Geometry of non-holonomic diffusion

Simon Hochgerner, Tudor S. Ratiu

Geometry of non-holonomic diffusion

Simon Hochgerner Tudor S. Ratiu

Abstract

We study stochastically perturbed non-holonomic systems from a geometric point of view. In this setting, it turns out that the probabilistic properties of the perturbed system are intimately linked to the geometry of the constraint distribution. For G-Chaplygin systems, this yields a stochastic criterion for the existence of a smooth preserved measure. As an application of our results we consider the motion planning problem for the noisy two-wheeled robot and the noisy snakeboard.

Keywords. non-holonomic system, symmetry, measure, reduction, diffusion, Brownian motion, generator, Chaplygin system, snakeboard, two-wheeled carriage

Contents

S. Hochgerner: Section de Mathématiques, Station 8, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne,Switzerland; e-mail: simon.hochgerner@gmail.com

T.S. Ratiu: Section de Mathématiques, Station 8 and Bernoulli Center, Station 15, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; e-mail: tudor.ratiu@epfl.ch

Mathematics Subject Classification (2010): Primary 37Jxx, 58J65 ; Secondary 93Exx

1 Introduction

The goal of this paper is the study of stochastic non-holonomic systems. This is a natural continuation of the work on stochastic Hamiltonian systems pioneered by Bismut [5] and revitalized, brought up to date, and expanded by Lázaro-Camí and Ortega [32] who also connected it to symmetries, momentum maps, and reduction.

1.A Motivation and basic idea

A non-holonomic system is, essentially, a rigid body together with a set of constraints on the velocities. A prototypical example is the Chaplygin ball ([10]; for a modern treatment see [13] and [12, Chapter 6]). Here, the configuration space is the direct product Lie group $G = \mathrm{SO}(3) \times \mathbb{R}^2$, describing orientation and position of the ball, and the kinetic energy is specified by a left-invariant metric μ; there are two (non-integrable) velocity constraints so that the ball does not slip, i.e., the point of contact of the ball and the plane has zero velocity. Without constraints (which is clearly not the case in the problem just presented), this would describe the motion of a rigid body in the plane, hence it would be a Hamiltonian system.

Stochastically perturbed versions of the latter setting (i.e., without constraints) have been considered by Lázaro-Camí and Ortega [33, Section 7.3]: Let h^0 be the kinetic energy Hamiltonian of a left invariant metric on the Lie group G, $\{y_i\}_i$ an orthonormal basis of the Lie algebra $\mathfrak{g}$ of G, $\{u_i\}_i$ its extension to a left invariant frame on G, and $h^i : T^*G \to \mathbb{R}$, $(q,p) \mapsto \langle p, u_i(q)\rangle$. Note that h^i is the component $\langle J^R, y_i\rangle$ of the momentum map $J^R : T^*G \to \mathfrak{g}^*$ defined by the lift to T^*G of right translation of G on itself. The $\mathbb{R}\times\mathfrak{g}^*$-valued function $H = (h_0, h^i)$ on T^*G is left invariant. Following [32, 33] and assuming that the perturbation is given by white noise, the stochastic rigid body is thus modeled by the Stratonovich equation

$$\delta\Gamma = X_{h^0}(\Gamma)\delta t + \sum X_{h^i}(\Gamma)\delta W^i, \tag{1.1}$$

where X_h denotes the Hamiltonian vector field of the function $h : T^*G \to \mathbb{R}$ and $W = \{W^i\}$ is Brownian motion in $\mathfrak{g} \cong \mathbb{R}^n$. A physical system modeled by this equation is that of a rigid body subject to small random impacts. Note that, since u_i is auto-parallel for the Levi-Civita connection, the equation $\delta\Gamma = \sum X_{h^i}(\Gamma)\delta W^i$ yields the Hamiltonian construction of Brownian motion, as in [32].

To pass to the nonholonomic setting, we note that the equations of motion of the constrained (Chaplygin) ball can be encoded in the vector field PX_{h^0} where P is the constraint force projection and is defined in (2.12) below. The effect of P is to force the dynamics generated by X_{h^0} to satisfy the constraints. Thus, the idea of 'the Hamiltonian construction of stochastic non-holonomic systems' is to apply P to (1.1). In fact, since PX_{h^0} is nothing but the non-holonomic vector field (see Section 2), we will focus on studying the effects of P on the second term in equation (1.1). This yields non-holonomic constraints on the operator which is used to construct Brownian motion, thus leading to 'constrained Brownian motion' described by

$$\delta\Gamma = \sum P(\Gamma)X_{h^i}(\Gamma)\delta W^i. \tag{1.2}$$

As it stands, this equation has some problems. It depends very much on the basis $\{u_i\}_i$ that was chosen in the definition of the h^i. For example, since the no-slip constraints are actually right invariant, one could have chosen a right invariant frame. But then the Hamiltonian description of Brownian motion needs a correction term involving the Levi-Civita connection of μ. This approach has been taken in [23]. However, the basis dependence implies that the generator of (1.2) also changes when we pass to a different frame, and there would be many natural choices depending on whether the frame should be left or right invariant, adapted to the constraint distribution, or the direct product structure of G, etc. Even if one ignores these issues, it is not clear what to do if the configuration space is not parallelizable. For all these reasons we transfer the construction to the bundle of orthonormal frames itself. It is only then that the generator of the resulting 'constrained

Brownian motion' is basis independent. This constrained Brownian motion has some interesting features:

- To visualize it, we can think of a microscopic robot (or ball, snakeboard, etc.) subject to molecular bombardment. The robot thus experiences small impacts from all sides (isotropic in space) which force it to move around, but it still has to respect the constraints.

- Now, it turns out, that the geometry of the constraints determines the probabilistic properties of the perturbed system. Indeed, if the constraints are integrable, then the robot's net drift will vanish. However, when the constraints are non-integrable and non-mechanical (which is the generic case) the Gaussian noise will induce a net drift on the robot. In Section 4 we quantify this drift in terms of the geometry of the constraint distribution. *Mechanical constraints* are given, by definition as level sets of conserved quantities, such as momentum maps. E.g., the constraints could be given by the horizontal bundle of the mechanical connection, which is just orthogonal to the vertical bundle in the case of a symmetry group action.

- This leads to a dictionary between probabilistic aspects of the perturbed system and classical properties of the original (deterministic) non-holonomic system. See Theorem 1.2 below for a preliminary statement of this dictionary and Section 4 for further details.

1.B Description of contents and results

Since this paper addresses both the geometric mechanics and the stochastic differential equations communities, we shall give the necessary background for all concepts and quote the main results that are used later on. The paper is self contained. We briefly present the main results and the structure of the paper.

Non-holonomic systems

We start by recalling the necessary facts, concepts, and results of non-holonomic systems and their geometry. This includes a careful presentation of symmetries, reduction, and conditions for the existence of a (smooth) preserved measure. We will have to rephrase some of the existing results in view of applying them to our stochastic study later on and develop the theory in the direction needed in subsequent sections in the paper.

Thus, we will have to give complete proofs not only for some of the known results, due to our reformulation, but we also need to establish new formulas. For example, the *global* formula (2.8) of the symplectic form on the tangent bundle given in terms of an underlying Riemannian metric on configuration space is new, as far as we know. In (2.12)

we introduce the above mentioned constraint force projection and explain its properties to prepare for Section 4. We also study Chaplygin systems, which are non-holonomic systems with a particularly rich geometric structure, and the symmetry reduction of such systems. One of the main points of Section 2 is the presentation of a certain one-from β which, according to Proposition 2.5, characterizes the existence of a (smooth) preserved measure for a given Chaplygin system. This result has been previously derived in [8] but both our proof and our interpretation of the relevant one-form β are different. In fact, our formulation of β in (2.25) is a prerequisite for Section 4.

Stochastic dynamics on manifolds

First, we recall some notions about manifold valued stochastic differential equations and diffusions from [26, 18].

Then we study symmetries of Stratonovich equations. We consider a manifold Q together with a proper action by a Lie group G and a diffusion Γ^Q generated by a Stratonovich operator $\mathcal{S}$ from $T\mathbb{R}^{k+1}$ to TQ satisfying the equivariance relation (3.11). In this setting, the Stratonovich operator does not (in general) induce a Stratonovich operator on the base Q/G; however, the diffusion Γ^Q and its generator A^Q are projectable to Q/G. Thus, there is an induced diffusion $\Gamma^{Q/G}$ with induced generator $A^{Q/G}$ on the base space Q/G. See Theorem 3.2.

Two examples for this procedure of 'equivariant reduction' are the Eells-Elworthy-Malliavin construction of Brownian motion (cf. equation (3.10)) on a Riemannian manifold and the stochastic Calogero-Moser systems (see [24]), as remarked in Subsection 3.B. In particular, we allow for non-free G-actions on Q and hence Q/G is, in general, not a smooth manifold but a stratified space. Thus, we extend the reduction theorem of [33, Theorem 3.1] to the case when the Stratonovich operator on the total space is not invariant but equivariant with respect to a symmetry group action.

This naturally leads to the introduction, in Subsection 3.C, of certain notions of equivariant diffusions, previously studied in [16, 17]. The material of this subsection will also be useful in Section 5. In particular, we prove a mean reconstruction equation for diffusions in principal bundles which is analogous to a concept by the same name in mechanics (see, e.g., [1, §4.3], [35, §3], [36, Theorem 11.8]) and uses that of [16, 17].

Non-holonomic diffusions

This section contains the main results of the paper. We introduce constrained Brownian motion as motivated above. This involves a careful analysis of the underlying geometry. Then we study the generator and symmetry reduction of the resulting diffusion process. The reduction relies on Theorem 3.2.

The surprising fact in this regard, is that there is a very strong interrelation of some probabilistic aspects of constrained Brownian motion and certain deterministic properties of the original non-holonomic system. A first instance of this relation is:

Theorem 1.1. *Constrained Brownian motion is a martingale with respect to the non-holonomic connection on the configuration space.*

A second result yields a probabilistic characterization of the existence of a preserved measure which is a very important concept in the theory of non-holonomic systems (see [3, 6, 10, 15, 25, 22, 30]):

Theorem 1.2. *Let $(Q, \mathcal{D}, L)$ be a G-Chaplygin system such that the base $M := Q/G$ is compact. Let Γ^M be the non-holonomic diffusion in M associated to these data. Then the following are equivalent:*

(1) *$(Q, \mathcal{D}, L)$ has a (smooth) preserved measure;*

(2) *Γ^M is time-reversible;*

(3) *Γ^M has vanishing entropy production rate.*

The compactness assumption on M is met in all classical examples such as the Chaplygin ball or the two-wheeled robot. This theorem sums up some of the results of Sections 4 and 3.D, where also the relevant notions are introduced.

Examples

As examples, we consider the two-wheeled robot and the snakeboard. The former is G-Chaplygin and does (in general) not allow for a preserved measure. The latter is not a Chaplygin system but does fit the general set-up of Section 4. For both of these examples we consider also the stochastic perturbation of deterministic trajectory planning. This emphasizes the way in which the noise couples with the constraints to produce a non-trivial drift vector field (the emergence of which is at the heart of the geometry of Section 4); this is in sharp contrast to stochastic Hamiltonian systems. Indeed, the Hamiltonian analogue of non-holonomic reduction is reduction at the 0-level set of the standard cotangent bundle momentum map, which reduces Brownian motion to Brownian motion in the base with respect to the induced metric. This is a manifestation of the idea that the amount by which a non-holonomic system differs from a Hamiltonian one can be measured by the amount by which the induced diffusion differs from Brownian motion – and vice versa.

However, in the non-holonomic setting, the constraints induce a drift giving rise to drifted Brownian motion on the base space. This drift is quantified in Section 4 and we use it to make the perturbed motion follow a given curve *on average*. We show how the

explicit form of the drift allows, in principle, for a simple numerical implementation to solve such a motion planning problem. It should be noted, though, that we have made no attempt to study stability or convergence properties of the resulting numerical algorithm. Similar problems have been treated, from a different perspective, in the engineering literature; see [2, 42] and the references therein.

2 Non-holonomic systems

We recall some facts about non-holonomic and, specifically, G-Chaplygin systems. Then we give a necessary and sufficient condition for the existence of a preserved measure that is suitable for our applications in Section 4.

A *non-holonomic system* is a triple $(Q, \mathcal{D}, \mathcal{L})$ consisting of a n-dimensional configuration manifold Q, a constraint distribution $\mathcal{D} \subset TQ$ which is smooth and of constant rank $r < n$ (i.e., it is a vector subbundle of TQ of rank r), and a smooth Lagrangian function $\mathcal{L} : TQ \to \mathbb{R}$. The dynamics of $(Q, \mathcal{D}, \mathcal{L})$ are given by the Lagrange-d'Alembert principle; see [3, 4, 6, 9, 10, 25, 30]. Throughout this paper, we assume that $\mathcal{L}$ is the kinetic energy of a Riemannian metric μ on Q.

2.A Almost Hamiltonian formulation

Since $TQ \ni u_q \mapsto \mu(q)(u_q, \cdot) \in T^*Q$ is a vector bundle isomorphism covering the identity on Q, we shall identify the vector bundles TQ with T^*Q. We follow [4] to give an almost Hamiltonian description of the dynamics of $(Q, \mathcal{D}, \mathcal{L})$. Let $\tau_Q : TQ \to Q$ be the tangent bundle projection and $\iota : \mathcal{D} \hookrightarrow TQ$ the inclusion. Define

$$\mathcal{C} := \left\{ X_{u_q} \in T\mathcal{D} \mid u_q \in \mathcal{D},\ T_{u_q}(\tau_Q \circ \iota)\left(X_{u_q}\right) \in \mathcal{D} \right\} = \left(T(\tau_Q \circ \iota)\right)^{-1}(\mathcal{D}). \tag{2.1}$$

In standard vector bundle charts of TQ and TTQ, we write u_q as $(q, \dot{q})$ and X_{u_q} as $(q, \dot{q}, \delta q, \delta\dot{q})$, respectively. Since $(\tau_Q \circ \iota)(q, \dot{q}) = q$, it follows that $T(\tau_Q \circ \iota)(q, \dot{q}, \delta q, \delta\dot{q}) = (q, \delta q)$ and hence $\mathcal{C} = \{(q, \dot{q}, \delta q, \delta\dot{q}) \mid (q, \dot{q}), (q, \delta q) \in \mathcal{D}\}$, $\ker\left(T(\tau_Q \circ \iota)(q, \dot{q}, \cdot, \cdot)\right) = \{(q, \dot{q}, 0, \delta\dot{q}) \mid \delta\dot{q} \in \mathbb{R}^n\}$. Thus $\mathcal{C}$ is a vector subbundle of $T\mathcal{D}$ of rank $2r$. (If $\mathcal{D}$ is the horizontal subbundle of a principal connection of some proper and free G-action on Q then $\mathcal{C}$ is the horizontal space of the tangent lifted G-action on $\mathcal{D}$. See (2.14) below.) According to [4, Section 5] we have

$$\big(T(TQ)\big)|\mathcal{D} = \mathcal{C} \oplus \mathcal{C}^{\Omega} \tag{2.2}$$

where $\mathcal{C}^{\Omega} := \{X_{u_q} \in T_{u_q}(TQ) \mid u_q \in \mathcal{D},\ \Omega(u_q)(X_{u_q}, Y_{u_q}) = 0, \forall Y_{u_q} \in \mathcal{C}\}$ is the Ω-orthogonal complement of $\mathcal{C}$ in $\big(T(TQ)\big)|\mathcal{D}$; Ω denotes the canonical symplectic form on $TQ \cong T^*Q$. We will prove identity (2.2) later on, after the proof of Proposition 2.1.

For reasons that will become clear in Section 4, we elaborate on (2.2). We use the Levi-Civita connection ∇^{μ} on $TQ \to Q$ to decompose $TTQ = \mathrm{Hor}^{\mu} \oplus \mathrm{Ver}(\tau_Q)$, where

$\mathrm{Ver}(\tau_Q) = \ker(T\tau_Q : TTQ \to TQ)$ is the *vertical* and $\mathrm{Hor}^\mu \subset TTQ$ is the *horizontal subbundle*. Recall that a curve $v(t)$ in TQ is *horizontal* if its covariant derivative $\frac{Dv(t)}{Dt} := \frac{d}{ds}\big|_{s=0} \mathbb{P}_t^{t+s} v(t+s)$ vanishes; here $\mathbb{P}_t^{t+s} : T_{q(t+s)}Q \to T_{q(t)}Q$ is the parallel transport operator of the Levi-Civita connection ∇^μ and $q(t) := \tau_Q(v(t))$. Alternatively, since $\frac{Dv(t)}{Dt} = \nabla^\mu_{dq(t)/dt} v(t)$, or in coordinates, $\frac{Dv^i(t)}{Dt} = \frac{dv^i(t)}{dt} + \Gamma^i_{jk}(q(t))\frac{dq^j(t)}{dt}v^k(t)$, the curve $v(t)$ is horizontal if and only if in any standard tangent bundle chart

$$\frac{dv^i(t)}{dt} + \Gamma^i_{jk}(q(t))\frac{dq^j(t)}{dt}v^k(t) = 0. \tag{2.3}$$

A vector $X_{u_q} \in T_{u_q}TQ$ is called *horizontal* if it is tangent to a horizontal curve. The *horizontal space* $\mathrm{Hor}^\mu_{u_q} \subset T_{u_q}TQ$ is the vector subspace formed by all horizontal vectors.

If $u_q = \dot{q}^i \frac{\partial}{\partial q^i} \in T_qQ$, the decomposition of a vector $X_{u_q} = A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial \dot{q}^i} \in T_{u_q}TQ$ in its horizontal and vertical part is

$$A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial \dot{q}^i} = \left(A^i \frac{\partial}{\partial q^i} - \Gamma^i_{jk}\dot{q}^j A^k \frac{\partial}{\partial \dot{q}^i} \right) + \left(\Gamma^i_{jk}\dot{q}^j A^k + B^i \right) \frac{\partial}{\partial \dot{q}^i}. \tag{2.4}$$

Indeed, since $T_{u_q}\tau_Q \left(R^i \frac{\partial}{\partial q^i} + S^i \frac{\partial}{\partial \dot{q}^i} \right) = R^i \frac{\partial}{\partial q^i}$ it follows that

$$\ker T_{u_q}\tau_Q = \left\{ S^i \frac{\partial}{\partial \dot{q}^i} \;\middle|\; S^i \in \mathbb{R} \right\}$$

which shows that the second summand in (2.4) is vertical. The first summand is horizontal since it verifies the horizontality condition (2.3) (with $v^i = \dot{q}^i$, $A^i = \frac{dq^i}{dt}$, and $\frac{dv^i}{dt} = -\Gamma^i_{jk}\dot{q}^j A^k$). In particular, note that $T_{u_q}\tau_Q : \mathcal{C}_{u_q} \cap \mathrm{Hor}^\mu_{u_q} \to \mathcal{D}_q$ is an isomorphism: $A^i \frac{\partial}{\partial q^i} - \Gamma^i_{jk}\dot{q}^j A^k \frac{\partial}{\partial \dot{q}^i} \in \mathrm{Hor}^\mu_{u_q}$ maps to the given vector $A^i \frac{\partial}{\partial q^i} \in \mathcal{D}_q$. Similarly $T_{u_q}\tau_Q : \mathrm{Hor}^\mu_{u_q} \to T_qQ$ is an isomorphism. Its inverse is the horizontal lift mapping which is often written as a map $\mathrm{hl}^\mu : TQ \times_Q TQ \cong \mathrm{Hor}^\mu$, $(u_q, v_q) \mapsto (T_{u_q}\tau_Q|\mathrm{Hor}^\mu_{u_q})^{-1}(v_q)$. Interpreting $\mathrm{pr}_1 : TQ \times_Q TQ \to TQ$ as a vector bundle over TQ with base the first factor, makes $\mathrm{hl}^\mu : TQ \times_Q TQ \xrightarrow{\sim} \mathrm{Hor}^\mu$ into a vector bundle isomorphism covering the identity on TQ.

