Journal of
Nonlinear
Science

Check for
updates

From Second-Order Differential Geometry to Stochastic Geometric Mechanics

Qiao Huang[1] · Jean-Claude Zambrini[1]

Abstract

Classical geometric mechanics, including the study of symmetries, Lagrangian and Hamiltonian mechanics, and the Hamilton–Jacobi theory, are founded on geometric structures such as jets, symplectic and contact ones. In this paper, we shall use a partly forgotten framework of second-order (or stochastic) differential geometry, developed originally by L. Schwartz and P.-A. Meyer, to construct second-order counterparts of those classical structures. These will allow us to study symmetries of stochastic differential equations (SDEs), to establish stochastic Lagrangian and Hamiltonian mechanics and their key relations with second-order Hamilton–Jacobi–Bellman (HJB) equations. Indeed, stochastic prolongation formulae will be derived to study symmetries of SDEs and mixed-order Cartan symmetries. Stochastic Hamilton's equations will follow from a second-order symplectic structure and canonical transformations will lead to the HJB equation. A stochastic variational problem on Riemannian manifolds will provide a stochastic Euler–Lagrange equation compatible with HJB one and equivalent to the Riemannian version of stochastic Hamilton's equations. A stochastic Noether's theorem will also follow. The inspirational example, along the paper, will be the rich dynamical structure of Schrödinger's problem in optimal transport, where the latter is also regarded as a Euclidean version of hydrodynamical interpretation of quantum mechanics.

Keywords Stochastic Hamiltonian mechanics · Stochastic Lagrangian mechanics · Hamilton–Jacobi–Bellman equations · Stochastic Hamilton's equations · Stochastic

✉ Jean-Claude Zambrini
jczambrini@fc.ul.pt

Qiao Huang
qhuang@fc.ul.pt

[1] Department of Mathematics, Faculty of Sciences, Group of Mathematical Physics (GFMUL), University of Lisbon, Campo Grande, Edifício C6, 1749-016 Lisbon, Portugal

Euler–Lagrange equation · Stochastic Noether's theorem · Schrödinger's problem ·
Second-order differential geometry

Mathematics Subject Classification 70L10 · 35F21 · 70H20 · 49Q22

Abbreviations

HJB equation	Hamilton–Jacobi–Bellman equation
MDE	Mean differential equations
SDE	Stochastic differential equations
S-EL equation	Stochastic Euler–Lagrange equation
S-H equations	Stochastic Hamilton's equations

List of Symbols

A	A general second-order differential operator or second-order vector field
A^X	Generator of the diffusion X
$\circ d$	Stratonovich stochastic differential
d	Exterior differential on the manifold M, or Itô stochastic differential
$\boldsymbol{d}$	Linear operator extended from the exterior differential on the tangent bundle TM
d^2	Second-order differential on M
d°	Mixed-order differential on $\mathbb{R} \times M$
d_x	Horizontal differential on the tangent bundle TM or cotangent bundle T^*M
$d_{\dot{x}}$	Vertical differential on TM
$(DX, QX), D_\nabla X, Q(X, Y)$	Mean derivatives
$\mathbf{D}_\mathrm{t}, \mathbf{Q}_\mathrm{t}$	Total mean derivatives
$\frac{\mathbf{D}}{dt}, \frac{\overline{\mathbf{D}}}{dt}$	Mean covariant derivative and damped mean covariant derivative
$\Delta, \Delta_\mathrm{LD}$	Connection Laplacian and Laplace–de Rham operator
F_*^S, F^{S*}	Second-order pushforward and pullback of a smooth map $F : M \to N$
F_*^R, F^{R*}	Mixed-order pushforward and pullback of F
Γ	Christoffel symbols or stochastic parallel displacement
$\overline{\Gamma}$	Damped parallel displacement
$I_t(M), I_{(t,q)}(M), I_{(t,q)}^{T,\mu}(M)$	Various sets of M-valued diffusions starting at time t
$j_q X, j_q^\nabla X, j_{(t,q)} X, j_t X$	Stochastic tangent vectors and stochastic jets
$\mathcal{L}$	Lie derivatives
∇	Linear connection, Levi–Civita connection, covariant derivative, or gradient on M
∇^2	Hessian operator
∇_p	Vertical gradient on T^*M

$(\Omega, \mathcal{F}, \mathbf{P})$	Probability space Ω with σ-field $\mathcal{F}$ and probability measure $\mathbf{P}$
$\{\mathcal{P}_t\}_{t\in\mathbb{R}}, \{\mathcal{F}_t\}_{t\in\mathbb{R}}$	Past (nondecreasing) filtration and future (nonincreasing) filtration
$\frac{\partial}{\partial t}, \partial_t$	Differential operator with respect to coordinate t
$\frac{\partial}{\partial x^i}, \partial_i$	Differential operator with respect to coordinate x^i
$\frac{\partial^2}{\partial x^j \partial x^k}, \partial_{jk}$	Second-order differential operator with respect to coordinates x^j and x^k
$\frac{\partial}{\partial p_i}, \partial_{p_i}$	Differential operator with respect to coordinate p_i
R, Ric	Riemann curvature tensor and Ricci $(1, 1)$-tensor
$\mathcal{T}^O M, \mathcal{T}^E M$	Second-order tangent bundle and second-order elliptic tangent bundle
$\mathcal{T}^S M$	Stochastic tangent bundle
$\mathcal{T}^{S*} M$	Second-order cotangent bundle
V	A general vector field
$(x^i, D^i x, Q^{jk} x)$	Canonical coordinates on $\mathcal{T}^S M$
(x^i, p_i, o_{jk})	Canonical coordinates on $\mathcal{T}^{S*} M$
X_*, X^*	Pushforward and pullback of the diffusion X
$\mathbf{X}$	A horizontal diffusion valued on a general bundle E or on $\mathcal{T}^{S*} M$
$\mathbb{X}$	A horizontal diffusion valued on $T^* M$

1 Introduction

Hamilton–Jacobi (HJ) partial differential equations and the associated theory lie at the center of classical mechanics (Abraham and Marsden 1978; Arnold 1989; Marsden and Ratiu 1999; Goldstein et al. 2002). Motivated by Hamilton's approach to geometrical optic where the action represents the time needed by a particle to move between two points and a variational principle due to Fermat, Jacobi extended this approach to Lagrangian and Hamiltonian mechanics. Jacobi designed a concept of "complete" solution of HJ equations allowing him to recover all solutions simply by substitutions and differentiations. Although, in general, it is more complicated to solve than a system of ODEs like Hamilton's ones, HJ equations proved to be powerful tools of integration of classical equations of motion. In addition, Jacobi's approach suggested him to ask what diffeomorphisms of the cotangent bundle, the geometric arena of canonical equations, preserve the structure of these first-order equations. Those are called today symplectic or canonical transformations, and Jacobi's method of integration is precisely one of them.

It is not always recognized as it should be that HJ equations were also fundamental in the construction of quantum mechanics. The reading of Schrödinger (1926), Fock (1978), Dirac (1933) and others until Feynman (1948) makes abundantly clear that most of new ideas in the field made use of HJ equations for the classical system to be "quantized," or some quantum deformation of them. There are at least two ways to express this deformation. On the one hand, one can exponentiate the L^2 wave func-

tion, call S its complex exponent and look for the equation solved by S (see Goldstein et al. 2002). When the system is a single particle in a scalar potential, one obtain the classical HJ equation with an additional Laplacian term and factor $i\hbar$, representing the regularization expected from the quantization of the system. This complex factor is symptomatic of the basic quantum probability problem, at least for pure states. In a nutshell, it is the reason why Feynman's diffusions, in his path integral approach, do not exist. On the other hand, there is an hydrodynamical interpretation of quantum mechanics, founded on Madelung transform, a polar representation of the wave function whose real part is the square root of a probability density. The argument solves another deformation of HJ equation. The geometry of this transform has been thoroughly investigated recently, highlighting its relations with optimal transport theory (Khesin et al. 2021; von Renesse 2012).

However, the probabilistic content of quantum mechanics, especially for pure states, remained a vexing mathematical mystery right from its beginning, despite several interesting (but unsuccessful) attempts (Nelson 2001). The current consensus is that regular probability theory and stochastic analysis have little or nothing to teach us about it. And, in particular, that all that can be saved from Feynman path integral theory is Wiener's measure and perturbations of it by potential terms. This is the "Euclidean approach," one of the starting points of mathematical quantum field theory.

In 1931, however, Schrödinger suggested in a paper almost forgotten until the 1980s (Schrödinger 1932) [but insightfully commented by the probabilist Bernstein (1932)] the existence of a completely different Euclidean approach to quantum dynamics. In short, a stochastic variational boundary value problem for probability densities characterizes optimal diffusions on a given time interval as having a density product of two positive solutions of time adjoint heat equations. This idea, revived and elaborated from 1986 (Zambrini 1986), is known today as "Schrödinger's problem" in the community of optimal transport, where it has proved to provide, among other results, very efficient regularization of fundamental problems of this field (Léonard 2014). In fact, Schrödinger's problem hinted toward the existence of a stochastic dynamical theory of processes, considerably more general than its initial quantum motivation. In it, various regularizations associated with the tools of stochastic calculus should play the role of those involved in quantum mechanics in Hilbert space, where the looked-for measures do not exist.

The variational side of the stochastic theory has been developed in the last decades, inspired by number of results in stochastic optimal control (Haussmann 1986; Fleming and Soner 2006) and stochastic optimal transport (Mikami 2021). In this context, the crucial role of (second-order) Hamilton–Jacobi–Bellman (HJB) equation has been known for a long time. It provides the proper regularization of the (first-order) HJ equation needed to construct well-defined stochastic dynamical theories. In contrast, for instance, with the notion of viscosity solution, whose initial target was the study of the classical PDE, HJB equation becomes central, there, as natural stochastic deformation of this one, compatible with Itô's calculus. It is worth mentioning that in any fields like AI or reinforcement learning, where HJB equations play a fundamental role (Peyré et al. 2019), it is natural to expect that such a stochastic dynamical framework, built on them, should present some interest.

The geometric, and especially, Hamiltonian side of the dynamical theory had resisted until now and constitutes the main contribution of this paper. It is our hope that it will be useful far beyond its initial motivation referred to, afterward, as its "inspirational examples." In this sense, it can clearly be interpreted as a general contribution to stochastic geometric mechanics. More precisely, we are trying to answer the following questions:

- Do we have any geometric interpretation of the Hamilton–Jacobi–Bellman equation? That is, can we derive the HJB equation from some sort of canonical transformations?
- Can we formulate some variational problem that leads to a Euler–Lagrange equation which is equivalent to the HJB equation?
- More systematically, can we develop some counterpart of Lagrangian and Hamiltonian mechanics that are associated with the HJB equation?

