal{F}_p\) denotes the \(\sigma\)-algebra generated by the family \(\{X_0, \ldots, X_p\}\), we assume that \(w^i_p, r^i_p\) and \(s^i_p\) are \(\mathcal{F}_p\) measurable, and satisfy...
where $R$ does not depend on $p \in \mathbb{N}$ and on $h \leq h_0$.

The condition (C1) will be fulfilled for projection schemes in general, while the condition (C2) is equivalent to the fact that the corresponding piecewise trajectory lies in the neighbourhood $\Omega^c$ of $\Sigma_0$; it will be fulfilled for $h$ sufficiently small in the case where the random variables $U_p^{[\ell]}$ are uniformly a.s. bounded.

The condition (C3) expresses the fact that the scheme is consistent with (1.1). Note that this condition implies, in particular, that for a given $\ell \in \{1, \ldots, D\}$, we have using (3.1),

$$\mathbb{E}(w_p^{[\ell]}|X_p) = 0.$$  

Note, moreover, that the first three terms in the expression of $X_{p+1} - X_p$ define a vector in $\Omega^c$ attached to the point $X_p \in \Sigma_0$. Using (2.10), (2.12) and (2.14), the fact that $\sigma_{[\ell]}^\alpha = f^\alpha = 0$ and the fact that $\Lambda_{\alpha\beta}^{\gamma\delta}(y, 0) = 0$, we see that in an orthogonal coordinate system, the tangential part of this vector is

$$\{X_{p+1} - X_p\}^\alpha = \sqrt{n} \sum_{\ell=1}^D \sigma_{[\ell]}^\alpha(X_p) U_p^{[\ell]}$$

while its normal part is

$$\{X_{p+1} - X_p\}^z = \frac{h}{2} \sum_{\ell=1}^D \Gamma_{\alpha\beta}^{\gamma\delta}(X_p) \sigma_{[\ell]}^\beta(X_p) \sigma_{[\ell]}^\delta(X_p)$$

3.1. Convergence result.

**Lemma 3.1.** Let $v$ be a function defined on $\Sigma_0$. Then we have

$$\forall p \in \mathbb{N}, \quad \mathbb{E}(v(X_{p+1}) - v(X_p)|X_p) = hL(X_p; D)v(X_p) + h^2 \xi_p$$

with

$$\mathbb{E}(|\xi_p|) \leq C \sup_{J=1, \ldots, 4} \|D^J v\|_{L^\infty(\Sigma_0)}$$

where $C$ does not depend on $p$ and $h \leq h_0$.

**Proof.** With the notation of (C3), the curve

$$(0, 1) \ni \theta \mapsto \alpha_p(\theta) = F_{-H(X_p)}(X_p(\theta))$$

is a well-defined curve on $\Sigma_0$ such that $\alpha_p(0) = X_p$ and $\alpha_p(1) = X_{p+1}$. We can write

$$v(X_{p+1}) - v(X_p) = \int_0^1 \frac{d}{d\theta}(v(\alpha_p(\theta))) d\theta$$
whence
\[ v(X_{p+1}) - v(X_p) = \frac{d}{d\theta} v(\alpha_p(\theta))|_{\theta=0} + \frac{1}{2} \frac{d^2}{d\theta^2} v(\alpha_p(\theta))|_{\theta=0} + \frac{1}{6} \frac{d^3}{d\theta^3} v(\alpha_p(\theta))|_{\theta=0} + \frac{1}{24} \int_0^1 (1 - \theta)^3 \frac{d^4}{d\theta^4} v(\alpha_p(\theta))d\theta. \]  

(3.6)

We can decompose the application \( \theta \rightarrow v(\alpha_p(\theta)) \) as
\[ v(\alpha_p(\theta)) = v \circ \Pi \circ X_p(\theta) \]
where \( \Pi : \Omega^\varepsilon \rightarrow \Sigma_0 \) is defined as \( \Pi(x) = F_{-H(x)}(x) \). Let us denote by \( (d\Pi)^\alpha_i(x) \) the Jacobian matrix of \( \Pi \) acting from \( T_x\Omega^\varepsilon \) to \( T_{\Pi(x)}\Sigma_0 \), and similarly, \( (d^2\Pi)_i^\alpha_j \) and \( (d^3\Pi)^\alpha_{ijk} \) the second and third differential of \( \Pi \). Here, the Greek indices are \((N - 1)\)-dimensional, while the Roman indices are \(N\)-dimensional.