Let $K : \mathrm{Ver}(\tau_Q) \to TQ \times_Q TQ$ be the inverse to the vertical lift mapping $\mathrm{vl} : TQ \times_Q TQ \xrightarrow{\sim} \mathrm{Ver}(\tau_Q)$ defined by $\mathrm{vl}(u_q, v_q) := \frac{d}{dt}\big|_{t=0}(u_q + tv_q)$, for all $u_q, v_q \in T_qQ$. In standard coordinates, $K(q, \dot{q}, 0, \delta\dot{q}) = (q, \dot{q}, q, \delta\dot{q})$. In particular, $K(X_{u_q}) \in T_qQ$. In addition, $T\tau_Q : \mathrm{Hor}^\mu \to TQ$ and $K : \mathrm{Ver}(\tau_Q) \to TQ$ restricted to each fiber over TQ are linear isomorphisms. Let P_{hor} and P_{ver} denote the horizontal and vertical projections associated to Hor^μ. By abuse of notation, we sometimes write K also for $K \circ P_{\mathrm{ver}} : TTQ \to \mathrm{Ver}(\tau_Q) \to TQ$. We have thus the vector bundle isomorphism over $\mathcal{D}$

$$\mathcal{C} \xrightarrow{\sim} (\mathcal{D} \times_Q \mathcal{D}) \oplus \ker T(\tau_Q \circ \iota), \tag{2.5}$$

$$X_{u_q} \longmapsto \left(u_q, T_{u_q}\tau_Q\left(X_{u_q}\right), K\left(P_{\mathrm{ver}}\left(X_{u_q}\right)\right) \right), \tag{2.6}$$

$$\mathrm{hl}^\mu_{u_q}(v_q) + \mathrm{vl}(u_q, w_q) \longleftarrow\!\shortmid (u_q, v_q, w_q), \tag{2.7}$$

where we regard $\mathcal{D} \times_Q \mathcal{D} \ni (u_q, v_q) \mapsto u_q \in \mathcal{D}$ as a vector bundle over $\mathcal{D}$. Notice also that $T\mathcal{D} \supset \ker T(\tau_Q \circ \iota) = \bigsqcup_{(q,u)\in\mathcal{D}} \mathrm{vl}_{(q,u)}\mathcal{D}_q$.

Proposition 2.1. *The canonical symplectic form $\Omega \in \Omega^2(TQ)$ has the expression*

$$\Omega(u_q)\left(X_{u_q}, Y_{u_q}\right) = \mu(q)(T_{u_q}\tau_Q(X_{u_q}), K(Y_{u_q})) - \mu(q)(T_{u_q}\tau_Q(Y_{u_q}), K(X_{u_q})), \quad (2.8)$$

for any $q \in Q$, $u_q \in T_qQ$, $X_{u_q}, Y_{u_q} \in T_{u_q}(TQ)$.

Proof. In an arbitrary standard tangent bundle chart, we have

$$\Omega = \frac{\partial \mu_{ik}}{\partial q^j}\dot{q}^k \mathbf{d}q^i \wedge \mathbf{d}q^j + \mu_{ij}\mathbf{d}q^i \wedge \mathbf{d}\dot{q}^j, \quad (2.9)$$

where the Riemannian metric is written as $\mu = \mu_{ij}\mathbf{d}q^i \otimes \mathbf{d}q^j$, with $\mu_{ij} = \mu_{ji}$. Thus, if

$$X_{u_q} = A^i\frac{\partial}{\partial q^i} + B^i\frac{\partial}{\partial \dot{q}^i}\,, \qquad Y_{u_q} = C^i\frac{\partial}{\partial q^i} + D^i\frac{\partial}{\partial \dot{q}^i}\,,$$

we get

$$\Omega(u_q)(X_{u_q}, Y_{u_q}) = \frac{\partial \mu_{ik}}{\partial q^j}\dot{q}^k(A^iC^j - A^jC^i) + \mu_{ij}(A^iD^j - C^iB^j). \quad (2.10)$$

On the other hand, $T_{u_q}\tau_Q(X_{u_q}) = A^i\frac{\partial}{\partial q^i}$, $T_{u_q}\tau_Q(Y_{u_q}) = C^i\frac{\partial}{\partial q^i}$ and

$$K\left(A^i\frac{\partial}{\partial q^i} + B^i\frac{\partial}{\partial \dot{q}^i}\right) = \left(\Gamma^i_{jk}\dot{q}^jA^k + B^i\right)\frac{\partial}{\partial q^i},$$
$$K\left(C^i\frac{\partial}{\partial q^i} + D^i\frac{\partial}{\partial \dot{q}^i}\right) = \left(\Gamma^i_{jk}\dot{q}^jC^k + D^i\right)\frac{\partial}{\partial q^i}$$

by (2.4) and the definition of K. Therefore,

$$\begin{aligned}
&\mu(q)(T_{u_q}\tau_Q(X_{u_q}), K(Y_{u_q})) - \mu(q)(T_{u_q}\tau_Q(Y_{u_q}), K(X_{u_q}))\\
&\quad= \mu_{ij}\Gamma^j_{rk}\dot{q}^r(A^iC^k - A^kC^i) + \mu_{ij}(A^iD^j - C^iB^j)\\
&\quad= \mu_{ij}\frac{1}{2}\mu^{js}\left(\frac{\partial \mu_{sk}}{\partial q^r} + \frac{\partial \mu_{sr}}{\partial q^k} - \frac{\partial \mu_{rk}}{\partial q^s}\right)\dot{q}^r(A^iC^k - A^kC^i) + \mu_{ij}(A^iD^j - C^iB^j)\\
&\quad= \frac{1}{2}\left(\frac{\partial \mu_{ik}}{\partial q^r} + \frac{\partial \mu_{ir}}{\partial q^k} - \frac{\partial \mu_{rk}}{\partial q^i}\right)\dot{q}^r(A^iC^k - A^kC^i) + \mu_{ij}(A^iD^j - C^iB^j)\\
&\quad= \frac{1}{2}\left(\frac{\partial \mu_{ik}}{\partial q^r} - \frac{\partial \mu_{ir}}{\partial q^k} - \frac{\partial \mu_{rk}}{\partial q^i}\right)\dot{q}^r(A^iC^k - A^kC^i) + \frac{\partial \mu_{ir}}{\partial q^k}\dot{q}^r(A^iC^k - A^kC^i)\\
&\qquad+ \mu_{ij}(A^iD^j - C^iB^j)\\
&\quad= -\Gamma^s_{ik}g_{rs}\dot{q}^r(A^iC^k - A^kC^i) + \frac{\partial \mu_{ir}}{\partial q^k}\dot{q}^r(A^iC^k - A^kC^i) + \mu_{ij}(A^iD^j - C^iB^j)\\
&\quad= \frac{\partial \mu_{ir}}{\partial q^k}\dot{q}^r(A^iC^k - A^kC^i) + \mu_{ij}(A^iD^j - C^iB^j)
\end{aligned}$$

because Γ^s_{ik} is symmetric and $(A^iC^k - A^kC^i)$ is skew-symmetric in (i,k). However, this expression coincides with (2.10) which proves (2.8). □

Thus by (2.8) we get

$$\begin{aligned}\mathcal{C}^{\Omega}_{u_q} &= \big\{X_{u_q} \in T_{u_q}(TQ) \mid u_q \in \mathcal{D}_q, \\ &\qquad \mu(q)(T_{u_q}\tau_Q(X_{u_q}), K(Y_{u_q})) - \mu(q)(T_{u_q}\tau_Q(Y_{u_q}), K(X_{u_q})) = 0,\ \forall Y_{u_q} \in \mathcal{C}_{u_q}\big\} \\ &= \big\{X_{u_q} \in T_{u_q}(TQ) \mid u_q \in \mathcal{D}_q,\ K(P_{\mathrm{ver}}(X_{u_q})) \in \mathcal{D}^{\perp}_q,\ T_{u_q}\tau_Q(P_{\mathrm{hor}}(X_{u_q})) \in \mathcal{D}^{\perp}_q\big\} \\ &\cong (\mathcal{D} \times_Q \mathcal{D}^{\perp}) \oplus \bigsqcup_{u_q \in \mathcal{D}} \mathrm{vl}_{u_q}(\mathcal{D}^{\perp})\end{aligned} \tag{2.11}$$

since $K, T_{u_q}\tau_Q : \mathcal{C}_{u_q} \to \mathcal{D}_q$ are surjective, where $\mathcal{D}^{\perp} \subset TQ$ is the μ-orthogonal of $\mathcal{D}$ and the vector bundle isomorphism in the last line of (2.11) is given by $X_{u_q} \mapsto (u_q, T_{u_q}\tau_Q(X_{u_q}), P_{\mathrm{ver}}(X_{u_q}))$. This expression of $\mathcal{C}^{\Omega}$ and (2.1) show that $\mathcal{C} \cap \mathcal{C}^{\Omega} = \{0\}$ which proves (2.2).

In particular, if

$$P : (T(TQ))|\mathcal{D} = \mathcal{C} \oplus \mathcal{C}^{\Omega} \to \mathcal{C} \tag{2.12}$$

is the projection along $\mathcal{C}^{\Omega}$ and $\Pi : TQ = \mathcal{D} \oplus \mathcal{D}^{\perp} \to \mathcal{D}$ is the orthogonal projection then it follows that

$$T(\tau_Q \circ \iota) \circ P = \Pi \circ T(\tau \circ \iota). \tag{2.13}$$

Indeed, using the above description of $\mathcal{C}$ and $\mathcal{C}^{\Omega}$, this follows immediately by decomposing $\big(T(TQ)\big)|\mathcal{D}$ into its horizontal and vertical parts.

Let $\mathcal{H}$ be the kinetic energy Hamiltonian on TQ which we regard as the Legendre transform of $\mathcal{L}$. Then the dynamics of the non-holonomic system $(Q, \mathcal{D}, \mathcal{L})$ are given by the vector field

$$X^{\mathcal{C}}_{\mathcal{H}} := PX_{\mathcal{H}} \in \mathfrak{X}(\mathcal{D})$$

where $X_{\mathcal{H}}$ is the Hamiltonian vector field of $\mathcal{H}$. More generally, for a function $f \in C^{\infty}(TQ)$ we regard $X^{\mathcal{C}}_f := PX_f \in \mathfrak{X}(\mathcal{D})$ as the non-holonomic vector field of f. Let $\Omega^{\mathcal{C}}$ denote the fiberwise restriction of $\iota^*\Omega$ to $\mathcal{C} \times \mathcal{C}$. Then (2.2) implies that $\Omega^{\mathcal{C}}$ is non-degenerate and we may rewrite the defining equation for $X^{\mathcal{C}}_f$ as

$$\mathbf{i}_{X^{\mathcal{C}}_f}\Omega^{\mathcal{C}} = (\mathbf{d}f)^{\mathcal{C}}$$

where $(\mathbf{d}f)^{\mathcal{C}}$ is the fiberwise restriction of $\iota^*(\mathbf{d}f)$ to $\mathcal{C}$.

2.B G-Chaplygin systems

Now we shall consider the case when the non-holonomic system is invariant under a group action such that the constraints are given by a principal bundle connection. A *G-Chaplygin system* consists of a Riemannian configuration space (Q, μ), a Lie group G with Lie algebra $\mathfrak{g}$ which acts freely and properly on (Q, μ) by isometries, and a principal

bundle connection $\mathcal{A} \in \Omega^1(Q;\mathfrak{g})$ on $\pi : Q \twoheadrightarrow Q/G =: M$. For $\xi \in \mathfrak{g}$ denote by $\xi_Q \in \mathfrak{X}(Q)$ the infinitesimal generator defined by

$$\xi_Q(q) := \frac{d}{dt}\Big|_{t=0} \exp(t\xi) \cdot q$$

for all $q \in Q$, where $\exp : \mathfrak{g} \to G$ is the exponential map.

The Lagrangian of this system is the kinetic energy $\mathcal{L} := \frac{1}{2}\|\cdot\|^2_\mu$. It is also assumed that the constraint distribution is the horizontal subbundle of the connection $\mathcal{A}$, i.e., $\mathcal{D} := \ker \mathcal{A} \subset TQ$. Thus the system $(Q, \mathcal{D}, \mathcal{L})$ is a non-holonomic system and the dynamics are determined by the Lagrange-d'Alembert equations; see [3, 4, 6, 9, 10, 25, 30]. It is *not* assumed that $\mathcal{D}$ is orthogonal to the vertical space $\ker T\pi$.

Since $\mathcal{D}$ is the horizontal subbundle, it is invariant with respect to the tangent lifted G-action on TQ. Thus we obtain a principal G-fiber bundle $\mathcal{D} \twoheadrightarrow \mathcal{D}/G = TM$. This bundle carries an induced connection $\iota^*\tau^*\mathcal{A}$, where $\iota : \mathcal{D} \hookrightarrow TQ$ is the inclusion and $\tau : TQ \to Q$ is the tangent bundle projection. Its associated horizontal bundle is

$$\ker(\tau \circ \iota)^*\mathcal{A} = \{u_q \in T\mathcal{D} \mid T(\tau \circ \iota)u_q \in \ker \mathcal{A} = \mathcal{D}\} = \mathcal{C}. \tag{2.14}$$

Let μ_0 denote the induced Riemannian metric on $M := Q/G$. Then the isomorphism

$$T_q\pi : (\mathcal{D}_q, \mu(q)|\mathcal{D}_q) \to \big(T_{\pi(q)}M, \mu_0(\pi(q))\big)$$

is an isometry for the indicated inner products for all $q \in Q$.

2.C The non-holonomic correction

In order to carry out non-holonomic reduction we need to introduce a two-form on TM induced by the momentum map and the curvature $\mathrm{Curv}^{\mathcal{A}} \in \Omega^2(Q;\mathfrak{g})$ of the connection $\mathcal{A}$. As we shall see in the next subsection, this form is the correction that one needs to subtract from the canonical symplectic form in order to give an almost Hamiltonian formulation of the reduced non-holonomic system. To define this form, we need three ingredients:

(i) The *adjoint bundle*: Let G act on $Q \times \mathfrak{g}$ by the (free and proper) action given by $g \cdot (q, \xi) := (g \cdot q, \mathrm{Ad}_g\, \xi)$, for all $g \in G$, $q \in Q$, $\xi \in \mathfrak{g}$, and let $\widetilde{\mathfrak{g}} := Q \times_G \mathfrak{g} = (Q \times \mathfrak{g})/G$ be the orbit space. Elements of $\widetilde{\mathfrak{g}}$ are denoted by $[q, \xi]_G$. The projection $\rho : \widetilde{\mathfrak{g}} \ni [q, \xi]_G \mapsto \pi(q) \in M$ defines the *adjoint vector bundle* whose fibers are Lie algebras.

(ii) The *curvature on the base*: $\mathrm{Curv}^{\mathcal{A}} \in \Omega^2(Q;\mathfrak{g})$ naturally induces a two-form on $\mathrm{Curv}^{\mathcal{A}}_0 \in \Omega^2(M;\widetilde{\mathfrak{g}})$ on the base M with values in the adjoint bundle $\widetilde{\mathfrak{g}}$ by

$$\mathrm{Curv}^{\mathcal{A}}_0(\pi(q))\,(T_q\pi(u_q), T_q\pi(v_q)) := \big[q, \mathrm{Curv}^{\mathcal{A}}(q)(u_q, v_q)\big]_G$$

for all $q \in Q$, $u_q, v_q \in T_qQ$.

(iii) The *momentum map of the tangent lifted G-action*: $\mathbf{J}_G : TQ \to \mathfrak{g}^*$ is defined by $\langle \mathbf{J}_G(u_q), \xi \rangle = \mu(q)(u_q, \xi_Q(q))$ for all $\xi \in \mathfrak{g}$, $u_q \in TQ$, and is equivariant.

To get a grip on the non-holonomic correction two-form, we begin describing it if G is a commutative group. Then the adjoint bundle is trivial: $\rho : \widetilde{\mathfrak{g}} = M \times \mathfrak{g} \to M$ is the projection on the first factor. Thus, $\mathrm{Curv}_0^{\mathcal{A}} \in \Omega^2(M; \mathfrak{g})$ and we define the *non-holonomic correction two-form* $\Xi \in \Omega^2(TM)$ by $\Xi := \left\langle \mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}, \tau_M^* \, \mathrm{Curv}_0^{\mathcal{A}} \right\rangle$, that is,

$$\Xi(u_x)\left(X_{u_x}, Y_{u_x}\right) := \left\langle \mathbf{J}_G\left(\mathrm{hl}^{\mathcal{A}}{}_q(u_x)\right), \mathrm{Curv}_0^{\mathcal{A}}(x)\left(T_{u_x}\tau_M(X_{u_x}), T_{u_x}\tau_M(Y_{u_x})\right)\right\rangle \quad (2.15)$$

for all $x \in M$, $u_x \in T_xM$, $X_{u_x}, Y_{u_x} \in T_{u_x}(TM)$, where $\mathrm{hl}^{\mathcal{A}}{}_q := \left(T_q\pi|\mathcal{D}_q\right)^{-1} : T_xM \to \mathcal{D}_q \subset TQ$, $x = \pi(q)$, is the horizontal lift operator associated to the connection $\mathcal{A}$. The pairing on the right hand side of this formula is between $\mathfrak{g}^*$ and $\mathfrak{g}$. The right hand side of this formula seems to depend on $q \in Q$. However, this is not the case because the horizontal lifts at two distinct points in Q are related by a group element and the momentum map is invariant under the G-action (since G is commutative).

As stated, this formula does not make sense for general Lie groups because the momentum map is $\mathfrak{g}^*$-valued and the curvature on the base is $\widetilde{\mathfrak{g}}$-valued so the pairing makes no sense. However, the idea for the general formula is based on (2.15). We define $\Xi \in \Omega^2(TM)$ by

$$\begin{aligned}&\Xi(u_x)(X_{u_x}, Y_{u_x})\\ &\quad := \langle \mathbf{J}_G(\mathrm{hl}^{\mathcal{A}}{}_q(u_x)), \mathrm{Curv}^{\mathcal{A}}(q)\left(\mathrm{hl}^{\mathcal{A}}{}_q\left(T_{u_x}\tau_M(X_{u_x})\right), \mathrm{hl}^{\mathcal{A}}{}_q\left(T_{u_x}\tau_M(Y_{u_x})\right)\right)\rangle \end{aligned} \quad (2.16)$$

for $X_{(x,u)}, Y_{(x,u)} \in T_{(x,u)}(TM)$ and $q \in \pi^{-1}(x)$; since both entries in this pairing are G-equivariant the ambiguity cancels out, that is, the right hand side in (2.16) does not depend on q but only on $\pi(q) = x$.

Due to the importance of this formula we make a few additional comments. Recall that the momentum map $\mathbf{J}_G : TQ \to \mathfrak{g}^*$ is equivariant with respect to the coadjoint action on $\mathfrak{g}^*$. The tangent lifted G-action restricts to an action on $\mathcal{D} \subset TQ$; indeed $\mathcal{D} = \ker \mathcal{A}$ is the horizontal subbundle and is hence G-invariant. Corresponding to the G-principal bundle projection $\mathcal{D} \twoheadrightarrow \mathcal{D}/G = TM$ there is a natural connection which is induced from the connection $\mathcal{A}$ on $Q \twoheadrightarrow Q/G$, namely $\iota^*\tau^*\mathcal{A}$ where $\iota : D \hookrightarrow TQ$ is the inclusion and $\tau : TQ \to Q$ is the tangent bundle projection. The curvature of $\iota^*\tau^*\mathcal{A}$ is $\iota^*\tau^*\mathrm{Curv}^{\mathcal{A}}$ which is equivariant: $l_g^*\iota^*\tau^*\mathrm{Curv}^{\mathcal{A}} = \mathrm{Ad}_g \circ \left(\iota^*\tau^*\mathrm{Curv}^{\mathcal{A}}\right)$, where $l_g : \mathcal{D} \to \mathcal{D}$ is the action of $g \in G$. Thus the two-form $\langle \mathbf{J}_G, \iota^*\tau_Q^*\mathrm{Curv}^{\mathcal{A}} \rangle$ defines a G-invariant two-form on $\mathcal{D}$. This two-form is, moreover, horizontal: since $\iota^*\tau^*\mathrm{Curv}^{\mathcal{A}}$ is a curvature form on $\mathcal{D} \twoheadrightarrow \mathcal{D}/G$ it vanishes upon insertion of vertical vectors, whence the same holds also for $\langle \mathbf{J}_G, \iota^*\tau^*\mathrm{Curv}^{\mathcal{A}} \rangle$. Thus the two-form $\langle \mathbf{J}_G, \iota^*\tau_Q^*\mathrm{Curv}^{\mathcal{A}} \rangle$ is basic and hence drops to a well-defined two-form Ξ on $\mathcal{D}/G = TM$. Implementing the computations suggested above gives (2.16).

2.D Non-holonomic reduction

Identify TQ with T^*Q by the metric μ and TM with T^*M by the metric μ_0. Consider the orbit projection map

$$T\pi|\mathcal{D} : \mathcal{D} \twoheadrightarrow \mathcal{D}/G = TM.$$

We may also associate a fiberwise inverse to this mapping which is given by the horizontal lift mapping $\mathrm{hl}^{\mathcal{A}} : Q \times_M TM \to \mathcal{D}$ associated to $\mathcal{A}$. The following statements are proved in [4, 15, 25].

Proposition 2.2 (Non-holonomic reduction)**.** *The following hold.*

(1) $\Omega^{\mathcal{C}}$ *descends to a non-degenerate two-form* Ω_{nh} *on* TM.

(2) $\Omega_{\mathrm{nh}} = \Omega_M - \Xi \in \Omega^2(TM)$, *where* $\Omega_M = -\mathbf{d}\theta_M$ *is the canonical symplectic form on* TM *and* Ξ *is the non-holonomic correction two-form given by* (2.16).