The first question indicates that canonical transformations should be somehow second-order, so that the corresponding symplectic and contact structures are also second-order. Meanwhile, the stochastic generalization of optimal control and optimal transport suggests that the variational problem of the second question should be formulated in stochastic sense. Combining these hints, the third question amounts to seeking a new theory of geometric mechanics that integrates stochastics and second-order together.

The cornerstone of stochastic analysis, the well-known Itô's formula, tells us that the generator of a diffusion process is a second-order differential operator. This provides a very natural way to connect the stochastics with the second-order. That is, in order to build a stochastic or second-order counterpart of geometric mechanics, we need to encode the rule of Itô's formula into the geometric structures.

There is a theory named second-order differential geometry ("stochastic differential geometry" is also used by some authors but we would like to keep the original terminology), which was devised by Schwartz and Meyer around 1980 (Schwartz 1980, 1982, 1984; Meyer 1979, 1981a), and later on developed by Belopolskaya and Dalecky (1990), Gliklikh (2011), Emery (1989), etc. See Emery (2007) for a survey of this aspect. Compared with the theory of stochastic analysis on manifolds (or geometric stochastic analysis) developed by Itô (1962, 1975), Malliavin (1997), Bismut (1981) and Elworthy (1982) etc., which focus on Stratonovich stochastic differential equations on classical geometric structures, like Riemannian manifolds, frame bundles and Lie groups, so that the Leibniz's rule is preserved, Schwartz' second-order differential calculus alter the underlying geometric structures to include second-order Itô correction terms, and provide a broader picture even though it loses Leibniz's rule and is much less known.

In this paper, we will adopt the viewpoint of Schwartz–Meyer and enlarge their picture to develop a theory of stochastic geometric mechanics. We first give an equivalent and more intuitive description for the second-order tangent bundle by equivalent classes of diffusions, via Nelson's mean derivatives. And then we generalize this idea to construct stochastic jets, from which stochastic prolongation formulae are proved and the stochastic counterpart of Cartan symmetries is studied. The second-order cotangent bundle is also studied, which helps us to establish stochastic Hamiltonian

mechanics. We formulate the stochastic Hamilton's equations, a system of stochastic equations on the second-order cotangent bundle in terms of mean derivatives. By introducing the second-order symplectic structure and the mixed-order contact structure, we derive the second-order HJB equations via canonical transformations. Finally, we set up a stochastic variational problem on the space of diffusion processes, also in terms of mean derivatives. Two kinds of stochastic principle of least action are built: stochastic Hamilton's principle and stochastic Maupertuis's principle. Both of them yield a stochastic Euler–Lagrange equation. The equivalence between the stochastic Euler–Lagrange equation and the HJB equation is proved, which exactly leads to the equivalence between our stochastic variational problem and Schrödinger's problem in optimal transport. Last but not least (actually vital), a stochastic Noether's theorem is proved. It says that every symmetry of HJB equation corresponds to a martingale that is exactly a conservation law in the stochastic sense. It should be observed, however, that the Schwartz–Meyer approach, together with the one of Bismut (1981), has also inspired a distinct, Stratonovich-type stochastic Hamiltonian framework (Lázaro-Camí and Ortega 2008) leading to a stochastic HJ equation (Lázaro-Camí and Ortega 2009), without relations with Schrödinger's problem or optimal transport.

The key results of the present paper and the dependence among them are briefly expressed in the following diagram:
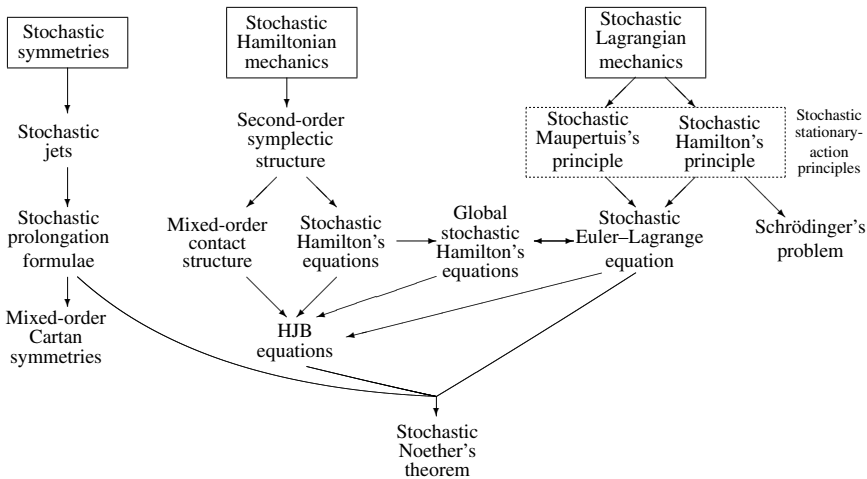
The organization of this paper is the following:

Section 2 is a summary on the theory of stochastic differential equations on manifolds, in the perspective appropriate to our goal. In particular, diffusions will be characterized by their mean and quadratic mean derivatives as in Nelson's stochastic mechanics (Nelson 2001) although the resulting dynamical content of our theory will have very little to do with his. In this way, we are able to rewrite Itô SDEs on manifolds as ODE-like equations that have better geometric nature. The notion of second-order tangent bundle answers to the question: the drift parts of Itô SDEs are sections of what?

Section 3 is devoted to the notion of Stochastic jets. In the same way as tangent vector on M are defined as equivalence classes of smooth curves through a given point and then generalized to higher-order cases to produce the notion of jets, the stochastic tangent vector is defined as equivalence classes of diffusions so that the stochastic tangent bundle is isomorphic to the elliptic subbundle of the second-order tangent bundle. Stochastic jets are also constructed. This provides an intrinsic definition of SDEs under consideration.

Section 4 illustrates the use of the above geometric formulation of SDEs for the study of their symmetries. Prolongations of M-valued diffusions are defined as new processes with values on the stochastic tangent bundle. Among all deterministic space-time transformations, bundle homomorphisms will be the only subclass to transform diffusions to diffusions. Total mean and quadratic derivative are defined in conformity with the rules of Itô's calculus. The prolongation of diffusions allows to define symmetries of SDEs and their infinitesimal versions. Stochastic prolongation formulae are derived for infinitesimal symmetries, which yield determining equations for Itô SDEs.

In Sect. 5, the second-order cotangent bundle, as dual bundle of second-order tangent bundle, is defined and analyzed. The properties of second-order differential operator, pushforwards and pullbacks are described. When time is involved, i.e., the base manifold is the product manifold $\mathbb{R} \times M$, the corresponding bundles are mixed-order tangent and cotangent bundles, where "mixed-order" means they are second-order in space but first-order in time. More about this topic, like mixed-order pushforwards and pullbacks, pushforwards and pullbacks by diffusions, and Lie derivatives, can be found in "Appendix A." An generalized notion to stochastic Cartan distribution and its symmetries are discussed in "Appendix B" based on the mixed-order contact structure.

The point of Sect. 6 is to use the tools developed before in the construction of the stochastic Hamiltonian mechanics which is one of the main goals of the paper. One of our inspirational example will be the one underlying the dynamical content of Schrödinger's problem. By analogy with Poincaré 1-form in the cotangent bundle of classical mechanics and its associated symplectic form, one can construct counterparts in the second-order cotangent bundle. Using the canonical second-order symplectic form on second-order cotangent bundles, one defines second-order symplectomorphisms. The generalization of classical Hamiltonian vector fields becomes second-order operators, for a given real-valued Hamiltonian function on the second-order cotangent bundle. The resulting stochastic Hamiltonian system involves pairs of extra equations compared with their classical versions. Bernstein's reciprocal processes inspired by Schrödinger's problem are described in this framework, corresponding to a large class of second-order Hamiltonians on Riemannian manifolds. A mixed-order contact structure describes time-dependent stochastic Hamiltonian systems. The last subsection of this section is devoted to canonical transformations preserving the form of stochastic Hamilton's equations. The corresponding generating function satisfies the Hamilton–Jacobi–Bellman equation.

Section 7 treats the stochastic version of classical Lagrangian mechanics on Riemannian manifolds. Itô's stochastic deformation of the classical notion of parallel displacements are recalled. Another one, called damped parallel displacement in the mathematical literature, involving the Ricci tensor, is also indicated. Each of

these displacements corresponds to a mean covariant derivative along diffusions. The action functional is defined as expectation of Lagrangian and the stochastic Euler–Lagrange equation involves the damped mean covariant derivative. The dynamics of Schrödinger's problem is, again, used as illustration. The equivalence between stochastic Hamilton's equations on Riemannian manifolds and the stochastic Euler–Lagrange one as well as the HJB equation are derived via the Legendre transform. Relations with stochastic control are also mentioned. The section ends with the stochastic Noether's theorem. The stochastic version of Maupertuis principle, as the twin of stochastic Hamilton's principle, is left into "Appendix C."

2 Stochastic Differential Equations on Manifolds

In this section, we will study several types of stochastic differential equations on manifolds which are weakly equivalent to Itô SDEs. We start with a d-dimensional smooth manifold M and a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and equip the latter with a filtration $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$, i.e., a family of nondecreasing sub-σ-fields of $\mathcal{F}$. We call $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$ a *past filtration*. Unless otherwise specified, the manifold M will not be endowed with any structures other than the smooth structure. In some cases, it will be endowed with a linear connection, a Riemannian metric, or a Levi–Civita connection.

Recall from Hsu (2002, Definition 1.2.1) that by an M-valued (forward) $\{\mathcal{P}_t\}$-*semimartingale*, we mean a $\{\mathcal{P}_t\}$-adapted continuous M-valued process $X = \{X(t)\}_{t \in [t_0, \tau)}$, where $t_0 \in \mathbb{R}$ and τ is a $\{\mathcal{P}_t\}$-stopping time satisfying $t_0 < \tau \leq +\infty$, such that $f(X)$ is a real-valued $\{\mathcal{P}_t\}$-semimartingale on $[t_0, \tau)$ for all $f \in C^\infty(M)$. The stopping time τ is called the lifetime of X. If we adopt the convention to introduce the one-point compactification of M by $M^* := M \cup \{\partial_M\}$, then the process X can be extended to the whole time line $[t_0, +\infty)$ by setting $X(t) = \partial_M$ for all $t \geq \tau$. The point ∂_M is often called the cemetery point in the context of Markovian theory.