We have
\[ \frac{d}{d\theta} v(\alpha_p(\theta)) = D_\alpha v(\Pi \circ X_p(\theta))(d\Pi)^\alpha_i(X_p(\theta))(X_{p+1} - X_p)^i. \]

For \( \theta = 0 \), this yields
\[ \frac{d}{d\theta} v(\alpha_p(\theta))|_{\theta=0} = D_\alpha v(X_p)(d\Pi)^\alpha_i(X_p)(X_{p+1} - X_p)^i. \]

In a normal coordinate system, the application \( \Pi \) simply reads \( (y, z) \mapsto y \), and thus its restriction to \( \Sigma_0 \) is the identity. Hence, if \( T \) is a tangent vector to \( \Sigma_0 \) viewed as a vector in \( \mathbb{R}^N \), we have as \( X_p \in \Sigma_0 \),
\[ (d\Pi)^\alpha_i(X_p) T^i = T^\alpha \]
where \( T^\alpha \) denote the components of \( T \) in a local basis of \( T_{X_p}\Sigma_0 \). Using (3.3) we have
\[ (d\Pi)^\alpha_i(X_p)(X_{p+1} - X_p)^i = \sqrt{h} \sum_{\ell=1}^D \sigma^\alpha_{[\ell]}(X_p) U_p^{[\ell]} \]
\[ + h \left( f^\alpha(X_p) + \frac{1}{2} \sum_{\ell=1}^D \sigma^\beta_{[\ell]}(X_p) D_\beta \sigma^\alpha_{[\ell]}(X_p) \right) \]
\[ + h (d\Pi)^\alpha_i(X_p) u^i_p + h^{3/2} (d\Pi)^\alpha_i(X_p) r^i_p + h^2 (d\Pi)^\alpha_i(X_p) s^i_p. \]

Denoting the conditional expectation \( \mathbb{E}(\cdot | X_p) \) by \( \mathbb{E}^{X_p} \), this implies that
\[ \mathbb{E}^{X_p} \frac{d}{d\theta} v(\alpha_p(\theta))|_{\theta=0} = \left( f^\alpha + \frac{1}{2} \sum_{\ell=1}^D \sigma^\beta_{[\ell]} D_\beta \sigma^\alpha_{[\ell]} \right) D_\alpha v + h^2 \xi_p^{(1)} \]
where all the functions are evaluated in \( X_p \), and where
\[ \mathbb{E}|\xi_p^{(1)}| \leq R\|Dv\|_{L^\infty}. \]
For the second derivative, we have

\[(3.7) \quad \frac{d^2}{d\theta^2} v(\alpha_p(\theta)) = D_\alpha v(\Pi \circ X_p(\theta))(d^2\Pi)^\alpha_{ij}(X_p(\theta))(X_{p+1} - X_p)^i(X_{p+1} - X_p)^j + D_\beta D_\alpha v(\Pi \circ X_p(\theta))[\{(d\Pi)^\alpha_{ij}(X_p(\theta))(X_{p+1} - X_p)^j\}[(d\Pi)^\beta_i(X_p(\theta))(X_{p+1} - X_p)^i].\]

Let $T$ be a vector in $\mathbb{R}^N$. In an orthogonal coordinate system around $X_p$, let $T^\alpha$ be its components in $T_{X_p} \Sigma_0$, and $T^\alpha$ its components along the normal to $T_{X_p} \Sigma_0$. We have

\[(d^2\Pi)^\alpha_{ij}(X_p)T^\alpha T^\beta = (d^2\Pi)^\alpha_{ij}(X_p)T^\alpha T^\beta + 2(d\Pi)^\beta_z(X_p)T^\beta T^z + (d^2\Pi)^\alpha_{zz}(X_p)T^z T^z;\]

but as the restriction of $\Pi$ to $\Sigma_0$ is the identity, we have $(d^2\Pi)^\alpha_{ij}(X_p)T^\alpha T^\beta = 0$. Using (C3), (3.3), (3.4) and (3.2) we easily see that

\[|E^X v(X_{p+1} - X_p)^\beta(X_{p+1} - X_p)^j| \leq C h^2\]

for a constant $C$ independent of $p$.

Hence, using the fact that $E^X v(X_{p+1} - X_p)^j = O(h)$ we get

\[E^X v(X_{p+1} - X_p)^\alpha(X_{p+1} - X_p)^j \leq C h^2\]

for a constant $C$ independent of $p$.

Similarly, we have for the second part of \((3.7)\), for $\theta = 0$,

\[E^X v(D_\beta D_\alpha v(X_p))(X_{p+1} - X_p)^\alpha(X_{p+1} - X_p)^j = \sum_{\ell=1}^D \sigma^\beta_{\ell|\ell}(X_p)\sigma^\alpha_{\ell|\ell}(X_p)D_\beta D_\alpha v(X_p) + O(h^2).\]

This implies that

\[E^X v \frac{d^2}{d\theta^2} v(\alpha_p(\theta))|_{\theta=0} = \sum_{\ell=1}^D \sigma^\beta_{\ell|\ell}\sigma^\alpha_{\ell|\ell}D_\beta D_\alpha v + h^2 \xi^{(2)}_{sp}\]

where the functions are evaluated in $X_p$, and with

\[E[\xi^{(2)}_{sp}] \leq C \sup_{J=1,2} \|D^J v\|_{L^\infty}.\]