(3) *Let* $h : TQ \to \mathbb{R}$ *be* G*-invariant. Then the vector field* $X_h^{\mathcal{C}}$ *is* $T\pi|\mathcal{D}$*-related to the vector field* $X_{h_0}^{\mathrm{nh}}$ *on* TM *defined by*

$$\mathbf{i}_{X_{h_0}^{\mathrm{nh}}}\Omega_{\mathrm{nh}} = \mathbf{d}h_0$$

where $h_0 : TM \to \mathbb{R}$ *is the induced Hamiltonian.*

In general, Ω_{nh} is an almost symplectic form, that is, it is non-degenerate and non-closed. We will denote the reduced Hamiltonian by $\mathcal{H}_{\mathrm{c}}$ and refer to the almost Hamiltonian system $(TM, \Omega_{\mathrm{nh}}, \mathcal{H}_{\mathrm{c}})$ as the *reduced data*. The identity $\Omega_{\mathrm{nh}} = \Omega_M - \Xi$ appears for the first time, albeit not completely explicitly, in [4]. A proof using moving frames is given in [15] where it is also called the "$\langle J, K\rangle$-formula". A different proof following the above outline is contained in [25, Prop 2.2].

2.E The preserved measure

Does $(TM, \Omega_{\mathrm{nh}}, \mathcal{H}_{\mathrm{c}})$ possess a preserved measure? This is an important question since it says something about the possible existence of asymptotic equilibria and also plays a prominent role in the theory of integration of non-holonomic systems. Correspondingly, this topic is touched upon in all of [3, 6, 10, 15, 25, 22, 30]. In [8], a necessary and sufficient condition for the existence of a preserved measure in terms of local coordinates on the base manifold M is given. We derive derive below an equivalent formulation of this result which is more closely adapted to the Riemannian structure on M. This point of view will then be exploited in Section 4 in the stochastic context.

For brevity we will denote $X = X_{\mathcal{H}_{\mathrm{c}}}^{\mathrm{nh}}$ in this subsection. Let Ω^m, $m = \dim M$, be the Liouville volume on TM. Then *there is a preserved measure for the flow of* X *if and only*

if there is a strictly positive function $\mathcal{N} : M \to \mathbb{R}$ such that $(\mathcal{N} \circ \tau_M)\Omega^m$ is preserved, that is,

$$\pounds_X((\mathcal{N} \circ \tau_M)\Omega^m) = 0; \tag{2.17}$$

see [8] for a proof. In such a case $\mathcal{N}$ is called the *density* of the preserved measure with respect to the Liouville volume. As shown in [8, Remark 7.4], it suffices to consider density functions on M.

To reformulate condition (2.17) we want to use the fact that (M, μ_0) is a Riemannian manifold. Hence we equip TM with the Sasaki metric σ associated to μ_0 (see, e.g., [20]),

$$\sigma(u_x)(X_{u_x}, Y_{u_x}) := \mu_0(x)\left(T_x\tau_M(X_{u_x}), T_x\tau_M(Y_{u_x})\right) + \mu_0(x)\left(K_M(X_{u_x}), K_M(Y_{u_x})\right) \tag{2.18}$$

for all $X_{u_x}, Y_{u_x} \in T_{u_x}(TM)$, where $\tau_M : TM \to M$ is the tangent bundle projection and $K_M : \mathrm{Ver}(\tau_M) \to TM \times_M TM$ is the inverse of the vertical lift map $\mathrm{vl}_M : TM \times_M TM \xrightarrow{\sim} \mathrm{Ver}(\tau_M)$; note, in particular that $K_M(X_{u_x}) \in T_xM$. We recall some of the key properties of the Sasaki metric; see [20] for proofs.

(i) The Sasaki metric σ is the unique Riemannian metric on TM such that $\tau_M : (TM, \sigma) \to (M, \mu_0)$ is a Riemannian submersion, that is, the isomorphism

$$T_{u_m}\tau_M : \Big((\ker T_{u_m}\tau_M)^{\perp_\sigma} = \mathrm{Hor}_{u_m}, \sigma(u_m)\Big) \longrightarrow (T_mM, \mu_0(m))$$

is an isometry (for the indicated inner products) for all $u_m \in TTM$, where $\perp_\sigma$ denotes the perpendicular relative to the Sasaki inner product $\sigma(u_m)$ on $T_{u_m}(TM)$.

(ii) Hor and Ver are σ-perpendicular complements of each other: $\mathrm{Hor} = \mathrm{Ver}^{\perp_\sigma}$.

(iii) The vertical lift map $\mathrm{vl} : TM \times_M TM \to \mathrm{Ver} \subset TTM$ is an isometry of vector bundles over TM, thinking of the projection onto the first factor $\mathrm{pr}_1 : TM \times_M TM \to TM$ as a vector bundle over TM and μ_0 as a vector bundle metric.

Given a vector field X on M we shall denote its horizontal lift relative to the Riemannian metric μ_0 by $X^h \in \mathfrak{X}(TM, \mathrm{Hor})$ and its vertical lift by $X^v \in \mathfrak{X}(TM, \mathrm{Ver})$.

Lemma 2.3 (The non-holonomic vector field)**.** *If $X_0 = X_{\mathcal{H}_c} = \Omega_M^{-1}(\mathrm{d}\mathcal{H}_c) \in \mathfrak{X}(TM)$ is the standard Hamiltonian vector field, $X = \Omega_{\mathrm{nh}}^{-1}(\mathrm{d}\mathcal{H}_c) \in \mathfrak{X}(TM)$ is the non-holonomic vector field, and $\{u_1, \dots, u_m\}$ is a local orthonormal frame on M, then the following hold:*

$$T_{u_x}\tau_M(X(u_x)) = T_{u_x}\tau_M(X_0(u_x)) = u_x \text{ for all } u_x \in TM, \tag{2.19}$$

$$\begin{aligned} P_{\mathrm{ver}}\Big(X - X_0\Big)(u_x) &= -\sum_{i=1}^m \Xi(u_x)\Big(X(u_x), u_i^h(u_x)\Big)\, u_i^v(u_x) \\ &= -\sum_{i=1}^m \langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}{}_q)(u_x), (\mathrm{Curv}_q^{\mathcal{A}} \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\rangle\, u_i^v(u_x) \end{aligned} \tag{2.20}$$

where the second equation holds locally in the domain of definition of the given frame, and is well-defined independently of the choice of $q \in \pi^{-1}(x)$.

Proof. We begin by noting that $\Omega_M(X_0, Y) = \mathbf{d}\mathcal{H}_{\mathrm{c}}(Y) = \Omega_{\mathrm{nh}}(X, Y)$ for all $Y \in \mathfrak{X}(TM)$. Hence by Propositions 2.2 and 2.1

$$\begin{aligned}\Xi(u_x)\Big(X(u_x), Y(u_x)\Big) &= \Omega_M(X - X_0, Y)_{u_x} \\ &= \mu_0(x)(T_{u_x}\tau_M((X - X_0)(u_x)), K_{u_x}Y(u_x)) \\ &\quad - \mu_0(x)(T_{u_x}\tau_M(Y(u_x)), K_{u_x}(X - X_0)(u_x)).\end{aligned}$$

This implies, firstly, that $T_{u_x}\tau_M(X(u_x)) = T_{u_x}\tau_M(X_0(u_x)) = u_x$ since $\mathcal{H}_{\mathrm{c}}$ is the kinetic energy Hamiltonian of the induced metric μ_0. Secondly, since $K_{u_x}\left(u_i^v(u_x)\right) = u_i(x) = T_{u_x}\tau_M\left(u_i^h(u_x)\right)$, we find locally

$$\begin{aligned}P_{\mathrm{ver}}(X - X_0)(u_x) &= \sum \sigma\Big((X - X_0)(u_x), u_i^v(u_x)\Big)\, u_i^v(u_x) \\ &= \sum \mu_0\Big(K_{u_x}(X - X_0)(u_x), u_i(u_x)\Big)\, u_i^v(u_x) \\ &= -\sum \Omega_M(X - X_0, u_i^h)_{u_x}\, u_i^v(u_x) \\ &= -\sum \Xi(u_x)\Big(X, u_i^h\Big)\, u_i^h(u_x) \\ &= -\sum \langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}})_q(u_x), (\mathrm{Curv}_q^{\mathcal{A}} \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\rangle\, u_i^v(u_x)\end{aligned}$$

where we have used $T_{u_x}\tau_M\left(X(u_x)\right) = u_x$ in the last line. □

Let vol_σ be the volume form on TM induced by the Riemannian metric σ. We shall prove the following formula:

$$\mathrm{vol}_\sigma = \tfrac{1}{m!}\Omega^m. \tag{2.21}$$

Indeed by (2.9), denoting by $\mathfrak{S}_m$ the permutation group of $\{1, \ldots, m\}$, we have

$$\begin{aligned}\Omega^m &= \left(\frac{\partial \mu_{ik}}{\partial q^j}\dot{q}^k \mathbf{d}q^i \wedge \mathbf{d}q^j + \mu_{ij}\mathbf{d}q^i \wedge \mathbf{d}\dot{q}^j\right)^m \\ &= (\mu_{ij}\mathbf{d}q^i \wedge \mathbf{d}\dot{q}^j)^m \\ &= \sum_{\pi \in \mathfrak{S}_m} \mu_{1\pi(1)} \cdots \mu_{m\pi(m)} \mathbf{d}q^1 \wedge \mathbf{d}\dot{q}^{\pi(1)} \wedge \ldots \wedge \mathbf{d}q^m \wedge \mathbf{d}\dot{q}^{\pi(m)} \\ &= m! \left(\sum_{\pi \in \mathfrak{S}_m} (-1)^{\mathrm{sign}\,\pi} \mu_{1\pi(1)} \cdots \mu_{m\pi(m)}\right) \mathbf{d}q^1 \wedge \mathbf{d}\dot{q}^1 \wedge \ldots \wedge \mathbf{d}q^m \wedge \mathbf{d}\dot{q}^m \\ &= m! \det(\mu_{ij}) \mathbf{d}q^1 \wedge \mathbf{d}\dot{q}^1 \wedge \ldots \wedge \mathbf{d}q^m \wedge \mathbf{d}\dot{q}^m.\end{aligned}$$

On the other hand, in the coordinates $(v^1, v^2, \ldots, v^{2m-1}, v^{2m})$ of TM, where $v^{2i-1} = q^i$ and $v^{2i} = \dot{q}^i$ for $i = 1, \ldots, m$, we have by the usual formula of the Riemannian volume, $\mathrm{vol}_\sigma = \sqrt{\det(\sigma_{IJ})}\mathbf{d}q^1 \wedge \mathbf{d}\dot{q}^1 \wedge \ldots \wedge \mathbf{d}q^m \wedge \mathbf{d}\dot{q}^m$, where $\sigma_{IJ} := \sigma\left(\frac{\partial}{\partial v^I}, \frac{\partial}{\partial v^J}\right)$. Since, by

the definition (2.18) of the Sasaki metric we have $\mu_{ij} = \sigma\left(\frac{\partial}{\partial q^i}, \frac{\partial}{\partial q^j}\right) = \sigma\left(\frac{\partial}{\partial \dot{q}^i}, \frac{\partial}{\partial \dot{q}^j}\right)$ and $\sigma\left(\frac{\partial}{\partial q^i}, \frac{\partial}{\partial \dot{q}^j}\right) = 0$, it follows that the matrix (σ_{IJ}) is of the form

$$(\sigma_{IJ}) = P \begin{pmatrix} (\mu_{ij}) & 0 \\ 0 & (\mu_{ij}) \end{pmatrix} P^{-1}$$

for a permutation matrix P. Therefore, $\sqrt{\det(\sigma_{IJ})} = \det(\mu_{ij})$ which proves (2.21).

By (2.21), condition (2.17) holds if and only if

$$\pounds_X((\mathcal{N} \circ \tau_M)\mathrm{vol}_\sigma) = 0 \iff \langle \mathbf{d}(\log \mathcal{N} \circ \tau_M), X\rangle + \mathrm{div}_{\mathrm{vol}_\sigma} X = 0. \tag{2.22}$$

Let us define

$$L(TM) := \{l \in C^\infty(TM) : l_x := l \mid T_xM : T_xM \to \mathbb{R} \text{ is linear for all } x \in M\}$$

and consider the prescription $\Phi : L(TM) \to \Omega^1(M)$, $\Phi(l)(u_x) = l_x(u_x)$, $u_x \in T_xM$, which is an isomorphism of $C^\infty(M)$-modules.

Lemma 2.4. *The following statements hold.*

(1) $\mathrm{div}_{\mathrm{vol}_\sigma} X \in L(TM)$.

(2) *Let* $\{u_i \mid i = 1, \dots, m\}$ *denote a local orthonormal frame on* M. *Then*

$$\begin{aligned} \mathrm{div}_{\mathrm{vol}_\sigma} X(u_x) &= -\sum_{i=1}^m \Xi(u_i(x))\Big(X(u_x), u_i^h(u_x)\Big) \\ &= -\sum_{i=1}^m \left\langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}{}_q)(u_i(x)), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\right\rangle. \end{aligned} \tag{2.23}$$

Proof. Clearly (2) implies (1) so we shall prove (2) below.

We use the Levi-Civita connection ∇^{μ_0} to split $TTM = \mathrm{Hor} \oplus \mathrm{Ver}$ where $\mathrm{Ver} = \ker T(\tau : TM \to M)$. Given a vector field X on M we shall, as before, denote its horizontal lift by $X^h \in \mathfrak{X}(TM, \mathrm{Hor})$ and its vertical lift by $X^v \in \mathfrak{X}(TM, \mathrm{Ver})$.

Let $\{u_i \in \mathfrak{X}(M) \mid i = 1, \dots, \dim M\}$, be a local orthonormal frame for TM. Then $\{(u_i^h, u_i^v) \mid i = 1, \dots, \dim M\}$ is a local orthonormal frame for TTM with respect to σ. By Lemma 2.3, if $X_0 = X_{\mathcal{H}_\mathrm{c}} = \Omega_M^{-1}(\mathbf{d}\mathcal{H}_\mathrm{c}) \in \mathfrak{X}(TM)$ is the standard Hamiltonian vector field, then we can locally express X as

$$X = X_0 - \sum \langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}{}_q)(u_x), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\rangle\, u_i^v(u_x).$$

Notice that X_0 preserves Ω^m whence $\mathrm{div}_{\mathrm{vol}_\sigma} X_0 = 0$. According to, e.g., [20, Proposition 7.2], it is true that $\nabla^\sigma_{u_i^h} u_i^h = (\nabla^{\mu_0}_{u_i} u_i)^h$ and $\nabla^\sigma_{u_i^v} u_i^v = 0$ where ∇^σ is the Levi-Civita

connection of σ. Therefore, sometimes suppressing the base point u_x for readability,

$$\begin{aligned}
\mathrm{div}_{\mathrm{vol}_\sigma} X = \mathrm{Tr}\,\nabla^\sigma_\cdot X &= \sum_{i=1}^m \big(\sigma(\nabla^\sigma_{u_i^h} X, u_i^h) + \sigma(\nabla^\sigma_{u_i^v} X, u_i^v)\big) \\
&= \sum_{i=1}^m \big(- \sigma(X, \nabla^\sigma_{u_i^h} u_i^h) + u_i^h \sigma(X, u_i^h) - \sigma(X, \nabla^\sigma_{u_i^v} u_i^v) + u_i^v \sigma(X, u_i^v)\big) \\
&= \sum_{i=1}^m \big(- \sigma(X_0, \nabla^\sigma_{u_i^h} u_i^h) + u_i^h \sigma(X_0, u_i^h) + u_i^v \sigma(X_0, u_i^v) \\
&\qquad - u_i^v \langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}{}_q)(u_x), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\rangle\big) \\
&= \mathrm{div}_{\mathrm{vol}_\sigma} X_0 - \sum_{i=1}^m \big\langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}})_q(u_i(x)), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u, u_i(x))\big\rangle \\
&= - \sum_{i=1}^m \big\langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}})_q(u_i(x)), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u, u_i(x))\big\rangle\,.
\end{aligned}$$

To see that the formula does not depend on the particular choice of $q \in \pi^{-1}(x)$ one uses equivariance of the involved expressions together with the observation that any G-ambiguity cancels out in the pairing. □

Therefore, $\mathrm{div}_{\mathrm{vol}_\sigma} X$ can be turned into a one-form on M through the canonical isomorphism $\Phi : L(TM) \to \Omega^1(M)$. Let us define

$$\beta := -\Phi\Big(\mathrm{div}_{\mathrm{vol}_\sigma} X\Big) \in \Omega^1(M) \tag{2.24}$$

for the non-holonomic vector field $X = X^{\mathrm{nh}}_{\mathcal{H}_c}$.

Proposition 2.5. *The system $(TM, \Omega_{\mathrm{nh}}, \mathcal{H}_c)$ admits a preserved measure if and only if $\beta \in \Omega^1(M)$ is exact. If $\beta = \mathbf{d}F$ for some function F on M then $\mathcal{N} = e^F$ is the density of the preserved measure for the Liouville volume.*

Proof. By Lemma 2.3, $\langle \mathbf{d}(\log \mathcal{N} \circ \tau_M), X\rangle(u_x) = \mathbf{d}(\log \mathcal{N})(u_x)$ for $u_x \in TM$ and hence $\Phi\Big(\langle \mathbf{d}(\log \mathcal{N} \circ \tau_M), X\rangle\Big) = \mathbf{d}(\log \mathcal{N})$. Now by (2.22) a preserved measure $(\mathcal{N} \circ \tau_M)\Omega^m_M$ exists if and only if

$$\mathbf{d}(\log \mathcal{N}) = \Phi\Big(\langle \mathbf{d}(\log \mathcal{N}), X\rangle\Big) = -\Phi(\mathrm{div}_{\mathrm{vol}_\sigma} X) = \beta,$$

i.e., $\beta \in \Omega^1(M)$ is exact. □

As stated above, this result is proved in [8, Theorem 7.5] but our interpretation of the form β is slightly different. The formula

$$\begin{aligned}
\beta(x)(u_x) &= \sum_{i=1}^m \Xi(u_i(x))\Big(X(u_x), u_i^h(u_x)\Big) \qquad (2.25) \\
&= \sum_{i=1}^m \big\langle (\mathbf{J}_G \circ \mathrm{hl}^{\mathcal{A}}{}_q)(u_i(x)), (\mathrm{Curv}^{\mathcal{A}}_q \circ \wedge^2 \mathrm{hl}^{\mathcal{A}}{}_q)(u_x, u_i(x))\big\rangle\,,
\end{aligned}$$

$u_x \in T_xM$, in a local orthonormal frame $\{u_1, \dots, u_m\}$ will be useful in Section 4 below.

3 Stochastic dynamics on manifolds

3.A Diffusions on manifolds

This subsection is a review of some necessary definitions and results which are all contained in the books [26, 18].

A diffusion is a continuous stochastic process which has the strong Markov property. This is a concept which can be formulated in any topological space.

Diffusion processes

Let X be a locally compact topological space with one-point compactification $\dot{X} = X \cup \{\infty\}$ and endow $\dot{X}$ with its Borel σ-algebra $\mathcal{B}(X)$. Define $W(X)$ to be the set of all maps $w : [0, \infty) \to \dot{X}$ such that there is a $\zeta(w) \in [0, \infty]$ satisfying

(1) $w(t) \in X$ for all $t \in [0, \zeta(w))$ and $w : [0, \zeta(w)) \to X$ is continuous;

(2) $w(t) = \infty$ for all $t \geq \zeta(w)$.

Let $l \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_l \in \mathbb{R}_+$, $A \subset \Pi_{i=1}^l \dot{X}$ a Borel set, and consider the evaluation mapping $\mathrm{ev}(t_1, \dots, t_l) : W(X) \to \Pi_{i=1}^l \dot{X}$, $w \mapsto (w(t_1), \dots, w(t_l))$. Then

$$S = \mathrm{ev}(t_1, \dots, t_l)^{-1}(A)$$

is called a *Borel cylinder set* in $W(X)$. If $t \geq 0$ and $t_l \leq t$ then S is a *Borel cylinder set up to time* t.

The set $W(X)$ is equipped with the σ-algebra $\mathcal{B}(W(X))$ generated by all Borel cylinder sets in $W(X)$. This σ-algebra has a natural filtration given by the family of

$$(\mathcal{B}_t(W(X)))_{t \geq 0}$$

which are the σ-algebras generated by Borel cylinder sets up to time t.

A family of probabilities $(P_x)_{x \in \dot{X}}$ on $(W(X), \mathcal{B}(W(X)))$ is said to be a *system of diffusion measures* on $(W(X), \mathcal{B}(W(X)), \mathcal{B}_t(W(X)))$ if it has the strong Markov property, the definition of which we will give shortly.

A $(\mathcal{B}_t(W(X)))_t$*-stopping time* is a random variable $\tau : W(X) \to \dot{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$ such that $\{w \in W(X) \mid \tau(w) \leq t\} \in \mathcal{B}_t(W(X))$ for all $t \in \mathbb{R}_+$.