2.1 Itô SDEs on Manifolds

Given $N + 1$ time-dependent vector field b, σ_r, $r = 1, \ldots, N$ on M, one can introduce a Stratonovich SDE in local coordinates, which has the same form as in Euclidean space (Hsu 2002, Section 1.2). The form of Stratonovich SDEs on M is invariant under changes of coordinates, as Stratonovich stochastic differentials obey the Leibniz's rule.

However, for Itô stochastic differentials this is not the case because of Itô's formula. Hence, we cannot directly write an Euclidean form of Itô SDE on M in local coordinates, since it is no longer invariant under changes of coordinates. Indeed, a change of coordinates will always produce an additional term. To balance this term, a common way is to add a correction term to the drift part of the Euclidean form of Itô SDE, by taking advantage of a linear connection. More precisely, under local coordinates (x^i), we consider the following Itô SDE (Gliklikh 2011, Section 7.1, 7.2):

$$dX^i(t) = \left[b^i(t, X(t)) - \frac{1}{2} \sum_{r=1}^{N} \Gamma_{jk}^i(X(t))(\sigma_r^j \sigma_r^k)(t, X(t)) \right] dt$$
$$+ \sigma_r^i(t, X(t)) dW^r(t), \tag{2.1}$$

where (Γ_{jk}^i) is the family of Christoffel symbols for a given linear connection ∇ on TM. When conditioning on $\{X(t) = q\}$ and taking (x^i) as normal coordinates at $q \in M$, (2.1) turns to the Euclidean form, since at q,

$$\sum_{r=1}^{N} \Gamma_{jk}^i \sigma_r^j \sigma_r^k = \frac{1}{2} \sum_{r=1}^{N} \left(\Gamma_{jk}^i + \Gamma_{kj}^i \right) \sigma_r^j \sigma_r^k = 0. \tag{2.2}$$

If we denote

$$\sigma \circ \sigma^* := \sum_{r=1}^{N} \sigma_r \otimes \sigma_r = \sum_{r=1}^{N} \sigma_r^j \sigma_r^k \frac{\partial}{\partial x^j} \otimes \frac{\partial}{\partial x^k}.$$

Then, clearly $\sigma \circ \sigma^*$ is a symmetric and positive semi-definite $(2,0)$-tensor field. We also introduce formally a modified drift $\mathfrak{b}$ which has the following coordinate expression

$$\mathfrak{b}^i = b^i - \frac{1}{2} \sum_{r=1}^{N} \Gamma_{jk}^i \sigma_r^j \sigma_r^k. \tag{2.3}$$

We change the coordinate chart from $(U, (x^i))$ to $(V, (\tilde{x}^j))$ with $U \cap V \neq \emptyset$. Since each σ_r transforms as a vector, we apply the change-of-coordinate formula for Christoffel symbols (e.g., Kobayashi and Nomizu 1963, Proposition III.7.2) to derive that

$$\Gamma_{jk}^i \sigma_r^j \sigma_r^k = \left(\tilde{\Gamma}_{mn}^l \frac{\partial \tilde{x}^m}{\partial x^j} \frac{\partial \tilde{x}^n}{\partial x^k} \frac{\partial x^i}{\partial \tilde{x}^l} + \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} \frac{\partial x^i}{\partial \tilde{x}^l} \right) \sigma_r^j \sigma_r^k$$
$$= \left(\tilde{\Gamma}_{mn}^l \tilde{\sigma}_r^m \tilde{\sigma}_r^n + \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} \sigma_r^j \sigma_r^k \right) \frac{\partial x^i}{\partial \tilde{x}^l}.$$

It follows that the coefficients of the modified drift $\mathfrak{b}$ in (2.3) transform as

$$\tilde{\mathfrak{b}}^l = \tilde{b}^l - \frac{1}{2} \sum_{r=1}^{N} \tilde{\Gamma}_{mn}^l \tilde{\sigma}_r^m \tilde{\sigma}_r^n = b^i \frac{\partial \tilde{x}^l}{\partial x^i} - \frac{1}{2} \sum_{r=1}^{N} \left(\Gamma_{jk}^i \frac{\partial \tilde{x}^l}{\partial x^i} - \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} \right) \sigma_r^j \sigma_r^k$$
$$\tag{2.4}$$
$$= \mathfrak{b}^i \frac{\partial \tilde{x}^l}{\partial x^i} + \frac{1}{2} \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} \sum_{r=1}^{N} \sigma_r^j \sigma_r^k.$$

Therefore, $\mathfrak{b}$ is *not* a vector field as it does not pointwisely transform as a vector.

Finally, using Itô's formula, we derive the transformation of (2.1) as follows:

$$
\begin{aligned}
d\tilde{x}^l &= \frac{\partial \tilde{x}^l}{\partial x^i} dx^i + \frac{1}{2} \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} d[x^j, x^k] \\
&= \left[\frac{\partial \tilde{x}^l}{\partial x^i} \left(b^i - \frac{1}{2} \sum_{r=1}^{N} \Gamma_{jk}^i \sigma_r^j \sigma_r^k \right) + \frac{1}{2} \sum_{r=1}^{N} \frac{\partial^2 \tilde{x}^l}{\partial x^j \partial x^k} \sigma_r^j \sigma_r^k \right] dt + \frac{\partial \tilde{x}^l}{\partial x^i} \sigma_r^i dW^r \\
&= \left(\tilde{b}^l - \frac{1}{2} \sum_{r=1}^{N} \tilde{\Gamma}_{mn}^l \tilde{\sigma}_r^m \tilde{\sigma}_r^n \right) dt + \tilde{\sigma}_r^l dW^r,
\end{aligned}
$$

where the bracket $[\cdot, \cdot]$ on the right-hand side (RHS) of the first equality denotes the quadratic variation. This shows that Eq. (2.1) is indeed invariant under changes of coordinates.

Remark 2.1 One can regard $\sigma = (\sigma_r)_{r=1}^{N} \in (\mathbb{R}^N)^* \otimes \mathfrak{X}(M)$ as an $(\mathbb{R}^N)^*$-valued vector field on M. In this way, the pair (b, σ) is called an Itô vector field in Gliklikh (2011, Chapter 7), while the pair $(\mathfrak{b}, \sigma)$ is called an Itô equation therein.

Now we present the definition of weak solutions to (2.1).

Definition 2.2 (*Weak solutions to Itô SDEs*) Given a linear connection on M, a weak solution of the Itô SDE (2.1) is a triple (X, W), $(\Omega, \mathcal{F}, \mathbf{P})$, $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$, where

(i) $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$ is a past (i.e., nondecreasing) filtration of $\mathcal{F}$ satisfying the usual conditions,

(ii) $X = \{X(t)\}_{t \in [t_0, \tau)}$ is a continuous, $\{\mathcal{P}_t\}$-adapted M-valued process with $\{\mathcal{P}_t\}$-stopping time $\tau > t_0$, W is an N-dimensional $\{\mathcal{P}_t\}$-Brownian motion, and

(iii) for every $q \in M, t \geq t_0$ and any coordinate chart $(U, (x^i))$ of q, it holds under the conditional probability $\mathbf{P}(\cdot | X(t_0) = q)$ that almost surely in the event $\{X(t) \in U\}$,

$$
\begin{aligned}
X^i(t) = X^i(t_0) &+ \int_{t_0}^{t} \left(b^i(s, X(s)) - \frac{1}{2} \sum_{r=1}^{N} \Gamma_{jk}^i(X(s)) \left(\sigma_r^j \sigma_r^k \right)(s, X(s)) \right) ds \\
&+ \int_{t_0}^{t} \sigma_r^i(s, X(s)) dW^r(s).
\end{aligned}
$$

Definition 2.3 (*Uniqueness in law*) We say that uniqueness in the sense of probability law holds for the Itô SDE (2.1) if, for any two weak solutions (X, W), $(\Omega, \mathcal{F}, \mathbf{P})$, $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$, and $(\hat{X}, \hat{W})$, $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{P}})$, $\{\hat{\mathcal{P}}_t\}_{t \in \mathbb{R}}$ with the same initial data, i.e., $\mathbf{P}(X(0) = x_0) = \hat{\mathbf{P}}(\hat{X}(0) = x_0) = 1$, the two processes X and $\tilde{X}$ have the same law.

Note that it is possible to change σ and W in the Itô SDE (2.1) but keep the same weak solution in law. In other words, the form of (2.1) does not univocally correspond to its weak solution in law. For this reason, we will reformulate SDEs in a fashion that makes them look more like ODEs and have better geometric nature. Moreover, we will see that it is the pair $(\mathfrak{b}, \sigma \circ \sigma^*)$ that univocally corresponds to the weak solution of (2.1).

2.2 Mean Derivatives and Mean Differential Equations on Manifolds

In this part, we will recall the definitions of Nelson's mean derivatives and extend them to M-valued processes. In Nelson's stochastic mechanics (Nelson 2001), the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is equipped with two different filtrations. The first one is just an usual nondecreasing filtration $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$, a past filtration. The second is a family of nonincreasing sub-σ-fields of $\mathcal{F}$, which is denoted by $\{\mathcal{F}_t\}_{t \in \mathbb{R}}$ and called a *future* filtration. For an $\mathbb{R}^d$-valued process $\{X(t)\}_{t \in I}$, its forward mean derivative DX and forward quadratic mean derivative QX are defined by conditional expectations as follows:

$$DX(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{X(t+\epsilon) - X(t)}{\epsilon}\bigg|\mathcal{P}_t\right],$$

$$QX(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{(X(t+\epsilon) - X(t)) \otimes (X(t+\epsilon) - X(t))}{\epsilon}\bigg|\mathcal{P}_t\right],$$

Their backward versions, i.e., the backward mean derivative and backward quadratic mean derivative, are defined as follows:

$$\overleftarrow{D} X(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{X(t) - X(t-\epsilon)}{\epsilon}\bigg|\mathcal{F}_t\right],$$

$$\overleftarrow{Q} X(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{(X(t) - X(t-\epsilon)) \otimes (X(t) - X(t-\epsilon))}{\epsilon}\bigg|\mathcal{F}_t\right].$$

In our present paper, we will only focus on the "forward" case, so that only the past filtration $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$ will be invoked. The "backward" case is analogous and every part of this paper can have its "backward" counterpart (cf. Zambrini 2015).

Denote by $\text{Sym}^2(TM)$ (and $\text{Sym}^2_+(TM)$) the fiber bundle of symmetric (and respectively, symmetric positive semi-definite) $(2, 0)$-tensors on M. Now we define quadratic mean derivatives for M-valued semimartingales, cf. Gliklikh (2011, Chapter 9).