Differentiating again the function $v(\alpha_p(\theta))$, we can see by similar computations that

\[E^X v \frac{d^3}{d\theta^3} v(\alpha_p(\theta))|_{\theta=0} \leq C h^2 \sup_{J=1,2,3} \|D^J v\|_{L^\infty}\]

and that for all $\theta \in (0,1),$

\[E^X v \frac{d^4}{d\theta^4} v(\alpha_p(\theta)) \leq C h^2 \sup_{J=1,\ldots,4} \|D^J v\|_{L^\infty}.\]

Collecting together the previous formulas, we get

\[E(v(X_{p+1} - X_p)) = h \left( f^\alpha D_\alpha v + \frac{1}{2} \sum_{\ell=1}^D (\sigma^\beta_{\ell|\ell} D_\beta \sigma^\alpha_{\ell|\ell})(D_\alpha v) + \frac{1}{2} \sum_{\ell=1}^D (\sigma^\beta_{\ell|\ell} \sigma^\alpha_{\ell|\ell} D_\beta D_\alpha v) + h^2 \xi_p\right)\]

where $\xi_p$ satisfies the estimate of the Lemma. This yields the result. \(\square\)
**Theorem 3.2.** Under Hypothesis [24], let $X(t)$ be the solution of (1.1) starting at $x_0 \in \Sigma_0$, and let $(X_p)_{p \in \mathbb{N}}$ be a family of random variables satisfying (C1)–(C3) for $h \leq h_0$ and such that $X_0 = x_0$. Let $g$ be a smooth function defined on $\Sigma_0$ and let $t_p = ph$ for $p \in \mathbb{N}$. Then we have for $h_0$ sufficiently small and $T > 0$, 

$$\forall \ t_p \leq T, \quad \left| \mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) \right| \leq C(T)h$$

where $C(T)$ depends on $T$ and $h_0$.

**Proof.** Let $u(t, x)$ be the solution of the equation $\partial_t u = Lu$ on $\Omega^c$ with the initial condition $u_0(y, z) = g(y)$ written in an orthogonal coordinate system (see Theorem [25]). Following [22], we write 

$$g(X_p) = u(0, X_p) = u(t_p, x_0) + \sum_{i=0}^{p-1} u(t_i, X_p - i) - u(t_i, X_p - i - 1) + \sum_{i=0}^{p-1} u(t_i, X_p - i - 1) - u(t_{i+1}, X_p - i - 1).$$

Hence we have

$$\mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) = \sum_{i=0}^{p-1} \mathbb{E} \left( \mathbb{E}(u(t_i, X_{p-i}) - u(t_i, X_{p-i-1})|X_{p-i-1}) \right) + \sum_{i=0}^{p-1} \mathbb{E} \left( \mathbb{E}(u(t_i, X_{p-i-1}) - u(t_{i+1}, X_{p-i-1})|X_{p-i-1}) \right).$$

Using the previous lemma, we have

$$\mathbb{E}(u(t_i, X_{p-i}) - u(t_i, X_{p-i-1})|X_{p-i-1}) = hL(X_{p-i-1}; D)u(t_i, X_{p-i-1}) + h^2 \xi_{p,i}$$

with

$$\mathbb{E} \xi_{p,i} \leq C \sup_{j=1, \ldots, 4} \|D^j u(t_i, \cdot)\|_{L^\infty(\Sigma_0)}.$$

Now, using Taylor expansion, we easily get

$$u(t_i, X_{p-i-1}) - u(t_{i+1}, X_{p-i-1}) = -hL(X_{p-i-1}; D)u(t_i, X_{p-i-1}) - \frac{1}{2} \int_0^h (h - s)L(X_{p-i-1}; D)^2 u(t_i + s, X_{p-i-1}) ds.$$

But as $L$ is an operator of order 2 in $D$, we have

$$\mathbb{E}(L(X_{p-i-1}; D)^2 u(t_i + s, X_{p-i-1})|X_{p-i-1}) \leq C \sup_{j=1, \ldots, 4} \|D^j u(t_{i+s}, \cdot)\|_{L^\infty(\Sigma_0)}$$

for some constant $C$ independent of $p \in \mathbb{N}$ and $i \in \mathbb{N}$. Collecting together the previous results, we see that

$$\mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) = h^2 \sum_{i=0}^{p-1} \zeta_{p,i}$$

where $|\zeta_{p,i}| \leq C \sup_{j=1, \ldots, 4} \|D^j u(t_{i+s}, \cdot)\|_{L^\infty(\Sigma_0)}$ and this yields the result. \hfill \Box
Remark 3.3. Combined with the techniques developed in [22], the previous result could be refined to obtain an asymptotic expansion of the error in (5.8) in powers of $h$ under more regularity assumptions.