For $s \in \mathbb{R}_+$ we define the *time shift operator*

$$\Sigma_s : W(X) \longrightarrow W(X), \quad w \longmapsto (\Sigma_s w : t \mapsto w(s + t)). \tag{3.1}$$

A family of probabilities $(P_x)_{x\in\dot{X}}$ on $(W(X),\mathcal{B}(W(X)))$ satisfies the *strong Markov property* if, for all $x\in X$, $(\mathcal{B}_t(W(X)))_t$-stopping times τ, bounded $\mathcal{B}_\tau(W(X))\times\mathcal{B}(W(X))$-measurable functions $F : W(X)\times W(X)\to\mathbb{R}$, and $s\in\mathbb{R}_+$, we have

$$\int_{\tau(w)<\infty} F(w,\Sigma_{\tau(w)}w)\,P_x(dw) = \int_{\tau(u)<\infty}\left(\int_{W(X)} F(u,w)\,P_{u(\tau(u))}(dw)\right)\,P_x(du). \tag{3.2}$$

See [41, p. 249].

Let $(\Omega,\mathcal{F},P)$ be a probability space and $\Gamma:\Omega\times\mathbb{R}_+\to\dot{X}$ a map. Then Γ is called a *stochastic process* if $\Gamma_t:\Omega\to\dot{X}$, $\omega\mapsto\Gamma_t(\omega)$, is a random variable for all $t\in\mathbb{R}_+$. Define $\check{\Gamma}:\omega\mapsto(t\mapsto\Gamma_t(\omega))$. Then Γ is said to be a *continuous stochastic process* in X if $\check{\Gamma}:(\Omega,\mathcal{F})\to(W(X),\mathcal{B}(W(X)))$ is a random variable. (Below, when X is a manifold, we will only be dealing with continuous processes.) Note that, for all $\omega\in\Omega$, the paths $[0,\zeta(\check{\Gamma}(\omega)))\ni t\mapsto\Gamma_t(\omega)\in X$ are continuous; the map $\zeta:W(X)\to[0,\infty]$ was part of the definition of $W(X)$.

The *law* of Γ is, by definition, the push-forward probability $\check{\Gamma}_*P$ on $(W(X),\mathcal{B}(W(X)))$, i.e., $\check{\Gamma}_*P(S)=P(\check{\Gamma}^{-1}(S))$ for all $S\in\mathcal{B}(W(X))$.

The process $\Gamma:\Omega\times\mathbb{R}_+\to\dot{X}$ defined on the probability space $(\Omega,\mathcal{F},P)$ is a *diffusion* in X if it is a continuous process and there is a system of diffusion measures $(P_x)_{x\in\dot{X}}$ such that $\check{\Gamma}_*P=P_\mu$ as probability laws on $(W(X),\mathcal{B}(W(X)))$; here

$$P_\mu(S) := \int_{\dot{X}} P_x(S)\mu(dx) \quad \text{for all} \quad S\in\mathcal{B}(W(X))$$

and $\mu=(\Gamma_0)_*P:\mathcal{B}(X)\to[0,1]$ is the initial distribution of Γ.

A diffusion Γ in X with associated system of diffusion measures $(P_x)_x$ is said to be *generated* by a linear operator A on the Banach space of continuous functions $C(\dot{X})$ with domain of definition $\mathcal{A}\subset C(\dot{X})$ if, for all $x\in X$, $t\geq 0$, and $f\in\mathcal{A}$, the stochastic process $M_t^f : W(X)\to\mathbb{R}$,

$$M_t^f(w) := f(w(t)) - f(w(0)) - \int_0^t (Af)(w(s))\,ds,$$

is a P_x-martingale on $(W(X),\mathcal{B}(W(X)))$ for the filtration $(\mathcal{B}_t(W(X)))_{t\geq 0}$. In this case, A is called the *generator* of Γ. See [26, Defs. IV.5.3 and IV.6.2]. The definition of a martingale is recalled below.

Let $(\Omega,\mathcal{F},P)$ be a probability space. A family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of sub-σ-algebras $\mathcal{F}_t\subset\mathcal{F}$ is called a *reference family* if it is increasing, i.e., $\mathcal{F}_t\subset\mathcal{F}_s$ for $0\leq t\leq s$, and right-continuous, i.e., $\cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}=\mathcal{F}_t$ for all $t\in\mathbb{R}_+$. Whenever we mention $(\mathcal{F}_t)$ we will suppress the index set $\mathbb{R}_+$, tacitly assume that it is a reference family, and refer to it as the filtration of $\mathcal{F}$ so that $(\Omega,\mathcal{F},(\mathcal{F}_t),P)$ becomes a filtered probability space.

A stochastic process $M:\Omega\times\mathbb{R}_+\to\dot{\mathbb{R}}_+$ is called a *martingale* on $(\Omega,\mathcal{F},(\mathcal{F}_t),P)$ if the following conditions are met:

(1) $M_t : \Omega \to \mathbb{R}$ is integrable for all $t \in \mathbb{R}_+$;

(2) $M_t : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{R}_+$, i.e., M is $(\mathcal{F}_t)$-*adapted*;

(3) $E[M_t|\mathcal{F}_s] = M_s$ for all $t \geq s \geq 0$, i.e., $E[(M_t - M_s)\chi_F] = 0$ for all $t \geq s \geq 0$ and all $F \in \mathcal{F}_s$, where χ_F is the characteristic function of the set F.

If $W : \Omega \times \mathbb{R}_+ \to \mathbb{R}^k$ is the (a fortiori continuous) diffusion defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ with initial condition $W_0 = 0$ a.s. and with generator $\frac{1}{2}\Delta = \frac{1}{2}\sum \partial_i \partial_i$, then W is called a$(\mathcal{F}_t)$-*adapted Brownian motion*. See [26, Example IV.5.2] or [41, Remark 7.1.23] for this characterization of Brownian motion. Below, we will be concerned with Brownian motion on a Riemannian manifold and then this aforementioned characterization will be taken to be the definition of Brownian motion.

Diffusions via Stratonovich equations

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space as above and suppose now that $X = Q$ is a manifold. From now on, all stochastic processes will be assumed to be continuous.

If N is manifold then a *Stratonovich operator* $\mathcal{S}$ from TN to TQ is a section of $T^*N \otimes TQ \to N \times Q$. Equivalently, we can view $\mathcal{S}$ as a smooth map $\mathcal{S} : Q \times TN \to TQ$ which is linear in the fibers and covers the identity on Q. Let $X_0, X_1, \ldots, X_k$ be vector fields on Q and define the associated Stratonovich operator $\mathcal{S} : Q \times T\mathbb{R}^{k+1} \longrightarrow TQ$ by

$$\mathcal{S}(x, w, w') := \sum_{i=0}^{k} X_i(x)\langle e_i, w' \rangle,$$

where $x \in Q$, $w \in \mathbb{R}^{k+1}$, $(w, w') \in T_w\mathbb{R}^{k+1} = \{w\} \times \mathbb{R}^{k+1}$, $\{e_i \mid i = 0, 1, \ldots, k\}$ is the orthonormal standard basis in $\mathbb{R}^{k+1}$, and $\langle\, , \rangle$ is the standard inner product in $\mathbb{R}^{k+1}$. We note that the number k is not related to the dimension of Q.

Consider the stochastic process $Y : \Omega \times \mathbb{R}_+ \to \mathbb{R}^{k+1}$, $(t, \omega) \mapsto (t, W_t(\omega))$ where W denotes $(\mathcal{F}_t)$-adapted Brownian motion in $\mathbb{R}^k$.

We will be concerned with Stratonovich equations of the form

$$\delta\Gamma = \mathcal{S}(Y, \Gamma)\delta Y; \tag{3.3}$$

a continuous $(\mathcal{F}_t)$-adapted process $\Gamma : \Omega \times \mathbb{R}_+ \to Q$ is called a *solution* to (3.3) if there is a $(\mathcal{F}_t)$-adapted Brownian motion $W = (W^i)$ in $\mathbb{R}^k$ such that, in the Stratonovich sense,

$$f(\Gamma_t) - f(\Gamma_0) = \int_0^t (X_0 f)(\Gamma_s)ds + \sum_{i=1}^{k} \int_0^t (X_i f)(\Gamma_s)\delta W_s^i \tag{3.4}$$

for all smooth functions $f \in C^\infty(Q)$ with compact support.

A few comments are in order. The definition of (3.3) is (3.4). The second integral in (3.4) is the Stratonovich integral (signaled by saying that the equation is to hold "in the Stratonovich sense"). We will not go into the elaborate definition and construction of the Stratonovich integral here and refer to [37, 26, 18] for the definition and an in depth study of the Stratonovich integral.

For the readers more familiar with Itô calculus we add the following remarks. Itô integrals are defined by a Riemann sum approximation with the rather essential difference that, in the sum, one evaluates the integrand *at the left end-points* of the partition intervals. The Stratonovich integral, on the other hand, can be obtained by evaluating the integrand *at the mid-points* of the sub-intervals. While Itô integrals give rise to a new transformation rule (the Itô formula), transformed Stratonovich integrals obey the same change of variables formula as Riemann integrals. This is the essential reason why, on manifolds, Itô calculus is replaced by Stratonovich calculus. Concretely, the Stratonovich integral is characterized in [26, Thm. III.1.4] by

$$\int_0^t Y\,\delta X = \text{l.i.p.}_{|\Lambda|\to 0} \sum_{i=1}^{n} \tfrac{Y(t_i)-Y(T_{i-1})}{2}(X(t_i)-X(t_{i-1}))$$

where Λ is a partition $0 = t_0 < t_1 < \ldots < t_n = t$ with maximal step size $|\Lambda|$ and l.i.p. is "limit in probability"; here X, Y are quasimartingales (a general class of semimartingales). We do not go into more details of the definition of the Stratonovich integral here and refer the reader to the above mentioned books.

Suppose Γ is a solution to (3.3) such that $\Gamma_0 = x$ a.s. and Γ satisfies (3.4) with respect to an $\mathbb{R}^k$-valued Brownian motion W defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. Then we will write $\Gamma = \Gamma^{x,W}$ to remember these data. The explosion time ζ of a solution $\Gamma^{x,W}$ is a stopping time on $(\Omega, \mathcal{F}, (\mathcal{F}_t))$ with the following property: the path $\Gamma^{x,W}_{[0,T]}(\omega)$ is contained in Q for all $T < \zeta(\omega)$ but if $\zeta(\omega) < \infty$ then $\Gamma^{x,W}_{[0,\zeta(\omega))}(\omega)$ is not contained in any compact subset of Q. The following is a partial account of [26, Theorems V.1.1 and V.1.2] and [18, Theorem (7.21)] that is sufficient for our purposes.

Theorem 3.1. *Let the assumptions be as above and consider equation* (3.3).

(1) *For each initial condition, $\Gamma_0 = x$ a.s., and continuous $(\mathcal{F}_t)$-adapted Brownian motion W, a solution $\Gamma^{x,W}$ exists and is unique up to explosion time.*

(2) *Let $P_x := \check{\Gamma}^{x,W}_* P$. Then P_x is independent of W and (P_x) is a system of diffusion measures generated by the second order differential operator*

$$A = X_0 + \tfrac{1}{2}\sum_{i=1}^{k} X_i x_i. \tag{3.5}$$

which acts on the space $C^\infty(Q)_0$ of smooth functions with compact support.

Assume that Q is endowed with a linear connection $\nabla : \mathfrak{X}(Q) \times \mathfrak{X}(Q) \to \mathfrak{X}(Q)$, where $\mathfrak{X}(Q)$ denotes the Lie algebra of smooth vector fields on Q.

If $f \in C^\infty(Q)$, its *Hessian* is defined by $\mathrm{Hess}(f) := \nabla\nabla f$, i.e., $\mathrm{Hess}(f)(X,Y) = X(Y(f)) - (\nabla_X Y)(f)$ for any $X, Y \in \mathfrak{X}(Q)$. The Hessian is bilinear in X and Y but is not symmetric, unless ∇ is torsion-free.

Let Γ be a diffusion in Q with generator A. Then the *drift* of Γ with respect to ∇ is defined to be the first order part of A which is determined by ∇. If A is of the form (3.5) then this is $X_0 + \frac{1}{2}\sum \nabla_{X_i} X_i$.

According to [18, Theorem 7.31], *the A-diffusion Γ is a martingale in (Q,∇) if and only if A is purely second order with respect to* ∇, i.e., the ∇-drift vanishes. In [18] this is stated for torsion-free connections but it is noted that one can use the same definition for connections with torsion.

If (Q,μ) is a Riemannian manifold then an A-diffusion is called *Brownian motion* if $A = \frac{1}{2}\Delta$ where $\Delta := \operatorname{div}\operatorname{grad} = -\delta\mathbf{d}$ is the metric Laplacian; see [26, Def. V.4.2] or [18, Def. 5.16].

To construct Brownian motion in (Q,μ), we need the principal connection

$$\omega \in \Omega^1\left(\mathfrak{F}; \mathfrak{so}(d)\right)$$

on the orthonormal frame bundle $\rho : \mathfrak{F} \to Q$ over (Q,μ), uniquely induced by the Levi-Civita connection ∇^μ on Q (whose Christoffel symbols in a chart are denoted by Γ^i_{kl}). We recall its construction and basic properties. Let $u \in \mathfrak{F}$ with base point $\rho(u) = q \in Q$. Then we define the horizontal bundle as $\mathrm{Hor}^\omega = \bigsqcup_{u\in\mathcal{F}} \mathrm{Hor}^\omega_u$, where the horizontal space at u, a vector subspace of $T_u\mathfrak{F}$, is given by

$$\mathrm{Hor}^\omega_u := T_q\sigma\left(T_qQ\right); \tag{3.6}$$

here σ is local section of $\rho : \mathfrak{F} \to Q$ such that $\sigma(q) = u$ and $\nabla^\mu_X \sigma_i = 0$ for all $X \in T_qQ$ and local vector fields $\sigma_i := \sigma(e_i)$ with $\{e_1,\dots,e_d\}$ being the standard basis in $\mathbb{R}^d$. We may express (3.6) in local coordinates (q^i, u^i_j) defined on a bundle coordinate patch $U \times V \hookrightarrow \mathfrak{F}$ as

$$\mathrm{Hor}^\omega_u = \left\{ A^k \frac{\partial}{\partial q^k} - \Gamma^i_{kl} u^l_j A^k \frac{\partial}{\partial u^i_j} \;\middle|\; A^k \in \mathbb{R} \right\}. \tag{3.7}$$

Restricting ω to this coordinate patch, it can be written as $\mathrm{pr}_2 + \omega_{\mathrm{loc}} : T(U\times V) \to \mathfrak{so}(d)$, where $\omega_{\mathrm{loc}} = (\omega^j_i) \in \Omega^1(U;\mathfrak{so}(d))$. It follows that, in terms of the Christoffel symbols,

$$\omega^i_j = \Gamma^i_{kl} u^l_j dq^k. \tag{3.8}$$

Thus, the local expression of the horizontal lift defined by ω is $\mathrm{hl}^\omega{}_u(\frac{\partial}{\partial q^k}) = \frac{\partial}{\partial q^k} - \Gamma^i_{kl} u^l_j \frac{\partial}{\partial u^i_j}$.

An orthonormal frame $u \in \mathfrak{F}$ can be regarded as an isometry $u : \mathbb{R}^d \to T_{\rho(u)}Q$, where $d = \dim Q$. Define the canonical horizontal vector fields $L_i \in \mathfrak{X}(\mathfrak{F}, \mathrm{Hor}^\omega)$, $i = 1,\dots,d$,

by

$$L_i(u) = \mathrm{hl}^{\omega}{}_u(u(e_i)), \tag{3.9}$$

where $\mathrm{hl}^{\omega} : \mathfrak{X}(Q) \to \mathfrak{X}(\mathfrak{F})$ is the horizontal lift map of ω. If (W^i) is Brownian motion in $\mathbb{R}^d$ and Γ solves the Stratonovich equation

$$\delta\Gamma = \sum L_i(\Gamma)\delta W^i \tag{3.10}$$

then $\rho \circ \Gamma$ is a diffusion in Q with generator $\frac{1}{2}\Delta^{\mu}$, that is, a Brownian motion. This is explained in [26, Chapter V.4] and follows also from Theorem 3.2 below; the essential observation in this context is that the Stratonovich operator $\mathcal{S}(u, w, w') = \mathcal{S}(u, w)w' = \sum L_i(u)\langle e_i, w'\rangle$ of (3.10) is equivariant:

$$\mathcal{S}(ug, g^{-1}w, g^{-1}w') = \sum \big(\mathrm{hl}^{\omega}{}_u(u(ge_i))\langle e_i, g^{-1}w'\rangle\big)g = \mathcal{S}(u, w, w')g$$

for the principal right action of the structure group, i.e., $g \in \mathrm{O}(d)$. Indeed, this follows since $\mathrm{hl}^{\omega} : \mathfrak{F} \times_Q TQ \to \ker\omega = \mathrm{Hor}^{\omega} \subset T\mathfrak{F}$ is $\mathrm{O}(d)$-equivariant.

To connect with Theorem 3.2, the principal right action can be turned to a left action via inversion in the group.

3.B Equivariant reduction

Equivariant reduction is a natural extension of the reduction theory of [33, Theorem 3.1]. While the results of [33] are stronger, in the sense that they provide a Stratonovich equation on the base space, they are only applicable when the original Stratonovich operator is G-invariant (i.e., equivariant with respect to the trivial action on the source space). By contrast, the observation in equivariant reduction is that although the upstairs Stratonovich operator is not projectable, the diffusion still factors to a diffusion in the base and the downstairs generator is induced from that of the original diffusion on the total space.

Two immediate examples are the construction of Brownian motion on a general Riemannian manifold as well as a stochastic version of Calogero-Moser systems (see below). For both of these cases, the diffusion upstairs is defined in terms of a Stratonovich operator which is equivariant but not projectable.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$, Q, $X_0, X_1, \ldots, X_k \in \mathfrak{X}(Q)$ and $\delta\Gamma = \mathcal{S}(Y, \Gamma)\delta Y$ be as before. Suppose there is a Lie group G which acts smoothly and properly on Q from the left. We continuously extend this action to the one point compactification $\dot{Q}$ by requiring ∞ to be a fixed point. Let $\pi : Q \twoheadrightarrow Q/G$ be the projection and $C^{\infty}(Q)^G$ denote the subspace of G-invariant smooth functions on Q. Note that Q/G need not be a manifold; in general Q/G is a topological space which is naturally stratified by smooth manifolds (see, e.g. [14, Chapter 2]).

In the following all actions are tangent lifted, where appropriate, without further notice. Generally, for Lie group actions, we will interchangeably use the notation $g \cdot q$ and gq.

Theorem 3.2. *Given is a group representation $\rho : G \to \mathrm{O}(k)$ and let $\mathrm{O}(k)$ act on $\mathbb{R}^{k+1} = \mathbb{R} \times \mathbb{R}^k$ such that the first factor is acted upon trivially. If the Stratonovich operator $\mathcal{S}$ satisfies the equivariance property*

$$\mathcal{S}(gx, \rho(g)y, \rho(g)y') = g\mathcal{S}(x, y, y') \tag{3.11}$$

for all $(x, y, y') \in Q \times T\mathbb{R}^{k+1}$, then the diffusion Γ induces a diffusion $\pi \circ \Gamma$ in Q/G. Moreover, the generator A of the diffusion Γ on Q preserves $C^\infty(Q)^G$ and the induced generator A_0 of the diffusion $\pi \circ \Gamma$ on Q/G is characterized by

$$\pi^*(A_0 f) = A(\pi^* f) \tag{3.12}$$

for all $f \in C^\infty(Q/G) := \{f \in C(Q/G) : \pi^* f \in C^\infty(Q)^G\}$.

Proof. Let us begin by noting that $g\Gamma^{x,W} = \Gamma^{gx,\rho(g)W}$. Indeed,

$$\delta(g\Gamma^{x,W}) = g\mathcal{S}(Y, \Gamma^{x,W})\delta Y = \mathcal{S}(\rho(g)Y, g\Gamma^{x,W})\delta(\rho(g)Y)$$

whence $\tilde{\Gamma} := g\Gamma^{x,W}$ satisfies $\tilde{\Gamma}_0 = gx$ a.s. and $\delta\tilde{\Gamma} = \mathcal{S}(\rho(g)Y, \tilde{\Gamma})\delta(\rho(g)Y)$. By existence and uniqueness of solutions the claim follows. In particular, we have $\pi \circ \Gamma^{x,W} = \pi \circ \Gamma^{gx,\rho(g)W}$.