Definition 2.4 (*Quadratic mean derivatives*) The (forward) quadratic mean derivative of the M-valued semimartingale $\{X(t)\}_{t \in [t_0, \tau)}$ is a $\text{Sym}^2_+(TM)$-valued process QX on $[t_0, \tau)$, whose value at time $t \in [t_0, \tau)$ in any coordinate chart $(U, (x^i))$ and in the event $\{X(t) \in U\}$ is given by

$$(QX)^{ij}(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{(X^i(t+\epsilon) - X^i(t))(X^j(t+\epsilon) - X^j(t))}{\epsilon}\bigg|\mathcal{P}_t\right], \quad (2.5)$$

where the limits are assumed to exist in $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

More generally, we can define the (forward) quadratic mean derivative for two M-valued semimartingales X and Y in local coordinates by

$$(Q(X, Y))^{ij}(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{(X^i(t+\epsilon) - X^i(t))(Y^j(t+\epsilon) - Y^j(t))}{\epsilon}\bigg|\mathcal{P}_t\right].$$

Due to Itô's formula for semimartingales, $QX(t)$ does transform as a $(2, 0)$-tensor and is obviously symmetric, so that the definition is independent of the choice of U. However, the formal limit $\mathbf{E}[\frac{1}{\epsilon}(X^i(t + \epsilon) - X^i(t))|\mathcal{P}_t]$ under any coordinates (x^i), no longer transforms as a vector, as can be guessed from (2.4). In order to turn it into a vector we need to specify a coordinate system. A natural choice is the normal coordinate system. For this purpose, we endow M with a linear connection ∇, which determines a normal coordinate system near each point on M.

Definition 2.5 (∇-*mean derivatives*) Given a linear connection ∇ on M, the (forward) ∇-mean derivative of the M-valued semimartingale $\{X(t)\}_{t\in[t_0,\tau)}$ is a TM-valued process $D_\nabla X$ on $[t_0, \tau)$, whose value at time $t \in [t_0, \tau)$ is defined in *normal* coordinates (x^i) on the normal neighborhood U of $q \in M$ and under the conditional probability $\mathbf{P}(\cdot|X(t) = q)$ as follows:

$$(D_\nabla X)^i(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{X^i(t + \epsilon) - X^i(t)}{\epsilon}\bigg|\mathcal{P}_t\right],$$

where the limits are assumed to exist in $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

As we force $D_\nabla X(t)$ to be vector-valued by definition, its coordinate expression under any other coordinate system can be calculated via Leibniz's rule. Let us stress that the notation D_∇ should not be confused with the one of covariant derivatives in geometry.

Now we formally take forward mean derivatives in Itô SDE (2.1), and note that the correction term in the modified drift involving Christoffel symbols vanishes by (2.2). Then, we get an ODE-like system:

$$\begin{cases} D_\nabla X(t) = b(t, X(t)), \\ QX(t) = (\sigma \circ \sigma^*)(t, X(t)). \end{cases} \tag{2.6}$$

We call Eq. (2.6) a system of mean differential equations (MDEs). Note that both MDEs (2.6) and Itô SDE (2.1) rely on linear connections on M.

Definition 2.6 (*Solutions to MDEs*) Given a linear connection on M, a solution of MDEs (2.6) is a triple X, $(\Omega, \mathcal{F}, \mathbf{P})$, $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$, where

(i) $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is a past filtration of $\mathcal{F}$ satisfying the usual conditions,
(ii) $X = \{X(t)\}_{t\in[t_0,\tau)}$ is a continuous, $\{\mathcal{P}_t\}$-adapted M-valued semimartingale with lifetime a $\{\mathcal{P}_t\}$-stopping time $\tau > t_0$, and
(iii) the ∇-mean derivative and quadratic mean derivative of X exist and satisfy (2.6).

2.3 Second-Order Operators and Martingale Problems

Definition 2.7 (*Second-order operators*) A second-order operator on M is a linear operator $A : C^\infty(M) \to C^\infty(M)$, which has the following expression in a coordinate

chart $(U, (x^i))$,

$$Af = A^i \frac{\partial f}{\partial x^i} + A^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j}, \quad f \in C^\infty(M), \tag{2.7}$$

where (A^{ij}) is a symmetric $(2, 0)$-tensor field, and the expression is required to be invariant under changes of coordinates. If (A^{ij}) is positive semi-definite, then we say the second-order operator A is elliptic; if (A^{ij}) is positive definite, we say A is nondegenerate elliptic.

There is a coordinate-free definition of second-order operators. A linear map $A_q : C^\infty(M) \to \mathbb{R}$ is called a second-order derivation at $q \in M$, if there is a symmetric $(2, 0)$-tensor Γ_{A_q} at q such that $A_q(fg) = f(q)A_q g + g(q)A_q f + (df \otimes dg)(\Gamma_{A_q})$ for all $f, g \in C^\infty(M)$. Then, a second-order operator is nothing but a smooth field of second-order derivations. From this, we see that for A in (2.7), $A^i = A(x^i)$, $A^{ij} = A(x^i x^j) - x^i A(x^j) - x^j A(x^i)$, and

$$\Gamma_A = A^{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}. \tag{2.8}$$

We call Γ_A the *squared field operator* (originally "opérateur carré du champ") associated with A. We also denote $\Gamma_A(f, g) := (df \otimes dg)(\Gamma_A)$. Clearly, for a classical vector field V, $\Gamma_V \equiv 0$ by Leibniz's rule.

It is easy to verify from the coordinate-change invariance that the coefficients A^i's and A^{ij}'s transform under the change of coordinates from (x^i) to $(\tilde{x}^j)$ by the following rule (e.g., Ikeda and Watanabe 1989, Section V.4),

$$\tilde{A}^i = \frac{\partial \tilde{x}^i}{\partial x^j} A^j + \frac{\partial^2 \tilde{x}^i}{\partial x^j \partial x^k} A^{jk}, \quad \tilde{A}^{ij} = \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial \tilde{x}^j}{\partial x^l} A^{kl}. \tag{2.9}$$

The formal generator of Itô SDE (2.1) is given by,

$$A_t^X = \mathfrak{b}^i(t) \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{r=1}^N \sigma_r^i(t)\sigma_r^j(t) \frac{\partial^2}{\partial x^i \partial x^j}, \tag{2.10}$$

which is a time-dependent second-order elliptic operator due to the change-of-coordinate formula (2.4).

Denote by $\mathcal{C}_{t_0}$ the subspace of $C([t_0, \infty), M^*)$ consisting of all paths always staying in M or eventually stopped at ∂_M. That is, $\omega \in \mathcal{C}_{t_0}$ if and only if there exists $\tau(\omega) \in (t_0, \infty]$ such that $\omega(t) \in M$ for $t \in [t_0, \tau(\omega))$ and $\omega(t) = \partial_M$ for $t \in [\tau(\omega), \infty)$. Let $\mathcal{B}(\mathcal{C}_{t_0})$ be the σ-field generated by Borel cylinder sets. Let $X(t) : \mathcal{C}_{t_0} \to M^*$, $X(t, \omega) = \omega(t)$, $t \geq t_0$ be the coordinate mapping. For each $t \in \mathbb{R}$, define a sub-σ-field by $\mathcal{B}_t = \sigma\{X(s) : t_0 \leq s \leq t_0 \vee t\}$. Then, $\{\mathcal{B}_t\}_{t \in \mathbb{R}}$ is a past filtration of $\mathcal{B}(\mathcal{C}_{t_0})$ and τ is a $\{\mathcal{B}_t\}$-stopping time.

Definition 2.8 (*Martingale problems on manifolds,* Hsu 2002, Definition 1.3.1) Given a time-dependent second-order elliptic operator $A = (A_t)_{t \geq t_0}$, a solution to the martingale problem associated with A is a triple X, $(\Omega, \mathcal{F}, \mathbf{P})$, $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$, where

(i) $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and $\{\mathcal{P}_t\}_{t \in \mathbb{R}}$ is a past filtration of $\mathcal{F}$ satisfying the usual conditions,
(ii) $X : \Omega \to \mathcal{C}_{t_0}$ is an M^*-valued $\{\mathcal{P}_t\}$-semimartingale, and
(iii) for every $f \in C^{\infty}(\mathbb{R} \times M)$, the process $M^{f,X}(t) := f(t, X(t)) - f(t_0, X(t_0)) - \int_{t_0}^t (\frac{\partial}{\partial t} + A_s) f(s, X(s)) \mathrm{d}s$, $t \in [t_0, \tau(X))$, is a real-valued continuous $\{\mathcal{P}_t\}$-martingale.

The process $\{X(t)\}_{t \in [t_0, \tau(X))}$ is called an M-valued $\{\mathcal{P}_t\}$-diffusion process with generator A (or simply an A-diffusion).

The uniqueness in the sense of probability law for both MDEs and martingale problems can be defined in a similar fashion to Definition 2.3. Note that unlike Itô SDEs or MDEs, the definition for martingale problems does not rely on linear connections.

When provided with a linear connection on M, one can see, in the same way as in Stroock and Varadhan's theory (e.g., Karatzas and Shreve 1991, Section 5.4), that the existence of a solution to the martingale problem associated with $A^X = (A_t^X)_{t \geq t_0}$ in (2.10) is equivalent to the existence of a weak solution to the Itô SDE (2.1), and also equivalent to the existence of a solution to MDEs (2.6); their uniqueness in law of are also equivalent.

2.4 The Second-Order Tangent Bundle

As we have seen, the modified drift $\mathfrak{b}$ in (2.3) is not a vector field. Is $\mathfrak{b}$ a section (and, in the affirmative, of what)? In fact, it is not a section of any bundle, as its changes-of-coordinate formula (2.4) involves σ. But if we look at the formal generator A^X in (2.10), or the pair $(\mathfrak{b}, \sigma \circ \sigma^*)$ of its coefficients, then we can construct a bundle whose structure group is governed by the changes-of-coordinate formulae (2.9), so that the sections are just second-order operators.

We denote by $\mathrm{Sym}^2(\mathbb{R}^d)$ the space of all symmetric $(2, 0)$-tensors on $\mathbb{R}^d$, and by $\mathrm{Sym}_+^2(\mathbb{R}^d)$ the subspace of it consisting of all positive semi-definite $(2, 0)$-tensors. Also denote by $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^d)$ the space of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}^d$.