3.2. The ergodic case. In this subsection, we assume that the hypothesis (2.19) is satisfied. We assume, moreover, that the family of random variables $\{X_p\}$ satisfies the additional hypothesis:

(C4) For all $x_0 \in \Sigma_0$ and for all open sets $W \subset \Sigma_0$ then

\[ \text{if } X_0 = x_0, \quad P(\exists p \in \mathbb{N} \mid X_p \in W) > 0. \]

As $\Sigma_0$ is compact, this hypothesis ensures that the process $\{X_p\}$ possesses a unique invariant probability measure $\mu^h$ for which it is ergodic: For all bounded functions $g$ on $\Sigma_0$,

\[ \lim_{N \to \infty} \frac{1}{N} \sum_{p=1}^{N} g(X_p) = \int_{\Sigma_0} g \, d\mu^h \quad \text{a.s.} \tag{3.11} \]

The following result refines the preceding convergence theorem. It shows that under the hypothesis (2.19) the constant in the convergence estimate (3.8) is uniform in $T$:

Lemma 3.4. Under Hypothesis (2.4) and (2.19), let $X(t)$ be the solution of (1.1) starting at $x_0 \in \Sigma_0$, and let $(X_p)_{p \in \mathbb{N}}$ be a family of random variables satisfying (C1)–(C3) for $h \leq h_0$ and such that $X_0 = x_0$. Let $g$ be a smooth function defined on $\Sigma_0$ and let $t_p = ph$ for $p \in \mathbb{N}$. Then we have for $h_0$ sufficiently small,

\[ \forall p \in \mathbb{N}, \quad \left| \mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) \right| \leq Ch \]

where $C$ does not depend on $p$ and $h \leq h_0$.

Proof. The proof is similar to the proof of Theorem 3.2. The changes are the following: equation (3.9) together with (2.35) imply that there exists positive constants $\gamma$ and $M$ independent of $h \leq h_0$, such that for all $p \in \mathbb{N}$ and $i \in \mathbb{N}$,

\[ \mathbb{E} |\xi_{p,i}| \leq M \exp(-\gamma t_i). \]

Using the same estimates in (3.10), we see that

\[ \mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) = h^2 \sum_{i=0}^{p-1} \zeta_{p,i} \]

where

\[ |\zeta_{p,i}| \leq M \exp(-\gamma t_i) \]

after a possible change of the constant $M$ which does not depend on $p$, $i$ and $h \leq h_0$. Hence we have

\[ \mathbb{E} g(X_p) - \mathbb{E} g(X(t_p)) \leq M h^2 \sum_{i=0}^{\infty} \exp(-\gamma t_i) \]

\[ = M h^2 \sum_{i=0}^{\infty} \exp(-\gamma h)^i \]

\[ = M h^2 \frac{1}{1 - \exp(-\gamma h)} \]

\[ \leq Ch \]
for some constant $C$ independent of $p$, and $h \leq h_0$ sufficiently small. This shows the result.

\begin{proof}
Using Lemma 3.4, we have
\begin{equation}
\left| \int_{\Sigma_0} g \, d\mu_h - \int_{\Sigma_0} g \, d\nu(0) \right| \leq Ch.
\end{equation}

Now, using Hypothesis 2.4 and (2.19), we have
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{p=0}^{N-1} g(X(t_p)) = \int_{\Sigma_0} g \, d\nu(0) \quad \text{a.s.}
\]

Hence, taking the limit $N \to \infty$ in (3.12) yields the result.
\end{proof}

4. Applications

In this Section, we give two classes of SDE of the form (1.1) satisfying the hypothesis (1.3), (2.19) with $\rho = 1$, and (2.16). We then give an example of numerical schemes in both of these cases, and show numerical results.

4.1. Two examples of conservative SDEs.

4.1.1. Stochastic shakers. In the spirit of [10], we introduce the following SDE, constructed as a Hamiltonian system with stochastic time dependent symplectic matrices.

Let $(G_{[t]}^i)_{1 \leq t \leq N(N-1)/2}$ be the set of $N \times N$ skew symmetric matrices with coefficients satisfying $G_{[t]}^{ij} = 0$ for all $(i, j) \in \{1, \ldots, N\}^2 \setminus \{(i(t), j(t)), (j(t), i(t))\}$, and $G_{[t]}^{i(t),j(t)} = -G_{[t]}^{j(t),i(t)} = 1$, where the indices $(i(t), j(t))$ are such that

\[
\bigcup_{t=1}^{N(N-1)/2} (i(t), j(t)) = \bigcup_{1 \leq i < j \leq N} (i, j).
\]

The set of matrices $(G_{[t]}^i)_{1 \leq t \leq N(N-1)/2}$ is a basis of the space of (real) skew symmetric matrices. In the case $N = 3$, we can choose, for instance, $(i(1), j(1)) = (1, 2)$, $(i(2), j(2)) = (1, 3)$ and $(i(3), j(3)) = (2, 3)$, which corresponds to

\[
G^1 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad G^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad G^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.
\]
We consider stochastic differential equations of the form

\begin{equation}
\label{eq:diff_eq}
\frac{dX(t)}{dt} = J^{ik} \partial_{k} H(X(t)) dt + \sum_{\ell=1}^{D} G^{i\ell}_{[\ell]} \partial_{k} H(X(t)) \circ dW_{\ell}(t)
\end{equation}

where \( J \) is an \( N \times N \) skew symmetric matrix and \( D = N(N - 1)/2 \). This SDE is of the form (1.1), with \( f^{i}(x) \) a Hamiltonian deterministic vector field, and \( \sigma^{i}_{[\ell]}(x) = G^{i\ell}_{[\ell]} \partial_{k} H(x) \). The skew-symmetry of the matrices \( J \) and \( G_{[\ell]} \) implies that the condition (1.3) is satisfied.