Claim:

$$P_{gx} = g_* P_x \tag{3.13}$$

where G acts on $W(Q)$ as $g : w \mapsto (t \mapsto gw(t))$. To see this, let $S \subset W(Q)$ be a Borel cylinder set. This means that there are $l \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_l \in \mathbb{R}_+$, and a Borel set $A \subset \Pi^l \dot{Q}$ such that $S = \mathrm{ev}(t_1, \ldots, t_l)^{-1}(A)$, where $\mathrm{ev}(t_1, \ldots, t_l) : W(Q) \to \Pi^l \dot{Q}$, $w \mapsto (w(t_i))_{i=1}^l$. From the identity $(\Gamma^{gx,\rho(g)W})^{\check{}}_* P = (\Gamma^{x,W})^{\check{}}_* P$ we find

$$\begin{aligned} P_{gx}(S) &= (\Gamma^{gx,\rho(g)W})^{\check{}}_* P(S) = P\{\omega : (\Gamma^{gx,\rho(g)W}_{t_i}(\omega))_{i=1}^l \in A\} \\ &= P_x(\mathrm{ev}(t_1, \ldots, t_l)^{-1}(g^{-1}A)) = P_x(g^{-1}S) \end{aligned}$$

which proves (3.13).

Consider the push forward map $\pi_* : W(Q) \to W(Q/G)$, $w \mapsto \pi \circ w$. It is straightforward to see that $\mathcal{B}(W(Q/G)) = \pi_* \mathcal{B}(W(Q))$. For $S_0 = \pi_*(S) \in \mathcal{B}(W(Q/G))$ we may write the law $\left(P_{[x]}\right)_{[x] \in \dot{Q}/G}$ of $\pi \circ \Gamma$ as

$$P_{[x]}(S_0) = (\pi \circ \Gamma^{gx,\rho(g)W})^{\check{}}_* P(S_0) = P_{gx}(\pi_*^{-1}(S_0)).$$

By (3.13) this does not depend on $g \in G$.

Let us show that the system $\big(P_{[x]}\big)_{[x]\in\dot{Q}/G}$ satisfies the strong Markov property. Let $p := \pi_* : W(Q) \to W(Q/G)$, $[x] \in Q/G$, $\tau : W(Q/G) \to \mathbb{R}_+$ be a $(\mathcal{B}_t(W(Q/G)))_t$-stopping time, and $F : W(Q/G) \times W(Q/G) \to \mathbb{R}$ a bounded $\mathcal{B}_\tau(W(Q/G)) \times \mathcal{B}(W(Q/G))$ measurable function. Then $p^{-1}(\mathcal{B}_t(W(Q/G))) \subset \mathcal{B}_t(W(Q))$ and $p^*\tau = \tau \circ p : W(Q) \to \mathbb{R}_+$ is a $(\mathcal{B}_t(W(Q)))_t$-stopping time. For $s \in \mathbb{R}_+$ let Σ_s be the time shift operator defined in (3.1) and observe that $(\Sigma_s pw)(t) = pw(s+t) = \pi(w(s+t)) = (p\Sigma_s w)(t)$. (We use the same notation for the time-shift on $W(Q)$ and that on $W(Q/G)$.) Now, since $P_{[x]}$ is the push forward of P_x via p, we can use the strong Markov property of $(P_x)_x$ to conclude that

$$\begin{aligned}
&\int_{\{w\in W(Q/G):\tau(w)<\infty\}} F(w, \Sigma_{\tau(w)}w)\, P_{[x]}(dw) \\
&\qquad = \int_{\{u\in W(Q):p^*\tau(u)<\infty\}} p^*F(u, \Sigma_{p^*\tau(u)}u)\, P_x(du) \\
&\qquad = \int_{\{p^*\tau(u)<\infty\}} \Big(\int_{W(Q)} p^*F(u,v)\, P_{u(p^*\tau(u))}(dv)\Big) P_x(du) \\
&\qquad = \int_{\{p^*\tau(u)<\infty\}} \Big(\int_{W(Q/G)} F(pu,w)\, P_{[u(p^*\tau(u))]}(dw)\Big) P_x(du) \\
&\qquad = \int_{\{\tau(u_0)<\infty\}} \Big(\int_{W(Q/G)} F(u_0,w)\, P_{u_0(\tau(u_0))}(dw)\Big) P_{[x]}(du_0)
\end{aligned}$$

which, according to (3.2), shows that $(P_{[x]})_{[x]}$ is strong Markov.

To show that $\sum X_iX_if \in C^\infty(Q)^G$ for all $f \in C^\infty(Q)^G$ consider the standard basis $\{e_0, e_1, \dots, e_k\}$ of $\mathbb{R}\times\mathbb{R}^k$. For $j = 1, \dots, k$ we find

$$g \cdot X_j(x) = g \cdot \mathcal{S}(x, y, e_j) = \mathcal{S}(gx, \rho(g)y, \rho(g)e_j) = \sum_k g_{kj} X_k(gx),$$

where $g_{kj} := \langle e_k, \rho(g)e_j\rangle$ is independent of $x \in Q$. Since $\sum_j g_{ij}g_{kj} = \delta_{ik}$,

$$X_i(gx) = \sum_{j,k} g_{ij}g_{kj}X_k(gx) = \sum_j g_{ij}\, g\cdot X_j(x).$$

Thus $\Big(df(X_i)\Big)(gx) = \sum_j g_{ij}\Big(df(X_j)\Big)(x)$ for $f \in C^\infty(Q)^G$ and also

$$d\Big(df(X_i)\Big)(gx)\circ T_xg = d\Big(\sum_j g_{ij}df(X_j)\Big)(x) = \sum_j g_{ij}d\Big(df(X_j)\Big)(x).$$

This implies that

$$\begin{aligned}
\sum_i \Big(X_iX_if\Big)(gx) &= \sum_i \Big\langle d\Big(df(X_i)\Big)(gx), X_i(gx)\Big\rangle \\
&= \sum_{i,j,k} \Big\langle g_{ij}d\Big(df(X_j)\Big)(x)\circ(T_xg)^{-1}, g_{ik}(T_xg)\cdot X_k(x)\Big\rangle \\
&= \sum_i \Big(X_iX_if\Big)(x).
\end{aligned}$$

Similarly, it is also easy to see that X_0 is G-invariant. Thus the generator $A = X_0 + \frac{1}{2}\sum X_i X_i$ acts on $C^\infty(Q)^G$, whence it induces a projected operator A_0 characterized by $A \circ \pi^* = \pi^* \circ A_0$.

Finally, to see that A_0 is the generator of $\pi \circ \Gamma$ we need to show that, for all $t \in \mathbb{R}_+$, $[x] \in Q/G$, and $f \in C^\infty(Q/G)_0$, the $\mathbb{R}$-valued process

$$\begin{aligned} &M_t^f : W(Q/G) \longrightarrow \mathbb{R}, \\ &M_t^f(w) := f(w(t)) - f(w(0)) - \int_0^t (A_0 f)(w(s))\, ds \end{aligned}$$

is a $P_{[x]}$-martingale on $(W(Q/G), \mathcal{B}(W(Q/G))$ for the filtration $(\mathcal{B}_t(W(Q/G)))_t$. See [26, Def. IV.5.3]. This means that for all $t \geq 0$, $s \in [0,t]$, and $A \in \mathcal{B}_s(W(Q/G))$ we should check that (see [41, Chapter V])

$$\int_A E^{P_{[x]}}\Big[M_t^f \Big| \mathcal{B}_s(W(Q/G))\Big](w)\, P_{[x]}(dw) = \int_A M_s^f(w)\, P_{[x]}(dw);$$

$E^{P_{[x]}}$ denotes the expectation on $\Big(W(Q/G), \mathcal{B}(W(Q/G))\Big)$ with respect to $P_{[x]}$. Indeed,

$$\begin{aligned} \int_A E^{P_{[x]}}\Big[M_t^f \Big| \mathcal{B}_s(W(Q/G))\Big](w)\, P_{[x]}(dw) &= \int_A M_t^f(w)\, P_{[x]}(dw) \\ &= \int_{p^{-1}A} (p^* M_t^f)(u)\, P_x(du) \\ &= \int_{p^{-1}A} (\hat{M}_t^{\pi^* f})(u)\, P_x(du) \\ &= \int_{p^{-1}A} E^{P_x}\Big[\hat{M}_t^{\pi^* f} \Big| \mathcal{B}_s(W(Q))\Big](u)\, P_x(du) \\ &= \int_{p^{-1}A} \hat{M}_s^{\pi^* f}(u)\, P_x(du) \\ &= \int_{p^{-1}A} (p^* M_s^f)(u)\, P_x(du) \\ &= \int_A M_s^f(w)\, P_{[x]}(dw). \end{aligned}$$

Here, $\hat{M}_t^{\pi^* f} : W(Q) \to \mathbb{R}$ is analogously defined to M_t^f. We have used that $\hat{M}_t^{\pi^* f}$ is a P_x-martingale with respect to $(\mathcal{B}_t(W(Q)))_t$ for all $x \in Q$ and that $p^* M_t^f = \hat{M}_t^{\pi^* f}$ which holds because of $(A_0 f) \circ \pi = A(\pi^* f)$. □

Stochastic Calogero-Moser systems

To construct classical trigonometric or rational Calogero-Moser models one can take the configurations space Q to be a (real or complex) semisimple Lie group G or a semisimple Lie algebra $\mathfrak{g}$, respectively. The metric μ on Q is then accordingly given by the

(essentially unique) bi-invariant (pseudo-)metric in the group or the Ad-invariant non-degenerate bilinear form in the Lie algebra. Thus one obtains a G-invariant Hamiltonian system $(T^*Q, \Omega^Q, \mathcal{H})$ where Ω^Q is the canonical symplectic form on T^*Q, $\mathcal{H}$ is the kinetic energy Hamiltonian, and G acts by the cotangent lift of the conjugation action or the adjoint action, respectively. The resulting Calogero-Moser system is then realized by passing to the (singular) symplectic quotient of $(T^*Q, \Omega^Q, \mathcal{H})$ with respect to the G-action. See [28, 19, 21]. In other words, Calogero-Moser systems are obtained by reducing the Hamiltonian description of geodesic motion on the Riemannian manifold (Q, μ) with respect to its obvious symmetry group.

Here we propose the stochastic analogue of this construction which should consist of reducing the Hamiltonian construction of Brownian motion on (Q, μ) with respect to the G-action. To this end, we consider the Hamiltonian version in [32] of (3.10). Using the left trivialization we may write $TQ = Q \times \mathfrak{g}$ (recall that $Q = G$ or $Q = \mathfrak{g}$) and choose an orthonormal basis L_i of $\mathfrak{g}$ with respect to the Ad-invariant inner product $\langle \cdot, \cdot \rangle$; suppose from now on, for simplicity of exposition, that G is compact. We obtain a $\mathfrak{g}$-valued Hamiltonian

$$H = (H^i) : T^*Q = Q \times \mathfrak{g}^* \longrightarrow \mathfrak{g}, \quad (q, p) \longmapsto \sum \langle p, L_i \rangle L_i.$$

The Hamiltonian version of Brownian motion is determined by the associated Stratonovich equation

$$\delta\Gamma = \sum_i X_{H^i}(\Gamma) \langle L_i, \delta W \rangle,$$

where W is Brownian motion in $\mathfrak{g} \cong \mathbb{R}^n$. It is shown in [32] that $\tau \circ \Gamma$ is Brownian motion in (Q, μ), where $\tau : T^*Q \to Q$ is the projection. In the left trivialization $T^*Q = Q \times \mathfrak{g}^*$ the Hamiltonian H is nothing but the projection onto the second factor when $\mathfrak{g}$ and $\mathfrak{g}^*$ are identified. Clearly, H is not G-invariant but it is G-equivariant for the Ad-action on $\mathfrak{g}$. It is easy to see that the same is true for the Stratonovich operator $(q, p; w, w') \mapsto \sum X_{H^i}(q, p) \langle L_i, w' \rangle$. In fact, we are ultimately concerned with the Stratonovich equation $\delta(\tau \circ \Gamma) = \mathcal{S}(W, \Gamma) \delta W = \sum \delta W^i L_i = \delta W$ and now it is evident that $\mathcal{S}(g \cdot q, g \cdot (w, w')) = \mathrm{Ad}(g) w'$ whence we need the $\mathrm{Ad}(G)$-action on (w, w') to make the Stratonovich operator $\mathcal{S} : Q \times T\mathfrak{g} \to TQ$ equivariant for the respective actions. Thus the above theorem applies and we obtain a diffusion $\pi \circ \tau \circ \Gamma$ in the (singular) space Q/G when $\pi : Q \twoheadrightarrow Q/G$ is the projection.

This construction has been carried out in [24] where it is shown that the associated stochastic Hamilton-Jacobi equation of [34] is related to the quantum Calogero-Moser Schrödinger equation of [38, 39].

The issue of equivariant reduction leads immediately to the setting of [16, 17]. There, one of the topics treated is that of a diffusion on the total space of a principal bundle such

that the diffusion factors through the projection and the generator induces also a generator on the base.

3.C Reconstruction of an equivariant diffusion

Proposition 3.3 that we shall state and prove in this subsection will be used in the examples considered later on. Before we can state it, we need to recall some notions of [16, 17].

Let $\pi : Q \twoheadrightarrow Q/G =: M$ be a left G-principal bundle with connection form $\mathcal{A} \in \Omega^1(Q;\mathfrak{g})$. Denote the horizontal and vertical spaces by Hor and Ver, respectively. Assume that Γ is a diffusion in Q generated by a Stratonovich equation

$$\delta\Gamma = X_0(\Gamma)\delta t + \sum_{a=1}^{m} X_a(\Gamma)\delta W^a + v_0(\Gamma)\delta t + \sum_{\alpha=1}^{k} v_\alpha(\Gamma)\delta B^\alpha \tag{3.14}$$

such that $X_0, X_1, \dots, X_m$ are basic, $v_0, v_1, \dots, v_k$ are vertical vector fields, and (W, B) is Brownian motion in $\mathbb{R}^{m+k}$ with respect to the underlying filtered probability space. Thus the generator of Γ is

$$A^Q = X_0 + \tfrac{1}{2}\sum X_a X_a + v_0 + \tfrac{1}{2}\sum v_\alpha v_\alpha.$$

By construction, this generator can be decomposed as follows. There are $Y_0, Y_1, \dots, Y_m \in \mathfrak{X}(M)$ such that $X_0 = \mathrm{hl}^{\mathcal{A}}(Y_0), \dots, X_m = \mathrm{hl}^{\mathcal{A}}(Y_m)$ and $x_t := \pi \circ \Gamma_t$ is a diffusion in M with generator $A^M = Y_0 + \frac{1}{2}\sum Y_a Y_a$. Note that $\pi^* \circ A^M = A^Q \circ \pi^*$, that is, A^Q is projectable. Moreover, A^Q decomposes into a horizontal part $A^h = X_0 + \frac{1}{2}\sum X_a X_a$ and a vertical part $A^v = v_0(\Gamma)\delta t + \sum_{\alpha=1}^{k} v_\alpha \delta B^\alpha$.

In [16, 17] one of the main points is that, assuming a non-degeneracy condition, the induced operator A^M gives rise to a connection in $\pi : Q \to M$ with respect to which the operator A^Q can be decomposed. In our applications the connection is given by the problem and the decomposition into horizontal and vertical part arises naturally.

We are going to use the observation of [16, 17] that, for $q \in Q$ and $\Gamma_0 = q$ a.s., the diffusion Γ can be written as

$$\Gamma_t = g_t^{x^h} \cdot x_t^h. \tag{3.15}$$

Here x_t^h is the diffusion in Q with generator A^h and $x_0^h = q$ a.s. That is, x_t^h is the horizontal lift of the A^M diffusion x_t. The process $g_t^{x^h}$ in G with $g_0^{x^h} = e$ a.s. can be written as the solution to a time-dependent Stratonovich equation: for $w \in W(Q)$ we define

$$\delta g_t^w = T_e R_{g_t^w}\Big(\mathcal{A}_{g_t^w \cdot w_t} v_0(g_t^w \cdot w_t) + \sum \mathcal{A}_{g_t^w \cdot w_t} v_\alpha(g_t^w \cdot x_t^w)\Big). \tag{3.16}$$

Here $R_g : G \to G$ is the action by right multiplication of G on itself. Equation (3.15) is reminiscent of a well-known concept in mechanics and can be viewed as a reconstruction equation (see, e.g., [1, §4.3], [35, §3], [36, Theorem 11.8]).

Let $Q_e(w)$ be the law on $W(G)$ of (3.16). This depends only on $\pi \circ w \in W(M)$. Let x_t be an A^M-diffusion path in M with horizontal lift x^h. Consider the evaluation map $\mathrm{ev}_t : W(G) \to G$, $g(\cdot) \mapsto g(t)$. We call

$$E^{Q_e(x^h)}[\mathrm{ev}] \cdot x^h : t \longmapsto E^{Q_e(x^h)}[\mathrm{ev}_t] \cdot x^h_t$$

the *mean reconstruction* of the sample path x_t.

From now on, we shall assume that G can be realized as a matrix group $G \subset \mathrm{GL}(N) \subset \mathbb{R}^{N^2}$. The flat connection on $\mathbb{R}^{N^2}$ thus induces a connection ∇ on G. For $X \in \mathfrak{g} \subset \mathfrak{gl}(N)$ we denote the associated left- and right-invariant vector field by $L_*X(g) = T_eL_g(X) = gX \in \mathfrak{gl}(N)$ and $R_*X(g) = T_eR_g(X) = Xg \in \mathfrak{gl}(N)$.

Proposition 3.3. *If $x \in W(M)$ is an A^M-sample path and $v_0, \dots, v_k$ are G-invariant vector fields then the expectation $E^{Q_e(x^h)}[\mathrm{ev}_t] =: c(t)$ associated to the mean reconstruction of an A^M-diffusion path x in M is given as the solution to the left-invariant time-dependent ODE*

$$T_eL^{-1}_{c(t)}\left(c'(t)\right) = \mathcal{A}_{x^h_t}v_0(x^h_t) + \tfrac{1}{2}\sum_{\alpha=1}^{k} \nabla_{L_*\left(\mathcal{A}_{x^h_t}v_\alpha(x^h_t)\right)} L_*(\mathcal{A}_{x^h_t}v_\alpha(x^h_t))(e).$$

Proof. We can use G-invariance of the vector fields together with the equivariance property $\mathcal{A}_{gq}(gu_q) = \mathrm{Ad}(g)\mathcal{A}_q u_q$, $q \in Q$, $u_q \in T_qQ$, of the principal bundle connection form $\mathcal{A}$ to rewrite the defining equation (3.16) as

$$\delta g^{x^h}_t = T_eR_{g^{x^h}_t}\left(T_e(R^{-1}_{g^{x^h}_t} \circ L_{g^{x^h}_t})\left(\mathcal{A}_{x^h_t}\left(v_0(x^h_t)\delta t + \sum v_\alpha(x^h_t)\delta B^\alpha\right)\right)\right).$$

Letting

$$\begin{aligned} g(t) &= (g(t)^l_n)_{nl} := g^{x^h}_t, \\ a(t)_\alpha &= (a(t)^{mn}_\alpha)_{mn} := \mathcal{A}_{x^h_t}v_\alpha(x^h_t) \in \mathfrak{gl}(N), \\ b(t) &= (b(t)^{mn})_{mn} := \mathcal{A}_{x^h_t}v_0(x^h_t) \in \mathfrak{gl}(N), \end{aligned}$$

this becomes with the summation convention, for $l = 1, \dots, N$, the Stratonovich equation

$$\delta g^l = \left(g^l_m a^{mn}_\alpha \delta B^\alpha + g^l_m b^{mn}\delta t\right)^N_{n=1}$$

in $\mathbb{R}^N$ when we think of g^l as a column vector and suppress the time-dependency. The associated Itô equation in $\mathbb{R}^N$ is, for $l = 1, \dots, N$,

$$dg^l = \left(g^l_m a^{mn}_\alpha dB^\alpha + (g^l_m b^{mn} + \tfrac{1}{2}a^{rn}_\alpha g^l_m a^{mr}_\alpha)dt\right)^N_{n=1}.$$

(See e.g. [37, Equ. (6.1.3)] for the conversion rule of Stratonovich equations to Itô equations.) This is a linear time-dependent Itô equation in $\mathbb{R}^N$. Hence, the mean motion is

found by erasing the martingale term in the corresponding integral equation. This implies that the expected motion of g is given by

$$c'(t) = \tfrac{d}{dt}E[g](t) = E[g](t)\left(b(t) + \tfrac{1}{2}a(t)_\alpha a(t)_\alpha\right)$$

which is an equation in $\mathrm{GL}(N)$. Since $a(t)_\alpha a(t)_\alpha = \nabla_{L_* a(t)_\alpha}(L_* a(t)_\alpha)(e)$ the claim follows. □

3.D Time reversible diffusions

The references for this section are [31, 29, 26]. Let (M, μ) be a Riemannian manifold and Γ an A-diffusion in M where

$$A = \tfrac{1}{2}\Delta + \tfrac{1}{2}b \tag{3.17}$$

with Δ the Laplace-Beltrami operator and b a vector field. Let $p(t, x, y)$ denote the transition probability density of Γ (the minimal fundamental solution – see [29]). If vol_μ is the Riemannian volume form on M then, for $(t, x, S) \in \mathbb{R}_+ \times M \times \mathcal{B}(M)$, the transition probability of Γ is

$$P(t, x, S) = \int_S p(t, x, y)\mathrm{vol}_\mu(y).$$

This quantifies the probability that a diffusion path starting at x is in S after time t. The diffusion Γ is said to be *symmetrizable* if there is a smooth function $\phi > 0$ such that

$$p(t, x, y)\phi(x) = p(t, y, x)\phi(y) \quad \text{for all } t, x, y \in \mathbb{R}_+ \times M \times M \tag{3.18}$$

in which case Γ is called ϕ-*symmetric*.