Definition 2.9 (*The second-order tangent bundle*)

(i) Gliklikh (2011, Definition 7.14) The Itô group G_I^d is the Cartesian product (but not direct product of groups) $\mathrm{GL}(d, \mathbb{R}) \times \mathcal{L}(\mathbb{R}^d \otimes \mathbb{R}^d, \mathbb{R}^d)$ equipped with the following binary operation:

$$(g_2, \kappa_2) \circ (g_1, \kappa_1) = (g_2 \circ g_1, g_2 \circ \kappa_1 + \kappa_2 \circ (g_1 \otimes g_1)),$$

for all $g_1, g_2 \in \mathrm{GL}(d, \mathbb{R})$, $\kappa_1, \kappa_2 \in \mathcal{L}(\mathbb{R}^d \otimes \mathbb{R}^d, \mathbb{R}^d)$.
(ii) The left group action of G_I^d on $\mathbb{R}^d \times \mathrm{Sym}^2(\mathbb{R}^d)$ is defined by

$$(g, \kappa) \cdot (\mathfrak{b}, a) = (g\mathfrak{b} + \kappa a, (g \otimes g)a), \tag{2.11}$$

for all $(g, \kappa) \in G_I^d$, $\mathfrak{b} \in \mathbb{R}^d$, $a \in \text{Sym}^2(\mathbb{R}^d)$.

(iii) The second-order tangent bundle $(\mathcal{T}^O M, \tau_M^O, M)$ is the fiber bundle with base space M, typical fiber $\mathbb{R}^d \times \text{Sym}^2(\mathbb{R}^d)$, and structure group G_I^d.

(iv) The fiber $\mathcal{T}_q^O M$ at $q \in M$ is called second-order tangent space to M at q. An element $(\mathfrak{b}, a)_q \in \mathcal{T}_q^O M$ is called a second-order tangent vector at q. A (global or local) section of τ_M^O is called a second-order vector field.

(v) Denote by $\mathcal{T}^E M$ the subbundle of $\mathcal{T}^O M$ consisting of all elements $(\mathfrak{b}, a)_q \in \mathcal{T}_q^O M, q \in M$, with a_q a positive semi-definite (2, 0)-tensors. Let $\tau_M^E = \tau_M^O|_{\mathcal{T}^E M}$. We call $(\mathcal{T}^E M, \tau_M^E, M)$ the second-order elliptic tangent bundle.

Remark 2.10 (i) We indulge in some abuse of notions. For example, the second-order vector fields should not be confused with the semisprays which are sections of the double tangent bundle $T^2 M$ (e.g., Saunders 1989, Section 1.4; Lang 1999, Section IV.3).

(ii) Some authors just defined second-order vector fields as second-order operators as in Definition 2.7 (Emery 1989, Definition 6.3 or Gliklikh 2011, Definition 2.74). As soon as we choose a frame for $\mathcal{T}^O M$, it will be clear that second-order vector fields are identified with second-order operators.

(iii) The authors in Belopolskaya and Dalecky (1990), Gliklikh (2011) define a bundle which has the Itô group as its structure group and has the pair $(\mathfrak{b}, \sigma)$ of coefficients in Itô SDE (2.1) as its section. They name it Itô's bundle and denote it as $\mathcal{I} M$. The difference is that, in our formulation, the pair $(\mathfrak{b}, \sigma \circ \sigma^*)$ of coefficients of the generator of Itô SDE (2.1) is a section of second-order elliptic tangent bundle τ_M^E. The advantage of the bundle τ_M^E is that it is a natural generalization of tangent bundle to second-order and has a good geometric interpretation, as we will see in Proposition 3.2.

(iv) Note that the typical fiber $\mathbb{R}^d \times \text{Sym}^2(\mathbb{R}^d)$ of τ_M^O is a vector space of dimension $d + \frac{d(d+1)}{2}$. But τ_M^E is not a vector bundle, since its structure group G_I^d is not a linear group (subgroup of general linear group). The typical fiber of τ_M^E is $\mathbb{R}^d \times \text{Sym}_+^2(\mathbb{R}^d)$, which is not even a vector space, so that τ_M^E is not a vector bundle either. Indeed, we may call them quadratic bundles, just as the way they call Itô's bundle in Belopolskaya and Dalecky (1990, Chapter 4).

(v) The Itô's bundle $\mathcal{I} M$ defined in Gliklikh (2011, Definition 7.17) is the fiber bundle over manifold M, with fiber $\mathbb{R}^d \times \mathcal{L}(\mathbb{R}^N, \mathbb{R}^d)$ and structure group G_I^d which acts on the fiber from the left by

$$(g, \kappa)(\mathfrak{b}, \sigma) = \left(g\mathfrak{b} + \tfrac{1}{2}\text{tr}\left(\kappa \circ (\sigma \otimes \sigma)\right), g \circ \sigma\right),$$

for all $(g, \kappa) \in G_I^d$, $\mathfrak{b} \in \mathbb{R}^d$, $\sigma \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^d)$. For the same reason as $\mathcal{T}^O M$ or $\mathcal{T}^E M$, Itô's bundle $\mathcal{I} M$ is not a vector bundle. There is a bundle homomorphism over M from $\mathcal{I} M$ to $\mathcal{T}^E M$, which maps in fibers from $\mathcal{I}_q M$ to $\mathcal{T}_q^E M$, $q \in M$, by $(\mathfrak{b}, \sigma) \to (\mathfrak{b}, \sigma \circ \sigma^*)$. It is easy to see that this bundle homomorphism is also a subjective submersion. If we identify $g \in \text{GL}(d, \mathbb{R})$ with $(g, 0) \in G_I^d$, then $\text{GL}(d, \mathbb{R})$ is a subgroup of G_I^d. We define the *Stratonovich's bundle* $\mathcal{S} M$ to be the reduction of $\mathcal{I} M$ to the structure group $\text{GL}(d, \mathbb{R})$, that is, the fiber bundle over M,

with fiber $\mathbb{R}^d \times \mathcal{L}(\mathbb{R}^N, \mathbb{R}^d)$ and structure group $\mathrm{GL}(d, \mathbb{R})$ which acts on the fiber from the left by

$$g(\mathfrak{b}, \sigma) = (g\mathfrak{b}, g \circ \sigma).$$

Unlike $\mathcal{T}^O M$ or $\mathcal{I}M$, Stratonovich's bundle $\mathcal{S}M$ is indeed a vector bundle, and the tangent bundle TM is a vector subbundle of $\mathcal{S}M$. It can be expected that Stratonovich's bundle is a natural bundle to formulate Stratonovich SDEs. But, in this paper, we mainly focus on Itô SDEs and their generators.

It is natural to regard the differential operators

$$\left\{ \frac{\partial}{\partial x^i}, \frac{\partial^2}{\partial x^j \partial x^k} : 1 \le i \le d, 1 \le j \le k \le d \right\} \tag{2.12}$$

as a local frame of $\mathcal{T}^O M$ over the local chart $(U, (x^i))$ on M. In the sequel, we will usually shorten them by

$$\left\{ \partial_i, \ \partial_j \partial_k : 1 \le i \le d, 1 \le j \le k \le d \right\}.$$

We make the convention that $\partial_k \partial_j = \partial_j \partial_k$ for all $1 \le j \le k \le d$. A second-order vector field $(\mathfrak{b}, a)$ is expressed in terms of this local frame by

$$(\mathfrak{b}, a) = \mathfrak{b}^i \partial_i + \tfrac{1}{2} a^{jk} \partial_j \partial_k.$$

In this way, every second-order vector field can be regarded as a second-order operator and vice versa. In particular, the generator A^X of an M-valued diffusion process X, for example the generator (2.10) of the Itô SDE, is a time-dependent second-order vector field, so that we can rewrite A^X as $A_t^X = (\mathfrak{b}(t), (\sigma \circ \sigma^*)(t))$.

The tangent bundle TM is a subbundle (but not a vector subbunddle) and also an embedded submanifold of $\mathcal{T}^O M$, as the bundle monomorphism

$$\iota : (TM, \tau_M, M) \to \left(\mathcal{T}^O M, \tau_M^O, M\right), \quad v_q \mapsto (v, 0)_q \tag{2.13}$$

is also an embedding. However, there is no canonical bundle epimorphism from $\mathcal{T}^O M$ to TM which is a left inverse of ι and linear in fiber. We call such a bundle epimorphism a *fiber-linear bundle projection* from $\mathcal{T}^O M$ to TM. The choice of such a bundle epimorphism is exactly the choice of a linear connection on M. More precisely, we have the following connection correspondence properties, the first of which can also be found in Gliklikh (2011, Section 2.9).

Proposition 2.11 (Connection correspondence) *Any linear connection on M induces a fiber-linear bundle projection from $\mathcal{T}^O M$ to TM. Conversely, any fiber-linear bundle projection from $\mathcal{T}^O M$ to TM induces a torsion-free linear connection on M.*

Remark 2.12 The connection correspondence is similar to the correspondence between horizontal subbundles of the tangent bundle of a vector bundle and connections on this vector bundle, cf. Saunders (1989, Section 3.1).

Proof Let (Γ^k_{ij}) be the Christoffel symbols of a linear connection ∇ on M. Define a projection by

$$\varrho_\nabla : \mathcal{T}^O M \to TM, \quad (\mathfrak{b}, a)_q \mapsto \left(\mathfrak{b}^i + \tfrac{1}{2}a^{jk}\Gamma^i_{jk}(q)\right)\partial_i\big|_q. \qquad (2.14)$$

Clearly, ϱ_∇ is linear in fiber and $\varrho_\nabla \circ \iota = \mathbf{Id}_{TM}$. Conversely, let $\varrho : \mathcal{T}^O M \to TM$ be a fiber-linear bundle projection. Then, on each coordinate chart $(U, (x^i))$ around $q \in M$, there exists a diffeomorphism $B_U : U \to \mathcal{L}(\mathrm{Sym}^2(\mathbb{R}^d), \mathbb{R}^d)$, such that

$$\varrho(\mathfrak{b}, a) = \left(\mathfrak{b}^i + B_U(q)(a)^i\right)\partial_i\big|_q, \quad (\mathfrak{b}, a) \in \mathcal{T}^O_q M, q \in U.$$

The family of diffeomorphisms (B_U) determines a spray and then a torsion-free linear connection on M (see, e.g., Lang 1999, Section IV.3). The torsion-freeness follows from the symmetry of B_U's. $\qquad\square$

Observe that a group action of $\mathrm{GL}(d, \mathbb{R})$ on $\mathrm{Sym}^2(\mathbb{R}^d)$ can be separated from (2.11), which is given by $g \cdot a = (g \otimes g)a$. Thus, the second component a of each element $(\mathfrak{b}, a) \in \mathcal{T}^O_q M$ can be regarded as a $(2, 0)$-tensor. Recall that we denote by $\mathrm{Sym}^2(TM)$ the bundle of $(2, 0)$-tensors on M, then there is a canonical bundle epimorphism

$$\hat{\varrho} : \mathcal{T}^O M \to \mathrm{Sym}^2(TM), \quad (\mathfrak{b}, a)_q \mapsto a_q, \qquad (2.15)$$

whose kernel is the image of ι. Conversely, we also have a similar connection correspondence property for $\mathrm{Sym}^2(TM)$, as in Proposition 2.11. That is, a linear connection ∇ on M induces a fiber-linear bundle monomorphism from $\mathrm{Sym}^2(TM)$ to $\mathcal{T}^O M$, which is a right inverse of $\hat{\varrho}$ and given by

$$\hat{\iota}_\nabla : \mathrm{Sym}^2(TM) \to \mathcal{T}^O M, \quad a_q \mapsto a^{ij}\left(\partial_i\partial_j\big|_q - \Gamma^k_{ij}(q)\partial_k\big|_q\right) = a^{ij}\nabla^2_{\partial_i, \partial_j}\big|_q \quad (2.16)$$

where ∇^2 is the second covariant derivative (Petersen 2016, Subsection 2.2.2.3) [which is also called the Hessian operator when acting on smooth functions (Jost 2017)]. In other words, $\nabla^2_{\partial_i, \partial_j}\big|_q = \hat{\iota}_\nabla(dx^i \odot dx^j\big|_q)$, where $\odot$ is the symmetrization operator on T^2M.