Moreover, it is easy to verify that this system satisfies the condition (1.9), and hence (2.10) with \( \rho = 1 \) and thus is volume preserving.

The following result shows that the condition (2.16) is fulfilled for the equation (4.1): \hfill 

**Lemma 4.1.** Assume that Hypothesis 2.21 is satisfied. Then there exists a constant \( c > 0 \) and \( \varepsilon > 0 \), such that, for all \( z \in (-\varepsilon, \varepsilon) \), \( x \in \Sigma_{z} \), and \( \xi = (\xi_{i})_{i=1}^{N} \in \mathcal{T}_{x} \Sigma_{z} \), we have

\begin{equation}
\sum_{1 \leq i < k \leq N} \left( \xi_{i} \partial_{k} H - \xi_{k} \partial_{i} H \right)^{2} 
\end{equation}

Proof. By the definition of the set of matrices \( (G^{i\ell}_{[\ell]})_{1 \leq \ell \leq N(N - 1)/2} \), the sum in equation (4.2) can be written as

\begin{equation}
\sum_{1 \leq i < k \leq N} \left( \xi_{i} \partial_{k} H - \xi_{k} \partial_{i} H \right)^{2} 
\end{equation}

Let us denote by \( n_{k} \) the \( k \)-th component of \( \nabla H/\|\nabla H\| \), i.e. \( n_{k} = \partial_{k} H/\|\nabla H\| \). We have

\[
\sum_{1 \leq i < k \leq N} (\xi_{i} n_{k} - \xi_{k} n_{i})^{2} = \sum_{1 \leq i < k \leq N} \xi_{i}^{2} n_{k}^{2} + \xi_{k}^{2} n_{i}^{2} - 2\xi_{i} n_{k} \xi_{k} n_{i} = \sum_{k \neq i} \xi_{i}^{2} n_{k}^{2} - 2 \sum_{1 \leq i < k \leq N} \xi_{i} n_{k} \xi_{k} n_{i}.
\]

Writing

\[
\sum_{k \neq i} \xi_{i}^{2} n_{k}^{2} = \sum_{k=1}^{N} \xi_{i}^{2} \sum_{i \neq k}^{N} n_{i}^{2}
\]

and using the fact that \( \sum_{i} n_{i}^{2} = 1 \), we find that

\[
\sum_{1 \leq i < k \leq N} (\xi_{i} n_{k} - \xi_{k} n_{i})^{2} = \sum_{k=1}^{N} \xi_{k}^{2} (1 - n_{k}^{2}) - 2 \sum_{1 \leq i < k \leq N} \xi_{i} n_{k} \xi_{k} n_{i} = \sum_{k=1}^{N} \xi_{k}^{2} - \sum_{k=1}^{N} \xi_{k}^{2} n_{k}^{2} - 2 \sum_{1 \leq i < k \leq N} \xi_{i} n_{k} \xi_{k} n_{i} = \|\xi\|^{2} - \left( \sum_{k=1}^{N} \xi_{k} n_{k} \right)^{2}.
\]
Now, if $\xi \in T_x\Sigma_z$, we have by definition,
\[
\sum_k \xi_k \partial_k H = 0.
\]
Gathering these results together, we find that
\[
\sum_{1 \leq i < k \leq N} \left( \xi_i \partial_k H - \xi_k \partial_i H \right)^2 = \| \nabla H \|^2 \| \xi \|^2
\]
and this yields the result with
\[
c = \min_{z \in (-\varepsilon, \varepsilon)} \min_{x \in \Sigma_z} \parallel \nabla H(x) \parallel,
\]
which is positive under Hypothesis 2.1 for sufficiently small $\varepsilon$. □

4.1.2. Projected gradient dynamics. We would now like to introduce another dynamics ergodic with respect to the microcanonical measure (see [7]).