A probability measure ν on M is an *equilibrium measure* if $\int_M \nu = 1$ and

$$P(t, x, S) \longrightarrow \nu(S) \text{ as } t \longrightarrow \infty$$

for all $(x, S) \in M \times \mathcal{B}(M)$. Equilibrium measures, if they exist, are unique. If $\nu = \phi\mathrm{vol}_\mu$ is an equilibrium measure then we refer to ϕ as the *equilibrium distribution*.

The diffusion Γ is called *time-reversible* if its law coincides with that of the time-reversed process; this means that for each $T > 0$ the law $P_{[0,T]}$ of $[0, T] \times \Omega \to M$, $(t, \omega) \mapsto \Gamma_t(\omega)$ is the same as the law $P^-_{[0,T]}$ of $[0, T] \times \Omega \to M$, $(t, \omega) \mapsto \Gamma_{T-t}(\omega)$.

The adjoint operator A^* associated to A is given by

$$A^* f = \tfrac{1}{2}\Delta f - \tfrac{1}{2}\mathrm{div}_\mu(fb)$$

where $f \in C^\infty(M)$. Here, the adjoint is with respect to the L^2 inner product $\langle f, g\rangle = \int_M fg\,\mathrm{vol}_\mu$. The following result is essentially due to Kolmogorov.

Theorem 3.4 ([31, 29, 26])**.** *With notation as above the following are true.*

(1) *The A-diffusion Γ is symmetric if and only if b is a gradient. Moreover, if $b = \mathrm{grad}(\log\phi)$ then Γ is ϕ-symmetric and $A^*\phi = 0$.*

(2) *Γ is time reversible if and only if it is symmetric and has an equilibrium distribution ϕ, in which case Γ is ϕ-symmetric.*

(3) *If M is compact then an equilibrium distribution always exists.*

(4) *If M is compact then the unique equilibrium distribution ϕ is characterized by the equations $\int_M \phi \mathrm{vol}_\mu = 1$ and $A^*\phi = 0$.*

Compactness of M is satisfied in important examples such the Chaplygin ball or the two-wheeled carriage studied in Section 5.

Assuming that M is compact, [27, Chapter 5] give various equivalent conditions for a diffusion of the form (3.17) to be time-reversible. One such condition is that the diffusion have vanishing entropy production rate

$$\lim_{T\to 0} \tfrac{1}{T} H(P_{[0,T]}, P^-_{[0,T]}). \tag{3.19}$$

The *relative entropy* $H(\mu,\nu)$ of two probability measures μ,ν on a measure space $(\mathcal{W},\mathcal{B})$ is defined as (see [27, Definition 1.4.3])

$$H(\mu,\nu) := \begin{cases} \int_\Omega \log \frac{d\mu}{d\nu}\mu(d\omega) & \text{if } \mu << \nu \text{ and } \log\frac{d\mu}{d\nu} \in L^1(d\mu); \\ +\infty & \text{otherwise.} \end{cases}$$

4 Non-holonomic diffusions

Consider a non-holonomic system $(Q,\mathcal{D},\mathcal{L})$ as in Section 2. This section is concerned with the study of non-holonomic diffusions on $\mathcal{D}$ which should be given by a Stratonovich equation of the form

$$\delta\Gamma = X^{\mathcal{C}}_{\mathcal{H}}\delta t + \mathcal{S}^{\mathcal{C}}(\Gamma, W)\delta W. \tag{4.1}$$

Here $X^{\mathcal{C}}_{\mathcal{H}}$ describes the dynamics of the deterministic system, W is Brownian motion in $\mathbb{R}^d$, $d = \dim Q$, and $\mathcal{S}^{\mathcal{C}}(\Gamma,W)\delta W$ should be interpreted as a noise term that stems from *constrained Brownian motion*. This is in analogy to [32, Section 3.1] and [5] where Hamiltonian diffusions are introduced. However, equation (4.1) does not make sense, in general, unless the configuration space is parallelizable. The problem to be considered below is to make this equation precise and to study the notion of constrained Brownian motion on manifolds.

4.A Constrained Brownian motion

Let $(Q, \mathcal{D}, \mathcal{L})$ be a non-holonomic system with symmetry group G as in Section 2.

Let $\rho : \mathfrak{F} \to Q$ be the orthonormal frame bundle over (Q, μ) and denote its structure group by $K := \mathrm{O}(d)$ and its Lie algebra by $\mathfrak{k} := \mathfrak{so}(d)$. The Levi-Civita connection ∇^μ on TQ gives rise to a uniquely determined principal connection $\omega \in \Omega^1(\mathfrak{F}; \mathfrak{k})$ on the principal bundle $\rho : \mathfrak{F} \to Q$; denote by $\mathrm{Hor}^\omega = \ker \omega \subset T\mathfrak{F}$ its horizontal subbundle. Equip $\mathfrak{F}$ with a K-invariant metric ν such that ρ becomes a Riemannian submersion and Hor^ω and $\mathrm{Ver}(\rho) := \ker T\rho$ are perpendicular. Since G acts by isometries on (Q, μ), it lifts to an action on $\mathfrak{F}$ and we may assume ν to be G-invariant; e.g., we could take ν to be the Sasaki-Mok metric (see the survey [40]). We can use the connection to lift the constraints to a subbundle $\mathcal{D}\mathfrak{F} \subset T\mathfrak{F}$ defined via the natural ω-dependent vector bundle isomorphism

$$\mathcal{D}\mathfrak{F} \cong_\omega (\mathfrak{F} \times_Q \mathcal{D}) \oplus \mathrm{Ver}(\rho).$$

To understand this definition and the isomorphism consider the bundle morphism over $\mathfrak{F}$ defined by

$$\begin{aligned} \mathfrak{F} \times_Q \mathcal{D} \oplus \mathrm{Ver}(\rho) &\hookrightarrow \mathrm{Hor}^\omega \oplus \mathrm{Ver}(\rho) \cong_\omega T\mathfrak{F}, \\ (u_q, X_q; \eta_{u_q}) &\longmapsto \mathrm{hl}^\omega_{u_q}(X_q) + \eta(u_q). \end{aligned}$$

As before, $\mathrm{hl}^\omega : \mathfrak{F} \times_Q TQ \to \mathrm{Hor}^\omega$ is the horizontal lift mapping associated to ω. Now the subbundle $\mathcal{D}\mathfrak{F}$ is defined as the image of this morphism.

Thus $(\mathfrak{F}, \mathcal{D}\mathfrak{F}, \frac{1}{2}||\cdot||_\nu)$ is a new G-invariant non-holonomic system covering $(Q, \mathcal{D}, \mathcal{L})$ in the following sense: The G-action lifts to an action on $\mathcal{D}\mathfrak{F}$ and there is an induced space $\mathcal{C}\mathfrak{F}$ defined by

$$\mathcal{C}\mathfrak{F} := \{\xi \in T(\mathcal{D}\mathfrak{F}) \mid T\tau_{\mathfrak{F}}(\xi) \in \mathcal{D}\mathfrak{F}\},$$

where $\tau_{\mathfrak{F}} : T\mathfrak{F} \to \mathfrak{F}$ is the tangent bundle projection. Again, we split $T(T\mathfrak{F})|(\mathcal{D}\mathfrak{F}) = \mathcal{C}\mathfrak{F} \oplus (\mathcal{C}\mathfrak{F})^{\Omega_{\mathfrak{F}}}$, where $\Omega_{\mathfrak{F}}$ is now the canonical symplectic form on $T\mathfrak{F} \cong_\nu T^*\mathfrak{F}$ (the tangent and cotangent bundles of $\mathfrak{F}$ are identified via the Riemannian metric ν on Q), and $P_{\mathfrak{F}} : \mathcal{C}\mathfrak{F} \oplus (\mathcal{C}\mathfrak{F})^{\Omega_{\mathfrak{F}}} \to \mathcal{C}\mathfrak{F}$ denotes the associated projection. The situation is summarized in the following commutative diagram:

$$\begin{array}{ccccc} T(T\mathfrak{F})|(\mathcal{D}\mathfrak{F}) & = & \mathcal{C}\mathfrak{F} \oplus (\mathcal{C}\mathfrak{F})^{\Omega_{\mathfrak{F}}} & \xrightarrow{P_{\mathfrak{F}}} & \mathcal{C}\mathfrak{F} \\ \downarrow{\scriptstyle TT\rho} & & \downarrow & & \downarrow \\ TTQ|\mathcal{D} & = & \mathcal{C} \oplus \mathcal{C}^\Omega & \xrightarrow{P} & \mathcal{C}. \end{array} \tag{4.2}$$

Indeed, this diagram is commutative since $TT\rho(\mathcal{C}\mathfrak{F}) = \mathcal{C}$ and we may regard (TQ, Ω) as the symplectic reduction of $(T\mathfrak{F}, \Omega_{\mathfrak{F}})$ with respect to the K-action at 0. In particular,

$$TT\rho\left(P_{\mathfrak{F}}\left(X_{\rho^* f}\right)\right) = P\left(TT\rho\left(X_{\rho^* f}\right)\right) = X^{\mathcal{C}}_f \in \mathfrak{X}(\mathcal{D})$$

for any $f \in C^\infty(TQ)$.

According to Section 3 we can construct Brownian motion on (Q, μ) by fixing Brownian motion $W = (W_i)$ in $\mathbb{R}^d$, $d = \dim Q$, and the Hamiltonian

$$H : T\mathfrak{F} \longrightarrow \mathbb{R}^d, \quad (u, \eta) \longmapsto \big(\nu(\eta, L_i(u))\big) = \big(H^i(u, \eta)\big), \tag{4.3}$$

where the L_i are defined by (3.9). This gives rise to the Stratonovich operator

$$\mathcal{S}^H : T\mathfrak{F} \times T\mathbb{R}^d \longrightarrow TT\mathfrak{F}, \quad (u, \eta, x, w) \longmapsto \sum X_{H^i}(u, \eta)\langle e_i, w\rangle,$$

where $\{e_1, \dots, e_d\}$ is the standard basis in $\mathbb{R}^d$. If Γ^H solves $\delta\Gamma^H = \mathcal{S}^H(W, \Gamma^H)\delta W$ then $\tau_{\mathfrak{F}} \circ \Gamma^H$ solves (3.10) and $\rho \circ \tau_{\mathfrak{F}} \circ \Gamma^H$ is Brownian motion in (Q, μ).

Definition 4.1. We define *constrained Brownian motion* to be the process

$$\Gamma^{\mathrm{nh}} := \rho \circ \tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}\mathfrak{F}}$$

in Q, where $\Gamma^{\mathcal{C}\mathfrak{F}}$ is a process in $\mathcal{D}\mathfrak{F}$ solving the Stratonovich equation

$$\delta\Gamma^{\mathcal{C}\mathfrak{F}} = P_{\mathfrak{F}}(\Gamma^{\mathcal{C}\mathfrak{F}})\mathcal{S}^H(W, \Gamma^{\mathcal{C}\mathfrak{F}})\delta W = \sum X^{\mathcal{C}\mathfrak{F}}_{H^i}(\Gamma^{\mathcal{C}\mathfrak{F}})\delta W^i. \tag{4.4}$$

Let $(\mathcal{D}\mathfrak{F})^\perp$ be the ν-orthogonal of $\mathcal{D}\mathfrak{F}$ and $\Pi_{\mathfrak{F}} : T\mathfrak{F} = \mathcal{D}\mathfrak{F} \oplus (\mathcal{D}\mathfrak{F})^\perp \to \mathcal{D}\mathfrak{F}$ be the orthogonal projection. Similarly, we define $\Pi : TQ = \mathcal{D} \oplus \mathcal{D}^\perp \to \mathcal{D}$ and we note that

$$\Pi \circ T\rho = T\rho \circ \Pi_{\mathfrak{F}}. \tag{4.5}$$

Equation (2.13) implies that $\tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}\mathfrak{F}}$ is a diffusion in $\mathfrak{F}$ generated by the Stratonovich equation

$$\delta(\tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}\mathfrak{F}}) = \sum \Pi_{\mathfrak{F}}(\tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}\mathfrak{F}})L_i(\tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}\mathfrak{F}})\delta W^i.$$

The associated Stratonovich operator will be denoted by $\mathcal{S}^{\mathcal{D}\mathfrak{F}}$. It is explicitly given by

$$\mathcal{S}^{\mathcal{D}\mathfrak{F}} : \mathfrak{F} \times T\mathbb{R}^d \longrightarrow \mathcal{D}\mathfrak{F}, \quad (u; w, w') \longmapsto \sum \Pi_{\mathfrak{F}}(u)L_i(u)\langle e_i, w'\rangle \tag{4.6}$$

where $\{e_1, \dots, e_d\}$ is the standard basis in $\mathbb{R}^d$.

Henceforth, local orthonormal frames on (Q, μ) and local sections of $\rho : \mathfrak{F} \to Q$ will be identified.

Theorem 4.2. *The process Γ^{nh} is a diffusion in Q (in the sense of Section 3) and its generator A has the form*

$$A = \tfrac{1}{2}\sum(\Pi u_i)(\Pi u_i) - \tfrac{1}{2}\sum \Pi\nabla^\mu_{\Pi u_i} u_i$$

in a local orthonormal frame $u = (u_i)$ on (Q, μ).

Proof. By Proposition 3.2, to see that Γ^{nh} is a diffusion we need to show that

$$\mathcal{S}^{\mathcal{D}\mathfrak{F}}(u\cdot k,w,w') = \mathcal{S}^{\mathcal{D}\mathfrak{F}}(u,kw,kw')\cdot k \tag{4.7}$$

for all $k\in K$. Here, $u\cdot k$ denotes the principal right action of $k\in K$ on $u\in\mathfrak{F}$ and condition (4.7) is equivalent to (3.11) since one can invert a right action to obtain a left action. Indeed, $\mathcal{D}\mathfrak{F}\cong_\omega(\mathfrak{F}\times_Q\mathcal{D})\oplus\mathrm{Ver}(\rho)$ and the definition of ν imply $\Pi_{\mathfrak{F}}\circ\mathrm{hl}^\omega=\mathrm{hl}^\omega\circ\Pi$ and therefore

$$\begin{aligned}\mathcal{S}^{\mathcal{D}\mathfrak{F}}(u\cdot k,w,w') &= \sum_{i=1}^d \Pi_{\mathfrak{F}}(u\cdot k)\mathrm{hl}^\omega_{u\cdot k}(u(ke_i))\langle e_i,w'\rangle\\ &= \left(\sum_{i=1}^d \mathrm{hl}^\omega_u(\Pi(\rho(u))u(e_i))\langle e_i,kw'\rangle\right)\cdot k\end{aligned}$$

which proves (4.7).

In order to calculate the generator A, let $f\in C^\infty(Q)_0$ and $u\in\mathfrak{F}$. Then

$$\begin{aligned}A(\rho^*f)(u) &= \tfrac12\sum_{i=1}^d(\Pi_{\mathfrak{F}}L_i)(\Pi_{\mathfrak{F}}L_i)(\rho^*f)(u)\\ &= \tfrac12\sum_{i=1}^d(\Pi u_i,-\omega(\Pi u_i))(\Pi u_i,-\omega(\Pi u_i))\rho^*f\\ &= \tfrac12\sum_{i=1}^d(\Pi u_i,-\omega(\Pi u_i))(\Pi u_i f)\\ &= \tfrac12\sum_{i=1}^d\Pi u_i(x\mapsto(\Pi u_i f)(x)) - \tfrac12\sum_{i=1}^d\omega(\Pi u_i)(v\mapsto(\Pi v(e_i)f))|_{v=u}\\ &= \tfrac12\sum_{i=1}^d(\Pi u_i)(\Pi u_i)f - \tfrac12\sum_{i=1}^d\left(\Pi\left.\frac{\partial}{\partial t}\right|_{t=0}\left(e^{t\omega(\Pi u_i)}u\right)e_i\right)f\\ &= \tfrac12\sum_{i=1}^d(\Pi u_i)(\Pi u_i)f - \tfrac12\sum_{i=1}^d(\Pi\omega(\Pi u_i)u_i)f\\ &= \tfrac12\sum_{i=1}^d(\Pi u_i)(\Pi u_i)f - \tfrac12\sum_{i=1}^d(\Pi\nabla^\mu_{\Pi u_i}u_i)f,\end{aligned}$$

where we have frequently dropped the base points to simplify the notation.

Alternatively, the above calculation can be done in local coordinates $u=(x^i,e^i_j)$ on $\mathfrak{F}$. Then, using the summation convention, $u_r=e^m_r\partial_m$ and $\Pi\partial_m=\Pi^n_m\partial_n$ with $\partial_m=\frac{\partial}{\partial x^m}$. The Christoffel symbols are given as usual by $\nabla^\mu_{\partial_k}\partial_j=\Gamma^i_{kj}\partial_i$. Using now the local coordinate description (3.7) of Hor^ω it follows that

$$\Pi_{\mathfrak{F}}L_r(u)=\mathrm{hl}^\omega(\Pi u_r)=\mathrm{hl}^\omega\Pi^n_m e^m_r\partial_n=\Pi^n_m e^m_r\Big(\partial_n-\Gamma^i_{nl}e^l_j\tfrac{\partial}{\partial e^i_j}\Big).$$

This yields

$$\begin{aligned}(\Pi_{\mathfrak{F}}L_r)(\Pi_{\mathfrak{F}}L_r)f &= (\Pi u_r)(\Pi u_r)f - \Pi^n_m e^m_r \Gamma^i_{nl} e^l_j \tfrac{\partial}{\partial e^i_j}(\Pi^b_a e^a_r \partial_b f)\\ &= (\Pi u_r)(\Pi u_r)f - \Pi^n_m e^m_r \Gamma^a_{nl} e^l_r \Pi^b_a \partial_b f\\ &= (\Pi u_r)(\Pi u_r)f - \Pi\nabla^\mu_{\Pi^n_m e^m_r \partial_n} e^l_r \partial_l f = (\Pi u_r)(\Pi u_r)f - \Pi\nabla^\mu_{\Pi u_r} u_r f\end{aligned}$$

which immediately yields the formula for A in the statement of the theorem. □

The second term in the above formula for A is reminiscent of the non-holonomic connection ∇^{nh}. This is defined as the linear connection

$$\nabla^{\mathrm{nh}} : \mathfrak{X}(Q) \times \mathfrak{X}(Q) \longrightarrow \mathfrak{X}(Q), \quad (X, Y) \longmapsto \nabla^\mu_X Y - (\nabla^\mu_X \Pi)Y. \tag{4.8}$$

If Y is a section of $\mathcal{D}$, one obtains the useful identity $\nabla^{\mathrm{nh}}_X Y = \Pi\nabla^\mu_X Y$. Let $\mathrm{Hess}^{\mathrm{nh}}$ be the Hessian of ∇^{nh}.

Corollary 4.3. *The non-holonomic diffusion Γ^{nh} is a martingale in Q with respect to the non-holonomic connection ∇^{nh}.*

Proof. Since $A(\rho^* f)(u)$ depends only on $\rho(u)$ (this follows from Theorem 3.2 but can also be checked directly), we may take a local frame $u = (u_i) = (u_a, u_\alpha)$ which is adapted to the decomposition $\mathcal{D} \oplus \mathcal{D}^\perp$ such that u_a are local sections of $\mathcal{D}$ and u_α are local sections of $\mathcal{D}^\perp$. Then $A = \frac{1}{2}\sum(u_a u_a - \nabla^{\mathrm{nh}}_{u_a} u_a) = \frac{1}{2}\sum \mathrm{Hess}^{\mathrm{nh}}(\cdot)(u_a, u_a)$ which is purely second order, by definition. □

4.B G-Chaplygin diffusions and stochastic non-holonomic reduction

Continue to assume that the non-holonomic system $(Q, \mathcal{D}, \mathcal{L})$ is invariant with respect to a free and proper action of a Lie group G. Denote the projection by $\pi : Q \twoheadrightarrow Q/G = M$.

The G- and K-action on $\mathfrak{F}$ commute. Thus, we may form the product action of $G \times K$ on $\mathfrak{F}$. Since the H^i from (4.3) are G-invariant, it follows that the Stratonovich operator (4.6) satisfies condition (3.11) with respect to the trivial G-representation on $\mathbb{R}^d$. Therefore, $\Gamma^{\mathcal{C}_{\mathfrak{F}}}$ induces a diffusion

$$\pi \circ \rho \circ \tau_{\mathfrak{F}} \circ \Gamma^{\mathcal{C}_{\mathfrak{F}}} = \pi \circ \Gamma^{\mathrm{nh}} =: \Gamma^M$$

on $M := \mathfrak{F}/(G \times K)$.