Combining (2.13) and (2.14) together, we have the following short exact sequence:

$$0 \longrightarrow TM \overset{\iota}{\longrightarrow} \mathcal{T}^O M \overset{\hat{\varrho}}{\longrightarrow} \mathrm{Sym}^2(TM) \longrightarrow 0. \qquad (2.17)$$

Proposition 2.11 and (2.15), (2.16) imply that when a linear connection ∇ is given, the sequence is also split, in the fiber-wise sense. The induced decomposition

$$\mathcal{T}^O M = \iota(TM) \oplus \hat{\iota}_\nabla \left(\mathrm{Sym}^2(TM) \right) \cong TM \oplus \mathrm{Sym}^2(TM), \qquad (2.18)$$

where both the first direct sum $\oplus$ and the isomorphism $\cong$ are in the fiber-wise sense (but not bundle isomorphism and Whitney sum), while the second direct sum is the Whitney sum, and is given by

$$(\mathfrak{b}, a)_q = b^i \partial_i \big|_q + \tfrac{1}{2} a^{ij} \nabla^2_{\partial_i, \partial_j} \big|_q \mapsto (b_q, a_q), \qquad (2.19)$$

for $b_q = (\mathfrak{b}^i + \tfrac{1}{2} a^{jk} \Gamma^i_{jk}(q)) \partial_i |_q \in T_q M$. A similar short exact sequence as (2.17) holds with $\mathcal{T}^E M$ and $\mathrm{Sym}^2_+(TM)$ in place of $\mathcal{T}^O M$ and $\mathrm{Sym}^2(TM)$, respectively.

Now we introduce a subclass of semimartingales on manifolds which contains diffusions. We call the M-valued process $X = \{X(t)\}_{t \in [t_0, \tau)}$ an *Itô process*, if there exists a $\{\mathcal{P}_t\}$-adapted continuous $\mathcal{T}^E M$-valued process $\{(\mathfrak{b}, a)(t)\}_{t \in [t_0, \tau)}$ satisfying $(\mathfrak{b}, a)(t) \in \mathcal{T}^E_{X(t)} M$ for each $t \in [t_0, \tau)$, such that for every $f \in C^\infty(\mathbb{R} \times M)$, $M^{f,X}(t) := f(t, X(t)) - f(t_0, X(t_0)) - \int_{t_0}^t (\frac{\partial}{\partial t} + \mathcal{A}^X) f(s, X(s)) \mathrm{d}s, \, t \in [t_0, \tau)$ is a real-valued $\{\mathcal{P}_t\}$-martingale, where $\mathcal{A}^X_t = (\mathfrak{b}, a)(t) = \mathfrak{b}^i(t)\partial_i + \tfrac{1}{2} a^{ij}(t)\partial_i\partial_j$. We call the process $\{(\mathfrak{b}, a)(t)\}_{t \in [t_0, \tau)} = \{\mathcal{A}^X_t\}_{t \in [t_0, \tau)}$ the *random generator* of X. A similar notion "Brownian semimartingale" is also used in the literature (e.g., Driver 1992). If X is a diffusion with generator $A^X_t = (\mathfrak{b}(t), a(t))$, then it is an Itô process with random generator $\mathcal{A}^X_t = A^X_{(t, X(t))} = (\mathfrak{b}(t, X(t)), a(t, X(t)))$. The difference between Itô processes and diffusions is that the randomness of the random generator of the former can not only appear on the base manifold M, but also on the fibers.

Then, we can define forward mean derivatives in a coordinate-free way, without relying on linear connections.

Definition 2.13 (*Mean derivatives*) For an M-valued Itô process $X = \{X(t)\}_{t \in [t_0, \tau)}$, we define its (forward) mean derivatives $(DX(t), QX(t))$ at time $t \in [t_0, \tau)$ by

$$(DX(t), QX(t)) = (\mathfrak{b}, a)(t) \in \mathcal{T}^E_{X(t)} M,$$

where $(\mathfrak{b}, a)$ is the random generator of X.

Comparing with forward mean derivatives defined in local coordinates before, we have the following relations. The proof follows the lines of Gliklikh (2011, Lemma 9.4).

Lemma 2.14 *Given an M-valued Itô process $X = \{X(t)\}_{t \in [t_0, \tau)}$ and a coordinate chart $(U, (x^i))$ centered at $q \in M$.*

(i) *In the event $\{X(t) \in U\}$, $QX(t)$ has the coordinate expression (2.5) and*

$$(DX)^i(t) = \lim_{\epsilon \to 0^+} \mathbf{E} \left[\frac{X^i(t + \epsilon) - X^i(t)}{\epsilon} \bigg| \mathcal{P}_t \right].$$

(ii) *Given a linear connection ∇ on M, we have, under the conditional probability* $\mathbf{P}(\cdot|X(t) = q)$, *that*

$$(D_\nabla X)^i(t) = (DX)^i(t) + \frac{1}{2}\Gamma^i_{jk}(X(t))(QX)^{jk}(t). \tag{2.20}$$

It follows from (2.20) that the map ϱ_∇ in (2.14) acts on the generator A^X of a diffusion X by

$$\varrho_\nabla\left(A^X_{(t,X(t))}\right) = \varrho_\nabla(DX(t), QX(t)) = D_\nabla X(t) \tag{2.21}$$

For a time-dependent second-order vector field $A_t = (\mathfrak{b}(t), a(t))$, we can take MDEs (2.6) to set up a new type of MDEs by using the mean derivatives as follows:

$$\begin{cases} DX(t) = \mathfrak{b}(t, X(t)), \\ QX(t) = a(t, X(t)). \end{cases} \tag{2.22}$$

Then, similarly to Definitions 2.6 and 2.3, we may also define solutions and uniqueness in law for MDEs (2.22). We call a solution of (2.22) an *integral process* of $A = (A_t)$. Note that the system (2.22) does not rely on linear connections. The equivalence of the well-posedness of (2.22) and the martingale problem in Definition 2.8 is easy to verify. When a linear connection is specified, the system (2.22) and martingale problem associated with A^X in (2.10) are both equivalent to the Itô SDE (2.1) and MDEs (2.6).

3 Stochastic Jets

In classical differential geometry, a tangent vector to a manifold may be defined as an equivalence class of curves passing through a given point, where two curves are equivalent if they have the same derivative at that point (Lee 2013, Chapter 3). This idea can be generalized to higher-order cases, which leads to the notion of jets. The jet structures allow us to translate a system of differential equations to a system of algebraic equations, and make it more intuitive to study the symmetries of systems of differential equations.

In this section we shall generalize these ideas to the stochastic case. We will first give an equivalent description to the second-order elliptic tangent bundle τ^E_M by constructing an equivalence relation on diffusions. Then, we will define the stochastic jets and figure out the "jet-like" bundle structure involved in the space of stochastic jets. Finally, we shall see that the bundle structure is the appropriate platform to formulate SDEs intrinsically. In the next section, we will apply stochastic jets to study stochastic symmetries.

3.1 The Stochastic Tangent Bundle

Recall that a tangent vector can be represented as a equivalence classes of smooth curves that have the same velocity at the base point. This leads to the following equivalent definition of tangent bundle TM:

$$TM \cong \left\{ [\gamma]_q : \gamma \in C_{(0,q)}^\infty(M), q \in M \right\}, \tag{3.1}$$

where $C_{(0,q)}^\infty(M)$ is the set of all smooth curves on M that pass through q at time $t = 0$, and the equivalence relation is defined as $\gamma, \tilde{\gamma} \in C_{(0,q)}^\infty(M)$ are equivalent if and only if $(f \circ \gamma)'(0) = (f \circ \tilde{\gamma})'(0)$ for every real-valued smooth function f defined in neighborhood q. If we replace smooth curves by diffusion processes, and time derivatives by mean derivatives, then we get the following definition.

Definition 3.1 (*The stochastic tangent bundle*) Two M-valued diffusion processes $X = \{X(t)\}_{t \in [0,\tau)}$, $Y = \{Y(t)\}_{t \in [0,\sigma)}$ are said to be *stochastically equivalent at* $(t, q) \in \mathbb{R} \times M$, if, almost surely, $X(t) = Y(t) = q$ and $D(f \circ X)(t) = D(f \circ Y)(t)$ for all $f \in C^\infty(M)$. The equivalence class containing X is called the *stochastic tangent vector* of X at q and is denoted by $j_{(t,q)}X$. When $t = 0$, we denote $j_q X := j_{(0,q)}X$ in short. Let $I_{(t,q)}(M)$ be the set of all M-valued diffusion processes starting from q at time t. The *stochastic tangent bundle* of M is the set

$$\mathcal{T}^S M = \{ j_q X : X \in I_{(0,q)}(M), q \in M \}.$$

Note that since X, Y are M-valued diffusion processes, $f(X)$ and $f(Y)$ are real-valued Itô processes, and hence their mean derivatives exists.

At this stage, we have not yet touched the jet-like formulation even though we used the jet-like notation $j_q X$. Indeed, if one follows strictly the definition of jet bundles over the trivial bundle $(\mathbb{R} \times M, \pi, \mathbb{R})$, it is more rational to use the time line $\mathbb{R}$ as "source" and the manifold M as "target" (cf. Saunders 1989, Example 4.1.16). But here we just assign the "target" to the manifold M, because, roughly speaking, one can talk about the velocity of a smooth curve at a moment t, but not about the generator of a diffusion at a moment t. Instead, we can talk about the generator of a diffusion at a position $q \in M$. Later on, we will define the "bona fide" stochastic jet space which possess the time line $\mathbb{R}$ as "source" and the manifold M as "target."