For any point $x \in \mathbb{R}^N$, let us define the orthogonal projector on $T_x\Sigma_H(x)$:
\[
P(x) = \text{Id} - \frac{\nabla^H \otimes \nabla^H}{\| \nabla^H \|^2}(x).
\]
In Cartesian coordinates, its components are written as
\[
P^j_i = \delta^j_i + \frac{(\nabla^H)^j_\ell (\nabla^H)^\ell_j}{\| \nabla^H \|}.
\]

Let us now consider the following SDE:
\[
dX^i(t) = -P^j_i(X(t))\nabla^j \left( g \circ H + \ln \| \nabla H \| \right)(X(t)) \, dt + \sqrt{2}P^j_i(X(t))\circ dW^j(t)
\]
where $X^i(t) \in \mathbb{R}^N$, $W^i(t)$ is an $N$-dimensional Brownian motion, and $g : \mathbb{R} \rightarrow \mathbb{R}$ denotes any smooth function. This SDE is of the form (1.1) with $\sigma(x) = \sqrt{2}P(x)$, and $f(x) = -P(x)\nabla(g(H(x)) + \ln \| \nabla H(x) \|)$. Using the definition of $P(x)$, it is clear that (4.3) is satisfied.

Let us introduce the function $V$ defined by
\[
V(x) = \left( g \circ H + \ln \| \nabla H \| \right)(x).
\]
The fact that (2.19) is satisfied for $\rho = 1$ is equivalent to the equation
\[
-\nabla_i P^\ell_j \nabla^\ell V + \nabla_i P^\ell_j \nabla^\ell P^\ell_j = 0;
\]
but we have (see Lemma A.1 in [7] for the first equality)
\[
P^\ell_j \nabla^\ell P^\ell_j = -P^\ell_j \nabla^\ell \ln \| \nabla H \|
\]
\[
= -P^\ell_j \nabla^\ell V,
\]
and hence the equation (4.4) is satisfied.

Finally, the equation (4.4) satisfies (2.16): We have for all $x \in \Sigma_z$ and for all $\xi \in T_x\Sigma_z$,
\[
\sum_{j=1}^N (\xi_j P^j_i)^2 = \sum_{j=1}^N \xi_j^2,
\]
which shows (2.16) with $\Lambda = 1/2$.

\footnote{Recall that in Cartesian coordinates we have $\nabla^i H = \nabla_i H = \partial_i H$, and similarly for the components of the projection operator $P^j_i = P^j_i = P^j_i$.}
Remark 4.2. More generally (see [7]), using such projected SDE, it is possible to sample any measure of the form $\exp(-\beta V)|\nabla \xi|^{-1}dS_\xi$ where $V : \mathbb{R}^N \to \mathbb{R}$ is an energy, $\xi : \mathbb{R}^N \to \mathbb{R}$ is a reaction coordinate and $S_\xi = \{x \in \mathbb{R}^N | \xi(x) = z\}$. One needs to consider the solution $X(t)$ of the SDE

$$dX^i(t) = -P^*_i(X(t))\nabla V(X(t)) \, dt + \sqrt{2}\beta^{-1}P^*_i(X(t)) \circ dW^j(t)$$

where $P(x)$ is the projector

$$P(x) = \text{Id} - \frac{\nabla \xi \otimes \nabla \xi}{\|\nabla \xi\|^2}(x).$$

4.2. Numerical schemes. We now give examples of schemes corresponding to both the previous systems of SDE.

4.2.1. Methods for stochastic shakers. A general method for constructing schemes adapted to (4.11) is the following: For each discrete time step $t_p = ph$, draw a symplectic matrix $(hJ + \sqrt{h}G[\ell]U^\ell_p)$, where $U^\ell_p$ satisfies the hypothesis of Section 3, and then apply any energy preserving method to the deterministic system associated with this matrix. In order to be consistent with the SDE, the scheme has to be of order 2 which will be automatically the case if the method is symmetric. For instance, the following two schemes can be considered:

1. Symmetric projection. This is the scheme defined as follows: Assume that $X_p \in \Sigma_0$, and define $X_{p+1}$ by

$$\begin{align*}
\dot{X}_p &= X_p + \nabla H(X_p)^T \mu, \\
\dot{X}_{p+1} &= \dot{X}_p + (hJ + G[\ell]U^\ell_p)\nabla H\left(\frac{\dot{X}_p + \dot{X}_{p+1}}{2}\right), \\
X_{p+1} &= \dot{X}_{p+1} + \nabla H(X_{p+1})^T \mu,
\end{align*}$$

(4.7)

where $\mu$ is chosen such that $X_{p+1} \in \Sigma_0$ (see [13]).

2. Symmetric discrete gradient. This is the scheme defined by

$$X_{p+1} = X_p + (hJ + G[\ell]U^\ell_p)\nabla H(X_p, X_{p+1}),$$

(4.8)

where $\nabla H(\tilde{y}, y)$ is a function satisfying:

$$\begin{align*}
\nabla H(\tilde{y}, y)^T (\tilde{y} - y) &= H(\tilde{y}) - H(y), \\
\nabla H(y, y) &= \nabla H(y).
\end{align*}$$

With these properties, we can easily prove that $H(X_{p+1}) = H(X_p)$. We can take, for instance,

$$\nabla H(\tilde{y}, y) = \nabla H(\overline{y}) + \frac{H(\tilde{y}) - H(y) - \nabla H(\overline{y})^T \Delta y}{\|\Delta y\|^2} \Delta y$$

with $\overline{y} = \frac{1}{h}(\tilde{y} + y)$ and $\Delta y = \tilde{y} - y$. This scheme, named midpoint discrete gradient, was introduced by Gonzalez [12].