Now we make the additional assumption that the constraints are of Chaplygin type, i.e., $\mathcal{D}$ is the kernel of a principal connection one-form $\mathcal{A} \in \Omega^1(Q; \mathfrak{g})$. The non-holonomic connection (4.8) on Q induces a connection on M which will be referred to as the non-holonomic connection ∇^M on M; it is given by

$$\nabla^M : \mathfrak{X}(M) \times \mathfrak{X}(M) \longrightarrow \mathfrak{X}(M), \quad (X, Y) \longmapsto T\pi\left(\Pi\nabla^\mu_{\mathrm{hl}^{\mathcal{A}}X}(\mathrm{hl}^{\mathcal{A}}Y)\right)$$

where $\mathrm{hl}^{\mathcal{A}} : \mathfrak{X}(M) \to \mathfrak{X}(Q, \mathcal{D})$ (the space of vector fields on Q with values in the vector subbundle $\mathcal{D} \subset TQ$) is the horizontal lift map of $\mathcal{A}$. Recall from §2.B that the Riemannian metric μ on Q naturally induces a Riemannian metric μ_0 on the quotient $M := Q/G$. To calculate the generator A^M of Γ^M, take a local orthonormal frame $u = (u_a)$ on M. Similarly as in the proof of Corollary 4.3, the generator becomes

$$A^M = \tfrac{1}{2}\sum (u_a u_a - \nabla^{\mu_0}_{u_a} u_a) + \tfrac{1}{2}\sum (\nabla^{\mu_0}_{u_a} u_a - \nabla^{M}_{u_a} u_a) = \tfrac{1}{2}\Delta^{\mu_0} + \tfrac{1}{2} b \tag{4.9}$$

where $b = \sum (\nabla^{\mu_0}_{u_a} u_a - \nabla^{M}_{u_a} u_a)$.

Lemma 4.4. *$b = \mu_0^{-1}\beta$ where β is defined by (2.24).*

Proof. Essentially this formula is a special case of [30, Proposition 8.5]. For convenience we provide a proof by using a local orthonormal frame (u_a) on M. Let $K = \zeta \circ \mathrm{Curv}^{\mathcal{A}} \in \Omega^2(Q, TQ)$ be the curvature of $\zeta \circ \mathcal{A} \in \Omega^1(Q, TQ)$ where $\zeta : \mathfrak{g} \ni \xi \mapsto \xi_Q \in \mathfrak{X}(Q)$ is the fundamental vector field mapping of the G-action. Then

$$\begin{aligned}
\mu_0(\nabla^{\mu_0}_{u_a} u_a, u_b) &= -\mu_0([u_a, u_b], u_a) = -\mu(\mathrm{hl}^{\mathcal{A}}[u_a, u_b], \mathrm{hl}^{\mathcal{A}} u_a) \\
&= \mu(K(\mathrm{hl}^{\mathcal{A}} u_a, \mathrm{hl}^{\mathcal{A}} u_b), \mathrm{hl}^{\mathcal{A}} u_a) - \mu([\mathrm{hl}^{\mathcal{A}} u_a, \mathrm{hl}^{\mathcal{A}} u_b], \mathrm{hl}^{\mathcal{A}} u_a) \\
&= -\mu(\zeta_{\mathrm{Curv}^{\mathcal{A}}(\mathrm{hl}^{\mathcal{A}} u_a, \mathrm{hl}^{\mathcal{A}} u_b)}, \mathrm{hl}^{\mathcal{A}} u_a) + \mu(\Pi \nabla^{\mu}_{\mathrm{hl}^{\mathcal{A}} u_a} \mathrm{hl}^{\mathcal{A}} u_a, \mathrm{hl}^{\mathcal{A}} u_b) \\
&= -\langle J(\mathrm{hl}^{\mathcal{A}} u_a), \mathrm{Curv}^{\mathcal{A}}(\mathrm{hl}^{\mathcal{A}} u_a, \mathrm{hl}^{\mathcal{A}} u_b)\rangle + \mu_0(\nabla^{M}_{u_a} u_a, u_b).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\check{\mu}_0(\nabla^{\mu_0}_{u_a} u_a - \nabla^{M}_{u_a} u_a) &= \langle J(\mathrm{hl}^{\mathcal{A}} u_a), \mathrm{Curv}^{\mathcal{A}}(\mathrm{hl}^{\mathcal{A}} u_b, \mathrm{hl}^{\mathcal{A}} u_a)\rangle \mu_0(u_b, _) \\
&= \Xi(u_a)\left(X^{\mathrm{nh}}_{\mathcal{H}_c}, u_a^h\right) = \beta
\end{aligned}$$

where u_a^h is the horizontal lift of the local vector field $u_a \in \mathfrak{X}_{\mathrm{loc}}(M)$ to $u_a^h \in \mathfrak{X}_{\mathrm{loc}}(TM)$ with respect to the Levi-Civita connection ∇^{μ_0}. (We have identified linear functions on TM and one-forms on M as we did in the definition (2.24).) □

Theorem 4.5. *The G-Chaplygin system $(Q, \mathcal{D}, \mathcal{L})$ has a preserved measure if and only if the associated diffusion Γ^M is symmetric. Moreover, if $b = \mathrm{grad}^{\mu_0}(\log \mathcal{N})$ then the diffusion is $\mathcal{N}$-symmetric and $\mathcal{N}$ is the density of the preserved measure of $X^{\mathrm{nh}}_{\mathcal{H}_c}$ with respect to the Liouville volume.*

Proof. Using (4.9) and Lemma 4.4 this is a direct consequence of Proposition 2.5 and Theorem 3.4. □

When M is compact, then we infer from Section 3.D that measure preservation of the deterministic system is equivalent to time-reversibility of Γ^M which in turn is equivalent to the vanishing of the entropy production rate (3.19) of Γ^M. Moreover, if $b = \mathrm{grad}^{\mu_0}(\log \mathcal{N})$ then

$$\left(\int_M \mathcal{N} \mathrm{vol}_{\mu_0}\right)^{-1} \mathcal{N} \mathrm{vol}_{\mu_0}$$

is the (unique) equilibrium distribution of Γ^M. For most systems of practical interest, such as the Chaplygin ball, the two-wheeled robot, or the snakeboard, the manifold M is compact.

5 Examples

5.A The two-wheeled robot

The configuration space of the two-wheeled robot is

$$Q = S^1 \times S^1 \times \mathrm{SE}(2) = \{(\psi^1, \psi^2, x, y, \theta)\}.$$

Here (ψ^1, ψ^2) measure the positions of the wheels with the orientation such that the robot goes forward when the wheels go backward, and (x, y, θ) give the overall configuration of the robot in the plane. Let $G = \mathrm{SE}(2)$ and $M := S^1 \times S^1 = Q/G$. We use almost exactly the same notation as [11, Section 5.2.2]. It is assumed that the two wheels can be controlled independently and roll without slipping and without lateral sliding on the plane. The Lagrangian $\mathcal{L}$ of the system is the kinetic energy corresponding to the metric

$$\begin{aligned}\mu = {} & J_w(d\psi^1 \otimes d\psi^1 + d\psi^2 \otimes d\psi^2) + m(dx \otimes dx + dy \otimes dy) \\ & + m_0 l \cos\theta(dy \otimes d\theta + d\theta \otimes dy) - m_0 l \sin\theta(dx \otimes d\theta + d\theta \otimes dx) + J_0 d\theta \otimes d\theta.\end{aligned}$$

Here $m = m_0 + 2m_w$, m_0 is the mass of the robot without the wheels, m_w is the mass of each wheel, J_w is the moment of inertia of each wheel, J_0 is the moment of inertia of the robot about the vertical axis, and l is the distance from the vehicle's center of mass to the midpoint of the axis which connects the two wheels. Let $2c$ denote the distance between the contact points of the two wheels with the ground, and R the radius of the wheels. The constraints are given by the kernel of the $\mathfrak{g} \cong \mathbb{R}^3$-valued one-form

$$\mathcal{A} = \begin{pmatrix} dx + yd\theta + \frac{R}{2}\cos\theta(d\psi^1 + d\psi^2) + y\frac{R}{2c}(d\psi^1 - d\psi^2) \\ dy - xd\theta + \frac{R}{2}\sin\theta(d\psi^1 + d\psi^2) - x\frac{R}{2c}(d\psi^1 - d\psi^2) \\ d\theta + \frac{R}{2c}(d\psi^1 - d\psi^2) \end{pmatrix}.$$

Thus the constraint distribution is $\mathcal{D} = \mathcal{A}^{-1}(0) = \mathrm{span}\{\xi_1, \xi_2\}$ where

$$\xi_1 := \partial_{\psi^1} - \tfrac{R}{2}(\cos\theta\partial_x + \sin\theta\partial_y + \tfrac{1}{c}\partial_\theta) \quad \text{and} \quad \xi_2 := \partial_{\psi^2} - \tfrac{R}{2}(\cos\theta\partial_x + \sin\theta\partial_y - \tfrac{1}{c}\partial_\theta). \tag{5.1}$$

Symmetry reduction

Since $\mathcal{A}$ is a connection one-form for the principal bundle $\pi : Q \twoheadrightarrow M = Q/G$, the system $(Q, \mathcal{D}, \mathcal{L})$ is of G-Chaplygin type. Let $J : TQ \to \mathfrak{g}^*$ be the momentum map of

the G-action. Then a calculation shows that

$$\langle J(q, v^1\xi_1 + v^2\xi_2), \mathrm{Curv}^{\mathcal{A}}(\partial_{\psi^1}, \partial_{\psi^2})\rangle = m_0 l \tfrac{R^3}{4c^2}(v^2 - v^1).$$

Note that this vanishes if $l = 0$. Let us apply the Gram-Schmidt orthonormalization scheme with respect to the reduced metric μ_0 to $\partial_{\psi^1}, \partial_{\psi^2}$ and denote the result by u_1, u_2. Thus,

$$u_1 = \left(J_w + m\tfrac{R^2}{4} + J_0\tfrac{R^2}{4c^2}\right)^{-\frac{1}{2}} \partial_{\psi^1} \tag{5.2}$$

and

$$u_2 = \left(J_w \frac{\left(J_w + m\frac{R^2}{2} + J_0\frac{R^2}{2c^2}\right) + mJ_0\frac{R^4}{4c^2}}{J_w + m\frac{R^2}{4} + J_0\frac{R^2}{4c^2}}\right)^{-\frac{1}{2}} \left(\partial_{\psi^2} - \frac{m\frac{R^2}{4} - J_0\frac{R^2}{4c^2}}{J_w + m\frac{R^2}{4} + J_0\frac{R^2}{4c^2}}\partial_{\psi^1}\right). \tag{5.2a}$$

Using the relation

$$\mu_0(b) = \beta = \mu_0\left(\sum_{i,j}\langle J(u_i), \mathrm{Curv}^{\mathcal{A}}(u_j, u_i)\rangle u_j, \_\right)$$

and expanding everything in terms of $\partial_{\psi^1}, \partial_{\psi^2}$, one finds that the drift vector b equals

$$b = l m_0 R^3\Big(J_w(4c^2 J_w + m2c^2R^2 + 2J_0R^2) + mJ_0R^4\Big)^{-1}(\partial_{\psi^1} + \partial_{\psi^2}). \tag{5.3}$$

Since M is compact, this b cannot be the gradient of a function for $l \neq 0$. Thus, we can conclude that the deterministic two-wheeled robot does not have a preserved volume for $l \neq 0$ and that the associated stochastic system is not time-reversible.

Kinematics of the noisy cart

Formula (5.3) seems to imply that the stochastic cart (with zero initial velocity) acquires a tendency to go backwards when the center of mass is displaced towards the rear. To see that this is indeed the case we should check that the horizontally lifted mean curve coincides with the expected motion of the cart.

Since $TQ \cong_\mu T^*Q$ (vector bundle isomorphism induced by the Riemannian metric μ on Q) and TTQ are trivial, we may view $TQ \subset TTQ$ as a vector subbundle, so that $b^h = \mathrm{hl}^{\mathcal{A}}(b)$ and u_a^h become vector fields on TTQ. Then the stochastic dynamics $\Gamma^{\mathcal{D}} = (q_t, p_t)$ on $\mathcal{D}$ are generated by the operator

$$A^{\mathcal{D}} = X^{\mathcal{C}}_{\mathcal{H}} + \tfrac{1}{2}b^h + \sum_a u_a^h u_a^h$$

or by the Stratonovich equation

$$\delta\Gamma^{\mathcal{D}} = X^{\mathcal{C}}_{\mathcal{H}}(\Gamma^{\mathcal{D}})\delta t + \tfrac{1}{2}b^h(q_t) + \sum_a u_a^h(q_t)\delta W^a.$$

Here $X^{\mathcal{C}}_{\mathcal{H}}$ was defined in Section 2. Now in local coordinates (q^i, p_i) on TQ the stochastic equations of motion are

$$\delta(p_i \circ \Gamma^{\mathcal{D}}_t) = dp_i \left(PX_{\mathcal{H}}(q_t, p_t)\right) \delta t,$$
$$\delta(q^i \circ \Gamma^{\mathcal{D}}_t) = (p_t)_i \delta t + \tfrac{1}{2} dq^i \left(b^h(q_t)\right) \delta t + dq^i \left(\sum_a u^h_a(q_t)\right) \delta W^a.$$

If the initial conditions are $(q_0, 0)$ then the solution is given by $(q_t, 0)$ where q_t satisfies

$$\delta q_t = \tfrac{1}{2} b^h(q_t)\delta t + \sum_a u^h_a(q_t)\delta W^a.$$

Let $q_t = (\psi^1_t, \psi^2_t, x_t, y_t, \theta_t) = (q^i_t)$. By [37, Lemma 7.3.2] we have

$$E\left[q^i_t\right] = q^t_0 + E\left[\int_0^t \left(\tfrac{1}{2}\left(\sum_a u^h_a u^h_a q^i\right)(q_s) + \tfrac{1}{2}(b^h q^i)(q_s)\right) ds\right].$$

Rewriting

$$\sum_a u^h_a u^h_a = A(\xi_1\xi_1 + \xi_2\xi_2) + B(\xi_1\xi_2 + \xi_2\xi_1),$$

with

$$A := \frac{\mu(\xi_1, \xi_1)}{\mu(\xi_1, \xi_1)^2 - \mu(\xi_1, \xi_2)^2}, \qquad B := \frac{\mu(\xi_1, \xi_2)}{\mu(\xi_1, \xi_1)^2 - \mu(\xi_1, \xi_2)^2},$$

and noting that $\mu(\xi_1, \xi_2)$ and $\mu(\xi_1, \xi_1) = \mu(\xi_2, \xi_2)$ are constants, implies that $\sum_a u^h_a u^h_a q^i = 0$ for $(q^i) = (\psi^1, \psi^2, x, y, \theta)$. Thus

$$\tfrac{\partial}{\partial t} E\left[q_t\right] = \tfrac{1}{2} b^h\left(E[q_t]\right),$$

i.e., $E[q_t]$ is the horizontal lift of the integral curve of $\frac{1}{2}b \in \mathfrak{X}(M)$.

Therefore, constraints and noise couple to produce a backwards drift of the robot. We emphasize that this is a stochastic non-holonomic effect which does not appear in a Hamiltonian setting. Indeed, the Hamiltonian reduction of Brownian motion at the zero-momentum level yields Brownian motion and this is consistent with the fact that the reduced two-wheeled robot system is actually Hamiltonian when $l = 0$.

Trajectory planning for noisy wheels

Generally speaking, consider a non-holonomic system (such as the cart) and assume that it is controlled so as to follow a predefined smooth curve $c(t) \in Q$, $t \in [0, T]$ when no noise is present. When the system is stochastically perturbed we may ask whether $c(t)$ is also the expected motion of the perturbed system.

Suppose we want to steer the robot so that it follows a predefined curve in the plane. As a curve we consider the circle C of radius $\rho \geq 0$ centered at the origin. The initial

configuration of the robot should be $(x_0, y_0, \theta_0) = (\rho, 0, \pi/2)$ and the vehicle should go around the circle in the positive sense. It is assumed that the wheel speeds can be individually controlled.

In this section we consider the example of [42]. Here the wheels are subject to a Gaussian white noise which is modeled by the Stratonovich equation

$$\delta\Gamma^M = \sqrt{D_1}\partial_{\psi^1}\delta W^1 + \sqrt{D_2}\partial_{\psi^2}\delta W^2 \tag{5.4}$$

in TM where (W^i) is Brownian motion in $\mathbb{R}^2$ and $D_i > 0$ are constants. In this setup one assumes that the controlled vehicle is not affected by the kinematics of the problem, thus effectively forgetting the metric μ. The generator of Γ^M is $\frac{1}{2}(D_1\partial^2_{\psi^1} + D_2\partial^2_{\psi^2})$. Equation (5.4) lifts to a Stratonovich equation

$$\delta\Gamma^Q = \sqrt{D_1}\xi_1\delta W^1 + \sqrt{D_2}\xi_2\delta W^2$$

in TQ. This is in accordance with the general theory of [16, 17]; the generator of Γ^Q is $A^Q = \frac{1}{2}(D_1\xi_1^2 + D_2\xi_2^2)$ which can be regarded as the horizontal lift of A^M. Consider the deterministic input vector field

$$u(t) := -\lambda(t)\left(\tfrac{\rho+c}{R}\partial_{\psi^1} + \tfrac{\rho-c}{R}\partial_{\psi^2}\right)$$

on M where the control

$$\lambda(t) = \begin{cases} 2t, & 0 \le t < \sqrt{\frac{\pi}{2}} =: t_1 \\ 2\frac{t-T}{t_1-T}, & t_1 \le t \le T := \frac{3\pi}{2} + t_1 \end{cases}$$

is chosen such that the unperturbed robot traverses the nominal curve C exactly once and initial and final speed are 0. The equation for the controlled noisy robot is thus

$$\delta\Gamma^u = -\lambda(t)(\tfrac{\rho+c}{R}\xi_1(\Gamma^u) + \tfrac{\rho-c}{R}\xi_2(\Gamma^u))\delta t + \sqrt{D_1}\xi_1(\Gamma^u)\delta W^1 + \sqrt{D_2}\xi_2(\Gamma^u)\delta W^2$$

and the corresponding (time-dependent) generator is $A^u = A^Q + \mathrm{hl}^{\mathcal{A}}(u)$ whence by [37, Lemma 7.3.2]

$$E\left[f(\Gamma^u_t)\right] = f(\Gamma^u_0) - E\left[\int_0^t (A^Q + \mathrm{hl}^{\mathcal{A}}(u))(f)(\Gamma^u_s)\, ds\right] \tag{5.5}$$

for $f \in C^\infty(Q)$. (The expectation is taken with respect to the underlying probability.) Let

$$\Gamma_0 = (0, 0, \rho, 0, \tfrac{\pi}{2}), \quad \Gamma^u_t =: (\psi^1_t, \psi^2_t, x_t, y_t, \theta_t), \quad \kappa := \tfrac{(D_2-D_1)R^2}{8c}.$$

Using (5.5) we find

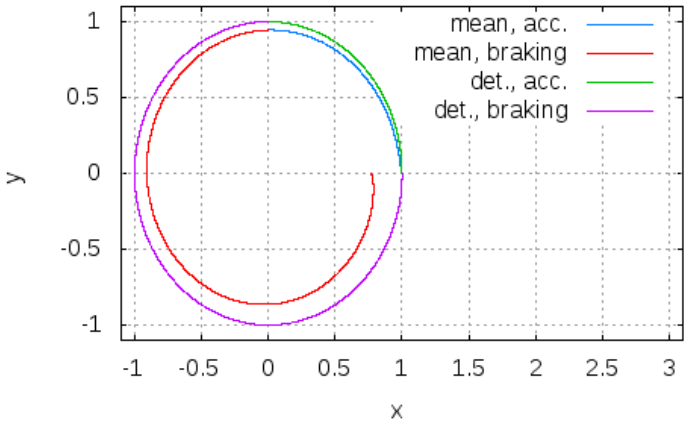
$$\begin{aligned} E[x_t] &= \kappa \int_0^t E[\cos(\theta_s)]\, ds + \rho \int_0^t \lambda(s) E[\sin(\theta_s)]\, ds \\ &= \kappa \int_0^t e^{\kappa t} \cos(\theta(s))\, ds + \rho \int_0^t \lambda(s) e^{\kappa t} \sin(\theta(s))\, ds \\ E[y_t] &= -\kappa \int_0^t E[\sin(\theta_s)]\, ds + \rho \int_0^t \lambda(s) E[\cos(\theta_s)]\, ds \\ &= -\kappa \int_0^t e^{\kappa t} \sin(\theta(s))\, ds + \rho \int_0^t \lambda(s) e^{\kappa t} \cos(\theta(s))\, ds \end{aligned}$$

where $\theta(t)$ differs from θ_t and is defined by

$$\theta(t) = \begin{cases} t^2 + \frac{\pi}{2}, & 0 \le t < \sqrt{\frac{\pi}{2}} =: t_1; \\ \frac{(t-T)^2}{t_1 - T} + \frac{5\pi}{2}, & t_1 \le t \le T := \frac{3\pi}{2} + t_1. \end{cases}$$

This determines the orientation of the vehicle.