Similarly to the one-to-one correspondence between tangent space and space of equivalence classes of smooth curves, we have the following:

Proposition 3.2 *There is a one-to-one correspondence between the stochastic tangent bundle $\mathcal{T}^S M$ and the second-order elliptic tangent bundle $\mathcal{T}^E M$.*

Proof For an M-valued diffusion process $X \in I_{(0,q)}(M), q \in M$, we denote by A^X its generator. Then, the map $j_q X \mapsto A_{(0,q)}^X = (DX(0), QX(0))$ defines a one-to-one correspondence between $\mathcal{T}^S M$ and $\mathcal{T}^E M$. The inverse map is $A_q = (\flat, a)_q \mapsto j_q X^A$, where A is a section of $\mathcal{T}^E M$ (i.e., an elliptic second-order operator) smoothly extending the element $A_q \in \mathcal{T}_q^E M$, and $X^A \in I_{(0,q)}(M)$ is a diffusion processes having A as its generator. □

Therefore, the stochastic tangent bundle $\mathcal{T}^S M$ admit a smooth structure which makes it to be a smooth manifold diffeomorphic to $\mathcal{T}^E M$, and hence it is a bona fide fiber bundle over M. In the sequel, we will identify $\mathcal{T}^S M$ with $\mathcal{T}^E M$ without ambiguity. And the projection map from $\mathcal{T}^S M$ to M will be denoted by τ_M^S, that is, $\tau_M^S(j_q X) = q$ for any $j_q X \in \mathcal{T}^S M$.

Definition 3.3 (*Canonical coordinate system on $\mathcal{T}^S M$*) Let $(U, (x^i))$ be an coordinate system on M. The induced canonical coordinate chart $(U^{(1)}, x^{(1)})$ on $\mathcal{T}^S M$ is defined by

$$U^{(1)} := \{j_q X : q \in U, X \in I_{(0,q)}(M)\}, \quad x^{(1)} := \left(x^i, D^i x, Q^{jk} x\right),$$

where $x^i(j_q X) = x^i(q)$, $D^i x(j_q X) = (DX)^i(0)$ and $Q^{jk}x(j_q X) = (QX)^{jk}(0)$.

Our slightly ambiguous notations $D^i x$ and $Q^{jk}x$ are chosen so as to avoid the worse one Qx^{jk}.

When a linear connection ∇ is provided, we can also define the coordinates via the ∇-mean derivative D_∇ instead of D, as follows:

$$D_\nabla^i x(j_q X) := (D_\nabla X)^i(0).$$

Then, $x_\nabla^{(1)} := (x^i, D_\nabla^i x, Q^{jk}x)$ also forms a coordinate system on $\mathcal{T}^S M$, which we call the ∇-canonical coordinate system. It follows from relation (2.20) that

$$D_\nabla^i x = D^i x + \tfrac{1}{2}\left(\Gamma_{jk}^i \circ x\right) Q^{jk}x. \tag{3.2}$$

Using the identification of elements $j_q X \in \mathcal{T}_q^S M$ and $(\mathfrak{b}, \mathfrak{a})_q \in \mathcal{T}_q^E M$ via Proposition 3.2, as well as their relations with the element $(b_q, a_q) \in TM \oplus \mathrm{Sym}^2(TM)$, via (2.19), we have $D^i x(j_q X) = \mathfrak{b}^i$, $D_\nabla^i x(j_q X) = b^i = \mathfrak{b}^i + \tfrac{1}{2}a^{jk}\Gamma_{jk}^i(q)$ and $Q^{jk}x(j_q X) = a^{jk}$. In this way the fiber-linear bundle projection ϱ_∇ of (2.14) maps, under the canonical coordinates $(x, \dot{x})$ on TM, as follows:

$$\dot{x}^i \circ \varrho_\nabla(j_q X) = \left(D^i x + \tfrac{1}{2}\left(\Gamma_{jk}^i \circ x\right) Q^{jk}x\right)(j_q X) = D_\nabla^i x(j_q X), \tag{3.3}$$

so that $D_\nabla^i x = \dot{x}^i \circ \varrho_\nabla$. Therefore, $(x^i, D_\nabla^i x)$ is a partial coordinate system on $\mathcal{T}^S M$ that coincides with $(x^i, \dot{x}^i)$ when restricted on TM. Moreover, the decomposition in (2.19) yields the following expressions for second-order vector fields:

$$(Dx, Qx) = D^i x \partial_i + \tfrac{1}{2} Q^{jk} x \partial_j \partial_k = D_\nabla^i x \partial_i + \tfrac{1}{2} Q^{jk} x \nabla_{\partial_j, \partial_k}^2. \tag{3.4}$$

Similarly to Definition 3.1, we define a ∇-dependent equivalence relation as follows:

Definition 3.4 Two M-valued diffusion processes $X = \{X(t)\}_{t \in [0,\tau)}, Y = \{Y(t)\}_{t \in [0,\sigma)}$ are said to be ∇-*stochastically equivalent at* $(t, q) \in \mathbb{R} \times M$, if, almost surely,

$X(t) = Y(t) = q$ and $D_\nabla X(t) = D_\nabla X(t)$. The equivalence class containing X is called the ∇-*tangent vector* of X at q and is denoted by $j_{(t,q)}^\nabla X$. When $t = 0$, we denote $j_q^\nabla X := j_{(0,q)}^\nabla X$ for short.

Then, similarly to Proposition 3.2, one can show that the tangent bundle TM can be identified with the following set of equivalent classes of diffusions:

$$\left\{ j_q^\nabla X : X \in I_{(0,q)}(M), q \in M \right\}, \tag{3.5}$$

via $j_q^\nabla X \mapsto D_\nabla X(0)$. Under this identification, it follows from (2.21) that $j_q^\nabla X = \varrho_\nabla(j_q X)$. Clearly, if we regard all smooth curves as special diffusions, then the partition determined by (3.1) is the restriction of the one determined by (3.5) to the set of all smooth curves.

Remark 3.5 In presence of a linear connection ∇ on M, one can easily follow Definition 3.1 and Proposition 3.2 with D_∇ in place of D, to verify the one-to-one correspondence between the set $\mathcal{T}^S M$ of equivalent classes and the Whitney sum $TM \oplus \mathrm{Sym}_+^2(TM)$, which brings back to the fiber-wise isomorphism (2.18). But since such kind of correspondence need to specify beforehand a linear connection, we still endow $\mathcal{T}^S M$ with the structure of $\mathcal{T}^E M$ instead of that of $TM \oplus \mathrm{Sym}^2(TM)$ in this paper, although the latter is also feasible and may provide easier calculations.

3.2 The Stochastic Jet Space

In classical jet theory, for the trivial bundle $(\mathbb{R} \times M, \pi, \mathbb{R})$, there is a one-to-one correspondence between 1-jets and tangent vectors, and there is a canonical diffeomorphism between the first-order jet bundle $J^1\pi$ and $\mathbb{R} \times TM$ (Saunders 1989, Example 4.1.16).

Now using similar ideas, we will introduce the "bona fide" stochastic jet space. The key is to modify the definition of stochastic tangent vectors, to involve the time line $\mathbb{R}$ as the "source" as well as to randomize the initial datum of the diffusion processes. Intuitively, an M-valued diffusion process X can be regarded as a random "section" of the trivial "bundle" $(\mathbb{R} \times M, \pi, \mathbb{R})$ which is merely continuous in time and depends on the sample point ω.

For a metric space (F, d), we denote by $L^0(\Omega, F)$ the quotient space of all F-valued random elements, by the following equivalence relation: two random elements are equivalent if and only if they are identical almost surely. We endow $L^0(\Omega, F)$ with the topology of the following **P**-essential metric (cf. Munkres 1975, Section 43):

$$\rho(\xi, \zeta) = \inf\{c > 0 : \mathbf{P}(d(\xi, \zeta) > c) = 0\} \wedge 1.$$

Definition 3.6 Two M-valued diffusion processes $X = \{X(s)\}_{s \in [t,\tau)}, Y = \{Y(s)\}_{s \in [t,\sigma)}$ starting at time t, are said to be *stochastically equivalent at* $t \in \mathbb{R}$, if, almost surely, $X(t) = Y(t)$ and $(DX(t), QX(t)) = (DY(t), QY(t))$. The equivalence class containing X is called the *stochastic jet* of X at t, denoted by $j_t X$. Let $I_t(M)$ be the set of all M-valued diffusion processes starting at time t. Then, the *stochastic jet space*

of M is the set

$$\mathcal{J}^S M = \{j_t X : X \in I_t(M), t \in \mathbb{R}\}.$$

The functions π_1^S and $\pi_{1,0}^S$, called stochastic source and target projections, are defined by

$$\pi_1^S : \mathcal{J}^S M \to \mathbb{R}, \quad j_t X \mapsto t,$$

and

$$\pi_{1,0}^S : \mathcal{J}^S M \to \mathbb{R} \times L^0(\Omega, M), \quad j_t X \mapsto (t, X(t)).$$

In the above definition, since $\pi_M \circ \phi = \mathbf{Id}_M$, we have $\pi(Y) = \pi_M \circ \phi(X) = X$ a.s., that is, X is the projection of Y.

To characterize the relation between $\mathcal{J}^S M$ and $\mathcal{T}^S M$ (or $\mathcal{T}^E M$), we need the following definitions.

Definition 3.7 (*Horizontal subspace*) Let (E, π_M, M) be a fiber bundle. The horizontal subspace of $L^0(\Omega, E)$ is defined by

$$L^h(\Omega; \pi_M) := \{\phi \circ \xi \in L^0(\Omega, E) : \phi \text{ is a section of } \pi_M, \xi \in L^0(\Omega, M)\}.$$

An element of the horizontal subspace $L^h(\Omega; \tau_M^E)$ of $L^0(\Omega, \mathcal{T}^E M)$ is then of the form $A \circ \xi$, where A is a section of τ_M^E and $\xi \in L^0(\Omega, M)$. Such an element $A \circ \xi$ will be denoted by A_ξ. By the correspondence of $\mathcal{T}^S M$ and $\mathcal{T}^E M$, one can easily get the following equivalent definition for $L^h(\Omega; \tau_M^E)$,

$$L^h\left(\Omega; \tau_M^E\right) = L^h\left(\Omega; \tau_M^S\right) := \{j_{X(0)} X : X \in I_0(M)\} \subset L^0(\Omega, \mathcal{T}^S M).$$

The correspondence is given explicitly by

$$j_{X(0)} X = A_{X(0)}^X = (DX(0), QX(0)), \quad \text{or} \quad A_\xi = j_\xi X^{A_\xi}.$$

where X^{A_ξ} is an M-valued diffusion with generator A and with $X^{A_\xi}(0) = \xi$ a.s..