It can be checked that these schemes satisfy the hypothesis (C1) to (C4) for $h$ sufficiently small provided the random variables $U^\ell_p$ are a.s. bounded.
4.2.2. Projected gradient dynamics schemes. For the discretization of the SDE \((4.4)\), we note that it can be rewritten in the form
\[
dX(t) = -\nabla V(X(t)) \, dt + \sqrt{2} \, dW(t) + \nabla H(X(t)) \, d\Lambda(t),
\]
where \(\Lambda(t)\) is a real-valued process, which can be interpreted as the Lagrange multiplier associated with the constraint \(H(X(t)) = H(X_0)\). This process can be decomposed into two parts:
\[
d\Lambda(t) = d\Lambda^m(t) + d\Lambda^f(t).
\]
The martingale part \(\Lambda^m(t)\) is
\[
d\Lambda^m(t) = -\sqrt{2} \frac{\nabla H}{\|\nabla H\|^2} (X(t)) \cdot dW(t),
\]
where \(\cdot\) implicitly denotes the Itô product. The bounded variation part \(\Lambda^f(t)\) is
\[
d\Lambda^f(t) = \frac{\nabla H}{\|\nabla H\|^2} (X(t)) \cdot \nabla V(X(t)) \, dt + \frac{\nabla H}{\|\nabla H\|^2} (X(t)) \cdot \Gamma(X(t)) \, dt,
\]
where \(\Gamma(x)\) denotes the mean curvature vector to \(\Sigma_{H(x)}\) at point \(x\):
\[
\Gamma = -\nabla \cdot \left( \frac{\nabla H}{\|\nabla H\|} \right) \frac{\nabla H}{\|\nabla H\|}.
\]
We can now consider two discretizations of \((4.4)\) which can be shown to be consistent (see \([7, 15]\)). The first one is
\[
\begin{align*}
X_{n+1} &= X_n - \nabla V(X_n) \, h + \sqrt{2h} \, U_n + \Delta \Lambda_{n+1} \nabla H(X_{n+1}), \\
\text{where } \Delta \Lambda_{n+1} \text{ is such that } H(X_{n+1}) &= H_0,
\end{align*}
\]
where \(h\) is the time step and \(U_n\) is a \(3N\)-dimensional standard Gaussian random vector satisfying the hypothesis in the beginning of Section 3. Notice that \((4.14)\) admits a natural variational interpretation, since \(X_{n+1}\) can be seen as the closest point on the submanifold \(\Sigma_z\) to the predicted position \(X_n - \nabla V(X_n) \Delta t + \sqrt{2h} \, U_n\).

The real \(\Delta \Lambda_{n+1}\) is then the Lagrange multiplier associated with the constraint \(H(X_{n+1}) = H_0\).

Another possible discretization of \((4.14)\) is
\[
\begin{align*}
X_{n+1} &= X_n - \nabla V(X_n) \, h + \sqrt{2h} \, U_n + \Delta \Lambda_{n+1} \nabla H(X_n), \\
\text{where } \Delta \Lambda_{n+1} \text{ is such that } H(X_{n+1}) &= H_0.
\end{align*}
\]
Although this scheme is not naturally associated with a variational principle, it may be more practical since its formulation is more explicit.

It can be checked that these schemes satisfy the hypotheses \((C1)\) to \((C4)\) for \(h\) sufficiently small provided the random variables \(U_n\) are a.s. bounded.

These numerical schemes for constrained Brownian dynamics are in the spirit of the so-called RATTLE \([3]\) and SHAKE \([19]\) algorithms classically used for constrained Hamiltonian dynamics, and also related to the algorithms proposed in \([23, 2, 18]\).
4.3. Numerical results. We consider the following Hamiltonian in dimension $N = 4$:

\begin{equation}
H(x_1, x_2, x_3, x_4) = \frac{1}{2}(x_1^2 + x_2^2) + V(x_3, x_4)
\end{equation}

where

\[ V(x_3, x_4) = (x_3^2 - 1)^2 + (x_4 + x_3^2 - 1)^2. \]

We are interested in computing the averages of the following eight observables,

\[(x_3, x_3^2, x_4, x_4^2, x_1^2, x_2^2, V(x_3, x_4))\]

with respect to the NVE measure (1.6). To this aim, we simulate approximations $X_p, p \geq 1$ of the previous systems, and compute the time average (3.11) for sufficiently large $N$.

The Hamiltonian (4.16) represents a particle in dimension 2, with momentum $(p_x, p_y) = (x_1, x_2)$ and position $(q_x, q_y) = (x_3, x_4)$ in the double-well potential $V$. It is possible to calculate analytically the exact averages of the observables mentioned above. We refer to Section 3.4 in [6].