We have solved for $(E[x_t], E[y_t])$ using Maxima and its built in Runge-Kutta scheme. Here is a plot:



The data are $\rho = 1$, $D_1 = 1.2$, $D_2 = 0.8$, $R = 0.3$, $c = 0.1$. we have plotted the accelerating and braking parts of $(E[x_t], E[y_t])$ as blue and red, and the accelerating and braking parts of the unperturbed controlled robot $(x(t), y(t))$ as green and magenta, respectively.

The discrepancy between the deterministic trajectory and the mean curve of the perturbed system is quite obvious. This phenomenon has also been observed in [42] by means of numerical simulations, and [42] have also proposed a trajectory planning algorithm which takes the perturbation into account. When comparing the above picture to that of [42], it should be noted that we have chosen a different convention for the orientation of the wheels.

5.B Microscopic snakeboard under molecular bombardment

This is not a G-Chaplygin system but does fit the set-up of Section 4.A.

In describing the snakeboard we follow mostly the presentation of [9]. There is, however, one difference: the metric which we use to define the kinetic energy is that of [7]. This considerably simplifies some of the formulas. We further assume that the angle of the front axis equals minus that of the back axis. Thus the configuration space of this system is

$$Q = S^1 \times S^1 \times \mathrm{SE}(3) = \{q = (\phi, \psi, x, y, \theta)\}.$$

The constraint distribution is the kernel of the $\mathbb{R}^2$-valued one-form $\omega = (\omega_1, \omega_2)$ given by

$$\begin{aligned}\omega_1(q) &= -\sin(\theta+\phi)dx + \cos(\theta+\phi)dy - r\cos(\phi)d\theta,\\ \omega_2(q) &= -\sin(\theta-\phi)dx + \cos(\theta-\phi)dy - r\cos(\phi)d\theta\end{aligned}$$

where $2r$ is the distance between the axes measured from their respective midpoints. Thus

$$\mathcal{D} = \text{span}\,\{\partial_\phi, \partial_\psi, s := a\partial_x + b\partial_y + c\partial_\theta\}$$

where the functions a, b, c are given by

$$\begin{aligned}a &= -r\big(\cos(\phi)\cos(\theta-\phi) + \cos(\phi)\cos(\theta+\phi)\big),\\ b &= -r\big(\cos(\phi)\sin(\theta-\phi) + \cos(\phi)\sin(\theta+\phi)\big),\\ c &= \sin(2\phi).\end{aligned}$$

Let m be the mass of the board, J_0 its moment of inertia, and J_ϕ, J_ψ, J_θ the moments of inertia corresponding to rotation about the angle ϕ, ψ, and θ respectively. Then the Lagrangian of the system is the kinetic energy of the metric

$$\mu = m(dx\otimes dx + dy\otimes dy) + K d\theta\otimes d\theta + J_\phi d\phi\otimes d\phi + J_\psi d\psi\otimes d\psi + J_\psi(d\psi\otimes d\theta + d\theta\otimes d\psi)$$

where $K := J_\theta + J_\psi + J_\phi$.

Let us assume that the snakeboard is perturbed by white noise. Using the left trivialization of TQ this can be modeled by a Stratonovich operator of the form

$$\begin{aligned}\mathcal{S} : Q \times T\mathbb{R}^6 &\longrightarrow TQ,\\ (q, w, w') &\longmapsto \sigma \sum \langle e_i, w'\rangle u_i \delta W^i\end{aligned}$$

where (u_i) is a left invariant orthonormal frame on Q and $\sigma \geq 0$ is a parameter specifying the field strength. According to the results of Section 4, constrained Brownian motion is a diffusion Γ^{nh} with generator

$$A = \tfrac{\sigma^2}{2} \sum (u_a u_a - \Pi \nabla^\mu_{\Pi u_a} u_a).$$

Here (u_a) is an orthonormal frame of $\mathcal{D}$ and $\Pi : TQ = \mathcal{D} \oplus \mathcal{D}^\perp \to \mathcal{D}$. We fix this frame to be

$$u_1 = J_\phi^{-\frac{1}{2}} \partial_\phi, \quad u_2 = \eta^{-\frac{1}{2}} (\partial_\psi - J_\psi \tfrac{c}{\varepsilon} s), \quad u_3 = \varepsilon^{-\frac{1}{2}} s$$

where

$$\varepsilon = m(a^2 + b^2) + Kc^2, \quad \eta = J_\psi (1 - \tfrac{J_\psi c^2}{\varepsilon}).$$

Note that η and ε are functions of ϕ only. A calculation now shows that we have, for the (trivial) connection ∇ associated to μ,

$$\begin{aligned}
&\nabla_{u_1} u_1 = 0, \\
&\nabla_{u_2} u_2 = \tfrac{J_\psi^2 c^3}{\eta \varepsilon^2} \big((\partial_\theta a) \partial_x + (\partial_\theta b) \partial_y \big) \in \mathcal{D}^\perp, \\
&\nabla_{u_3} u_3 = \tfrac{c}{\varepsilon} \big((\partial_\theta a) \partial_x + (\partial_\theta b) \partial_y \big) \in \mathcal{D}^\perp.
\end{aligned}$$

Thus $\Pi \nabla^\mu_{\Pi u_a} u_a = 0$ for this frame and Γ^{nh} is given by the Stratonovich equation

$$\delta \Gamma^{\text{nh}} = \sigma \sum u_a(\Gamma^{\text{nh}}) \delta W^a. \tag{5.6}$$

As in the theory of [9], we fix the horizontal space of the principal bundle $\pi : Q \twoheadrightarrow Q/G = T^2 = M$ associated to the distribution $\mathcal{D}$ to be given by the span of $\{u_1, u_2\}$. The corresponding connection form is denoted by $\mathcal{A}$. Consider the control vector fields

$$\begin{aligned}
U_\phi(t) &= u'_\phi(t) \partial_\phi, \qquad & u_\phi(t) &= a_\phi \sin(\omega_\phi t), \\
U_\psi(t) &= u'_\psi(t) \partial_\psi, \qquad & u_\psi(t) &= a_\psi \sin(\omega_\psi t)
\end{aligned}$$

in the control space TM. Their horizontal lifts are $\text{hl}^{\mathcal{A}}(U_\phi) = u'_\phi(t)\partial_\phi$ and $\text{hl}^{\mathcal{A}}(U_\psi) = u'_\psi(t)(\partial_\psi - J_\psi \frac{c}{\varepsilon} s)$. Combining this with (5.6) yields

$$\delta \Gamma^u = \text{hl}^{\mathcal{A}}_{\Gamma^u}(U_\phi + U_\psi) \delta t + \sigma u_a(\Gamma^u) \delta W^a \tag{5.7}$$

which describes the stochastic perturbation of the controlled snakeboard with deterministic gait input $(\phi, \psi) = (u_\phi(t), u_\psi(t))$.

Since the variables (ϕ, ψ) are also the ones which can be controlled, we are interested in estimating Γ^u given that the projected process $X_t = \pi \circ \Gamma^u_t$ satisfies the projected equation

$$\delta X = (U_\phi(t) + U_\psi(t)) \delta t + J_\psi^{-\frac{1}{2}} \delta W^1 \partial_\phi + \eta(X)^{-\frac{1}{2}} \delta W^2 \partial_\psi \,. \tag{5.8}$$

This is the filtering problem $E[\Gamma^u_t | \pi \circ \Gamma^u_t = X_t] =: Z_t$ and the solution is provided by [16, 17]: The process Γ^u can be decomposed as

$$\Gamma^u_t = g^X_t \cdot X^h_t$$

where X^h_t is the horizontal lift of $X_t = \pi \circ \Gamma^u_t$ and g^X is the reconstruction process. These satisfy the Stratonovich equations

$$\delta X^h = \mathrm{hl}^{\mathcal{A}}_{X^u}(U_\phi + U_\psi)\delta t + \sigma(u_1(X^h)\delta W^1 + u_2(X^h)\delta W^2), \qquad X^h_0 = \Gamma^u_0 = q_0 \in Q$$

and

$$\delta g^X_t = \sigma T_e L_{g^X_t}.\mathcal{A}_{X^h_t} u_3(X^h_t)\delta W^3, \qquad g^X_0 = e \in G.$$

(See also Section 3.C.) By [17] we have that

$$Z_t = E[g^X_t] \cdot X^h_t. \tag{5.9}$$

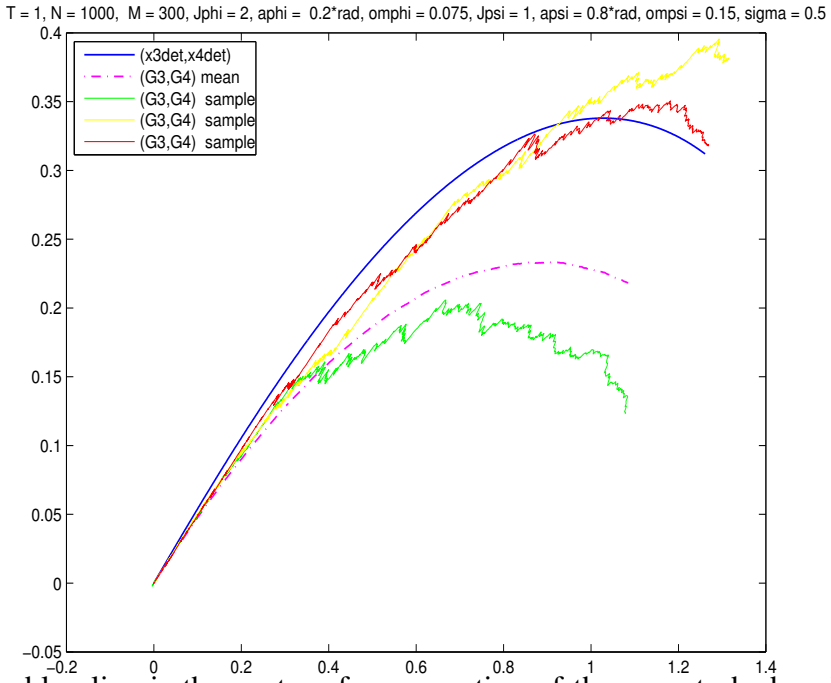
Let $X^h_t = (\phi_t, \psi_t, x_t, y_t, \theta_t)$ and $E[g^X_t] = (a_t, b_t, \gamma_t) \in G$. It follows from Proposition 3.3 that the mean reconstruction curve $E[g^X_t]$ is determined by the time- and $\omega \in \Omega$-dependent ODE

$$\begin{aligned} \tfrac{\partial}{\partial t} E[g^X_t] &= \tfrac{\partial}{\partial t} \begin{pmatrix} a_t \\ b_t \\ \gamma_t \end{pmatrix} \\ &= \tfrac{\sigma^2 c(\phi_t)}{2\sqrt{\varepsilon(\phi_t)}} \begin{pmatrix} -\big(a(\phi_t,\theta_t) + y_t c(\phi_t)\big)\sin(\gamma_t) - \big(b(\phi_t,\theta_t) - x_t c(\phi_t)\big)\cos(\gamma_t) \\ \big(a(\phi_t,\theta_t) + y c(\phi_t)\big)\cos(\gamma_t) - \big(b(\phi_t,\psi_t) - x_t c(\phi_t)\big)\sin(\gamma_t) \\ 0 \end{pmatrix}. \end{aligned} \tag{5.10}$$

Using the rule for transforming Stratonovich equations to Itô type, we can characterize X^h_t by the Itô equation

$$\begin{aligned} dX^h_t = &\begin{pmatrix} u'_\phi(t) \\ u'_\psi(t) \\ -J_\psi \frac{a(\phi_t,\theta_t)c(\phi_t)}{\varepsilon(\phi_t)} u'_\psi(t) + \sigma^2 \frac{c(\phi_t)^3}{2\eta(\phi_t)\varepsilon(\phi_t)^2}(\partial_\theta a)(\phi_t,\theta_t) \\ -J_\psi \frac{b(\phi_t,\theta_t)c(\phi_t)}{\varepsilon(\phi_t)} u'_\psi(t) + \sigma^2 \frac{c(\phi_t)^3}{2\eta(\phi_t)\varepsilon(\phi_t)^2}(\partial_\theta b)(\phi_t,\theta_t) \\ -J_\psi \frac{c(\phi_t)^2}{\varepsilon(\phi_t)} u'_\psi(t) \end{pmatrix} dt \\ &+ \sigma u_1(X^h) dW^1 + \sigma u_2(X^h) dW^2. \end{aligned} \tag{5.11}$$

Equation (5.11) involves an iterated dependence on trigonometric functions, and hence numerical simulation is not straightforward. A naive approach would involve to run an Euler-Maruyama and an Euler simulation for (5.11) and (5.10) respectively, and to multiply the results together according to (5.9) which is the action of G on Q. This yields Z_t. Running the simulation sufficiently many times and computing the average yields the mean $E[Z_t]$. We have implemented this scheme and the results seem reasonably stable up to time 1, according to a first order test. Beyond that time, the trajectories blow up very quickly, which is a strong indication that the method is unstable and a more detailed analysis of the numerical implementation is necessary. Our preliminary results are contained in the plot below.



The blue line is the center of mass motion of the unperturbed snakeboard and and the dotted magenta line is the mean motion of the stochastic snakeboard with the same deterministic input. Additionally 3 sample plots have been included. The data are as indicated above: T is the runtime, $1/N$ the step size, M the number of experiments, rad $= \frac{180}{\pi}$ and σ the parameter specifying the strength of the white noise. The initial conditions are $q_0 = (0, 0, 0, 0, 0.5)$.

Acknowledgments. T.S. Ratiu was partially supported by Swiss NSF grant 200021-140238 and by the government grant of the Russian Federation for support of research projects implemented by leading scientists, Lomonosov Moscow State University under the agreement No. 11.G34.31.0054.

References

[1] Abraham, R. and Marsden, J.E.: Foundations of Mechanics. 2nd ed., Addison-Wesley, 1978

[2] Alterovitz, R., Branicky, M., Goldberg, K.: Constant-curvature motion planning under uncertainty with applications in image-guided medical needle steering. In "Algorithmic Foundation of Robotics VII", S. Akella et al. (Eds.), vol. **47**, Springer-Verlag, Berlin, 319–334 (2008)

[3] Arnold, V.I. Kozlov, V.V., Neishtadt, A.I.: Mathematical Aspects of Classical and Celestial Mechanics. Springer-Verlag, Berlin, (2002)

[4] Bates, L., and Śniatycki, J.: Nonholonomic reduction. Rep. Math. Phys. **32**(1), 99–115 (1993)

[5] Bismut, J.-M.: Mecanique Aléatoire. Lecture Notes in Math., vol. **866**, Springer-Verlag, Berlin, (1981)

[6] Bloch, A.M.: Nonholonomic Mechanics and Control. Springer-Verlag, Berlin, (2003)

[7] Bullo, F., Lewis, A.: Kinematic controllability and motion planning for the snakeboard. IEEE Transactions on Robotics and Automation, vol. **19**(3), 494–498 (2003)

[8] Cantrijn, F., Cortes, J., de Leon, M., Martín de Diego, D.: On the geometry of generalized Chaplygin systems. Math. Proc. Camb. Phil. Soc. **132**, 323–351 (2002)

[9] Cendra, H., Marsden, J.E., Ratiu, T.S.: Geometric mechanics, Lagrangian reduction and nonholonomic systems. In "Mathematics Unlimited: 2001 and Beyond", B. Enguist and W. Schmid (eds.), Springer-Verlag, Berlin, 221–273 (2001)

[10] Chaplygin, S.A.: On a ball's rolling on a horizontal plane. Regul. Chaotic Dyn. **7**, 131–148 (2002); translation of original from: Mathematical collection of the Moscow Mathematical Society **24**, 139–168 (1903) (Russian)

[11] Cortes Monforte, J.: Geometric Control and Numerical Aspects of Non-holonomic Systems. Lecture Notes in Math. **1793**, Springer-Verlag, Berlin (2002)

[12] Cushman, R., Duistermaat, J.J., Śniatycki, J.: Geometry of Nonholonomically Constrained Systems. Advanced Series in Nonlinear Dynamics, **26**, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2010)

[13] Duistermaat, J.J.: Chaplygin's sphere. `arXiv:math/0409019v1`.

[14] Duistermaat, J.J., Kolk, J.A.C.: Lie Groups, Universitext, Springer-Verlag, Berlin (2000)

[15] Ehlers, K., Koiller, J., Montgomery, R., Rios, P.M.: Nonholonomic systems via moving frames: Cartan equivalence and Chaplygin Hamiltonization. In "The Breadth of Symplectic and Poisson Geometry", Progress in Mathematics, **232** Birkhäuser, Boston, 75–120 (2004)

[16] Elworthy, K.D., Le Jan, Y., Li, X-M.: Equivariant diffusions on principal bundles. Stochastic analysis and related topics in Kyoto, vol. **41** of Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, 31–47 (2004)

[17] Elworthy, K.D., Le Jan, Y., Li, X-M.: The Geometry of Filtering. Springer-Verlag, Basel (2010)

[18] Emery, M.: Stochastic Calculus in Manifolds. Universitext, Springer-Verlag, Berlin (1989)

[19] Fehér, L., Pusztai, B.G.: Spin Calogero models obtained from dynamical r-matrices and geodesic motion. Nuclear Physics B **734**, 304–325 (2006)

[20] Gudmundsson, S., Koppas, E.: On the geometry of tangent bundles. Expos. Math. **20**, 1–41 (2002)

[21] Hochgerner, S.: Singular cotangent bundle reduction and spin Calogero-Moser systems. Diff. Geom. Appl. **26**(2), 169–192 (2008)

[22] ———, Chaplygin systems associated to Cartan decompositions of semi-simple Lie groups. Diff. Geom. Appl. **28**(4), 436–453 (2010)

[23] ———, Stochastic Chaplgin systems. Rep. Math. Phys. **66**(3), 385–401 (2010)

[24] ———, Symmetry reduction of Brownian motion and quantum Calogero-Moser systems. Stoch. Dyn., to appear.

[25] Hochgerner, S., and García-Naranjo, L.: G-Chaplygin systems with internal symmetries. Truncation, and an (almost) symplectic view of Chaplygin's ball. J. Geom. Mech. **1**(1), 35–53 (2009)

[26] Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland Publishing Company, Kodansha, 2nd ed. (1989)

[27] Jiang, D.-Q. Qian, M., Qian, M.-P.: Mathematical Theory of Nonequilibrium Steady States. On the Frontier of Probability and Dynamical Systems. Lecture Notes in Math., vol. **1833**, Springer-Verlag, Berlin (2004)

[28] Kazhdan, D., Kostant, B., and Sternberg, S.: Hamiltonian group actions and dynamical systems of Calogero type. Comm. Pure Appl. Math. **31**, 481–507 (1978)

[29] Kent, J.: Time-reversible diffusions. Adv. Appl. Prob. **10**(4), 819–835 (1978)

[30] Koiller, J.: Reduction of some classical nonholonomic systems with symmetry. Arch. Rat. Mech. Anal., **118**, 113–148 (1992)

[31] A. Kolmogorov, A.: Zur Umkehrbarkeit der statistischen Naturgesetze. Math. Ann. **113**, 766–772 (1936) (German)

[32] Lázaro-Camí, J.-A., Ortega, J.-P.: Stochastic Hamiltonian dynamical systems. Rep. Math. Phys. **61**, 65–122 (2008)

[33] _____, Reduction, reconstruction, and skew-product decomposition of symmetric stochastic differential equations. Stoch. Dyn. **9**(1), 1–46 (2009)

[34] _____, The stochastic Hamilton-Jacobi equation. J. Geom. Mech. **1**(3), 295-315 (2009)

[35] Marsden, J.E., Montgomery, R., Ratiu, T.S.: Reduction, symmetry, and phases in mechanics. Memoirs of the American Mathematical Society, Volume 88, Number 436 (1990)

[36] Montgomery, R.: A tour of SubRiemannian Geometries, their Geodesics and Applications. Math. Surveys and Monographs vol. **91**, Amer. Math. Soc., Providence (2002)

[37] Øksendal, B.: Stochastic Differential Equations. Universitext, Springer-Verlag, Berlin (2007)

[38] Olshanetsky, M.A., Perelomov, A.M.: Quantum systems related to root systems, and radial parts of Laplace operators. Funct. Anal. Appl. **12**, 121–128 (1978)

[39] _____, Quantum integrable systems related to Lie algebras. Phys. Rep. **94**(6), 313–404 (1983)

[40] Sekizawa, M.: On Riemannian geometry of orthonormal frame bundles. Note Mat. **1**, suppl. 1, 383–394, (2008)

[41] Stroock, D.W.: Probability Theory, an Analytic View. Cambridge Univ. Press (1993)

[42] Zhou, Y., Chirikjian, G.S.: Planning for Noise-induced Trajectory Bias in Nonholonomic Robots with Uncertainty. In ICRA pp. 4596–4601 (2004).