In Tables 1 and 2, we show the result for the method (4.8). The initial conditions are such that $x_1 = 0.5, x_2 = -0.5, x_3 = -1$ and $x_4 = 0.5$, so that the initial energy level is $H = 0.5$. The 95% confidence interval is obtained with 30 independent runs. The time step is $h$, and the number of points in the trajectory is represented by the number $N$.

**Table 1. Midpoint discrete gradient algorithm**

<table>
<thead>
<tr>
<th>Observables</th>
<th>Exact</th>
<th>Approx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_3$</td>
<td>-0.94459</td>
<td>-0.94580</td>
</tr>
<tr>
<td>$x_3^2$</td>
<td>0.92843</td>
<td>0.93064</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.98964</td>
<td>0.9937</td>
</tr>
<tr>
<td>$x_4^2$</td>
<td>0.07156</td>
<td>0.07075</td>
</tr>
<tr>
<td>$x_1^2$</td>
<td>0.25517</td>
<td>0.25470</td>
</tr>
<tr>
<td>$x_2^2$</td>
<td>0.24482</td>
<td>0.24496</td>
</tr>
<tr>
<td>$V(x_3, x_4)$</td>
<td>0.25517</td>
<td>0.25488</td>
</tr>
</tbody>
</table>

**Table 2. Midpoint discrete gradient algorithm**

<table>
<thead>
<tr>
<th>Observables</th>
<th>Exact</th>
<th>Approx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_3$</td>
<td>-0.94459</td>
<td>-0.94580</td>
</tr>
<tr>
<td>$x_3^2$</td>
<td>0.92843</td>
<td>0.92850</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.98964</td>
<td>0.99009</td>
</tr>
<tr>
<td>$x_4^2$</td>
<td>0.07156</td>
<td>0.07075</td>
</tr>
<tr>
<td>$x_1^2$</td>
<td>0.25517</td>
<td>0.25550</td>
</tr>
<tr>
<td>$x_2^2$</td>
<td>0.24482</td>
<td>0.24458</td>
</tr>
<tr>
<td>$V(x_3, x_4)$</td>
<td>0.25517</td>
<td>0.25488</td>
</tr>
</tbody>
</table>
In Table 3, we give the results obtained with a slight modification of the dynamics (4.14), namely:

\[
\begin{align*}
X_{n+1} &= X_n - \nabla V(X_n) h + \sqrt{2h} U_n \\
+ \frac{\nabla H(X_n)}{\|\nabla H\|} \nabla H(X_n) - \sqrt{2h} \frac{\nabla H(X_n)}{\|\nabla H\|} \nabla H(X_n) + \Delta \Lambda_{n+1} \nabla H(X_{n+1}),
\end{align*}
\]

where \(\Delta \Lambda_{n+1}\) is such that \(H(X_{n+1}) = H_0\).

The two additional terms are easy to compute, do not modify the consistency of the scheme, and improve the convergence in the projection step. In practice, we use an Uzawa algorithm to perform the projection step. The initial values are the same as in the previous section.

### Table 3. Projected gradient algorithm

<table>
<thead>
<tr>
<th>Observables</th>
<th>Exact</th>
<th>Approx</th>
<th>(h = 5.10^{-4}, N = 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_3)</td>
<td>-0.94459</td>
<td>-0.94094</td>
<td>[-0.94288, -0.93900]</td>
</tr>
<tr>
<td>(x_2^2)</td>
<td>0.92843</td>
<td>0.92051</td>
<td>[0.91688, 0.92414]</td>
</tr>
<tr>
<td>(x_3^4)</td>
<td>0.98964</td>
<td>0.97029</td>
<td>[0.96342, 0.97715]</td>
</tr>
<tr>
<td>(x_4)</td>
<td>0.07156</td>
<td>0.07372</td>
<td>[0.06666, 0.08080]</td>
</tr>
<tr>
<td>(x_4^2)</td>
<td>0.25517</td>
<td>0.26230</td>
<td>[0.26071, 0.26390]</td>
</tr>
<tr>
<td>(x_4^4)</td>
<td>0.24482</td>
<td>0.25074</td>
<td>[0.24910, 0.25238]</td>
</tr>
<tr>
<td>(x_2^2)</td>
<td>0.24482</td>
<td>0.25223</td>
<td>[0.25074, 0.25372]</td>
</tr>
<tr>
<td>(V(x_3,x_4))</td>
<td>0.25517</td>
<td>0.24851</td>
<td>[0.24766, 0.24936]</td>
</tr>
</tbody>
</table>

We observe that the schemes described above provide satisfactory numerical results. Note that the numerical simulation tends to show that the variance of the result with the midpoint discrete gradient algorithm applied to the stochastic shakers is smaller than in the case of the projected gradient dynamics.

**Acknowledgment**

This work is supported by the ANR INGEMOL of the French Ministry of Research. The authors are glad to thank Arnaud Debussche and Denis Talay for helpful comments during the preparation of this work.

**References**