Proposition 3.8 *The stochastic jet space $\mathcal{J}^S M$ is trivial. More precisely, we have the homeomorphism*

$$\mathcal{J}^S M \cong \mathbb{R} \times L^h\left(\Omega; \tau_M^S\right),$$

given by $j_t X \mapsto (t, j_{X(t)}(\theta_t X))$, for any $X \in I_t(M)$, where θ_t is the shift operator on $\mathcal{C}$, that is, $\theta_t \omega(\cdot) = \omega(\cdot + t)$.

Proof The homeomorphism $\mathcal{J}^S M \cong \mathbb{R} \times \mathcal{J}^S_0 M$ is given by $j_t X \mapsto (t, j_0(\theta_t X))$. The homeomorphism $\mathcal{J}^S_0 M \cong L^h(\Omega; \tau^S_M)$ is given by $j_0 X \mapsto j_{X(0)} X$, whose inverse map is $A_\xi \mapsto j_0 X^{A_\xi}$. $\hfill\square$

Definition 3.9 (*Stochastic fibered space*)

(i) Given a fiber bundle (E, π_M, M) with total space E, base space M and typical fiber manifold F, the stochastic fibered space associated with it is the triplet (E^S, π^S_M, M) where

$$E^S := \{(q, \xi) : q \in M, \xi \in \hat{L}(\Omega, E_q)\},$$

$\pi^S_M : E^S \to M$ is the natural projection given by $\pi^S_M(q, \xi) = q$, and $\hat{L}(\Omega, F)$ is a subspace of $L^0(\Omega, F)$, with E_q denoting the fiber of π_M over q. The fiber bundle E is called model bundle of E^S. There is a family of projections $\{\pi_\omega\}_{\omega \in \Omega}$ from the stochastic fiber manifold E^S to its model bundle E, defined by

$$\pi_\omega : E^S \to E, \quad (q, \xi) \mapsto (q, \xi(\omega)).$$

(ii) A global section of (E^S, π^S_M, M) is called a random global section. A random local section is a map $\sigma : U \to E$ defined on some measurable subset $U \subset \Omega \times M$ and such that, for almost all $\omega \in \Omega$, $\sigma(\omega) : U_\omega \to E$ is a local section of (E, π_M, M), where $U_\omega = U \cap (\{\omega\} \times M)$.

Note that a random global section is a random local section defined on all $\Omega \times M$.

It follows from Proposition 3.8 that the stochastic jet space $(\mathcal{J}^S M, \pi^S_1, \mathbb{R})$ is a stochastic fibered space, whose associated model bundle is $(\mathbb{R} \times \mathcal{T}^S M, \pi_1, \mathbb{R})$. Just like the first-order jet bundle $J^1\pi$ which is diffeomorphic to $\mathbb{R} \times TM$, the model bundle $\mathbb{R} \times \mathcal{T}^S M$ is itself a jet bundle and also has two bundle structures, with base space $\mathbb{R}$ and $\mathbb{R} \times M$, respectively. The corresponding source and target projections are defined, respectively by

$$\pi_1 : \mathbb{R} \times \mathcal{T}^S M \to \mathbb{R}, \quad (t, j_q X) \mapsto t,$$

and

$$\pi_{1,0} : \mathbb{R} \times \mathcal{T}^S M \to \mathbb{R} \times M, \quad (t, j_q X) \mapsto (t, q).$$

Moreover, we will denote the natural projection from $\mathbb{R} \times \mathcal{T}^S M$ to $\mathcal{T}^S M$ by $\pi_{0,1}$. This projection map is indeed a bundle homomorphism from $(\mathbb{R} \times \mathcal{T}^S M, \pi_{1,0}, \mathbb{R} \times M)$ to $(\mathcal{T}^S M, \tau^S_M, M)$, whose projection is the natural projection from $\mathbb{R} \times M$ to M, denoted by $\hat{\pi}$.

Similarly to Proposition 3.8, we have the following diffeomorphisms for the model bundle $\mathbb{R} \times \mathcal{T}^S M$:

$$\{j_{(t,q)} X : X \in I_{(t,q)}(M), t \in \mathbb{R}, q \in M\} \cong \mathbb{R} \times \mathcal{T}^S M \cong \mathbb{R} \times \mathcal{T}^E M,$$

which is given by

$$j_{(t,q)}X \mapsto (t, j_q(\theta_t X)) \mapsto A_{(t,q)}^X = (t, DX(t), QX(t)), \qquad (3.6)$$
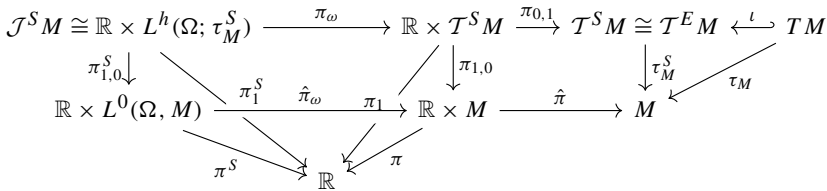
for any $X \in I_{(t,q)}(M)$, where A^X is the generator of X as a section of $\mathbb{R} \times \mathcal{T}^E M$ (i.e., a time-dependent elliptic second-order differential operator). Furthermore, the proof of Proposition 3.2 allows us to find simply the inverse maps, especially for the second diffeomorphism. That is, for any $(t, A_q) = (t, \mathfrak{b}, a) \in \pi_{1,0}^{-1}(t, q)$,

$$(t, A_q) = (t, \mathfrak{b}, a) \mapsto \left(t, j_q\left(\theta_t X^A\right)\right) \mapsto j_{(t,q)}X^A, \qquad (3.7)$$

where A is a section of $\mathbb{R} \times \mathcal{T}^E M$ such that $A_{(t,q)} = A_q$, and $X^A \in I_{(t,q)}(M)$ is a diffusion process having A as its generator.

The "stochastic target" of $\mathcal{J}^S M$, i.e., the trivial bundle $(\mathbb{R} \times L^0(\Omega, M), \pi^S, M)$, is another example of stochastic fibered spaces. Its model bundle is the trivial bundle $(\mathbb{R} \times M, \pi, \mathbb{R})$. The graph of an M-valued stochastic process defined on a random time interval $[0, \tau)$ is a random (local) section of $(\mathbb{R} \times L^0(\Omega, M), \pi^S, \mathbb{R})$. The projection of π_ω on the targets from $\mathbb{R} \times L^0(\Omega, M)$ to $\mathbb{R} \times M$ is denoted by $\hat{\pi}_\omega$.

We may summarize how all these maps fit together by the following diagram:

$$
\begin{array}{ccccccc}
\mathcal{J}^S M \cong \mathbb{R} \times L^h(\Omega; \tau_M^S) & \xrightarrow{\pi_\omega} & \mathbb{R} \times \mathcal{T}^S M & \xrightarrow{\pi_{0,1}} & \mathcal{T}^S M \cong \mathcal{T}^E M & \xleftarrow{\iota} & TM \\
\pi_{1,0}^S \downarrow \quad \searrow & \swarrow \quad \downarrow & \downarrow \pi_{1,0} & & \downarrow \tau_M^S & \swarrow \tau_M & \\
\mathbb{R} \times L^0(\Omega, M) & \xrightarrow[\pi_1^S \quad \hat{\pi}_\omega \quad \pi_1]{} & \mathbb{R} \times M & \xrightarrow{\hat{\pi}} & M & & \\
& \searrow \pi^S \quad \swarrow \pi & & & & & \\
& \mathbb{R} & & & & &
\end{array}
$$

When a linear connection is specified on M, one can easily obtain, similarly to (3.6), the following homeomorphism:

$$\left\{j_t^\nabla X : X \in I_t(M), t \in \mathbb{R}\right\} \cong \mathbb{R} \times L^h(\Omega; \tau_M), \quad j_t^\nabla X \mapsto \left(t, j_{X(t)}^\nabla(\theta_t X)\right),$$

and the following diffeomorphisms:

$$\left\{j_{(t,q)}^\nabla X : X \in I_{(t,q)}(M), t \in \mathbb{R}, q \in M\right\} \cong \mathbb{R} \times \left\{j_q^\nabla X : X \in I_{(0,q)}(M), q \in M\right\}$$
$$\cong \mathbb{R} \times TM \cong J^1\pi,$$

where the first two diffeomorphisms are given by

$$j_{(t,q)}^\nabla X \mapsto \left(t, j_q^\nabla(\theta_t X)\right) \mapsto (t, D_\nabla X(t)),$$

and the last one is due to the classical theory.

3.3 Intrinsic Formulation of SDEs

With the classical machinery of jet structures, it is possible to translate differential equations into algebraic equations on jet bundle (Saunders 1989). In this subsection, we follow this way to formulate intrinsic SDEs.

For a subset S of the model bundle $\mathbb{R} \times \mathcal{T}^S M$ and $t \in \mathbb{R}$, we denote by S_t the intersection of S with the fiber $\{t\} \times \mathcal{T}^S M$.

Definition 3.10 A stochastic differential equation on M is a closed embedded submanifold S of the model jet bundle $\mathbb{R} \times \mathcal{T}^S M$ with $S_0 \neq \emptyset$. A (local) solution of the stochastic differential equation S is a triple $X, (\Omega, \mathcal{F}, \mathbf{P}), \{\mathcal{P}_t\}_{t \geq 0}$, where

(i) $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and $\{\mathcal{P}_t\}_{t \geq 0}$ is a past filtration of $\mathcal{F}$ satisfying the usual conditions,

(ii) $X = \{X(t)\}_{t \in [0, \tau)}$ is a $\{\mathcal{P}_t\}$-adapted M-valued diffusion process over $[0, \tau)$, where τ is a $\{\mathcal{P}_t\}$-stopping time, and

(iii) almost surely $j_t X = (t, j_{X(t)}(\theta_t X)) \in S$ for every $t \in [0, \tau)$.

Remark 3.11 (i) The condition that $S_0 \neq \emptyset$ is just for convenience, in order to set the initial time at $t = 0$.

(ii) There is an equivalent way to formulate the solution of a stochastic differential equation S. That is, a (local) solution is a pair (P, τ), where P is a probability measure on $(\mathcal{C}, \mathcal{B}(\mathcal{C}), \{\mathcal{B}_t\})$ and τ is a $\{\mathcal{B}_t\}$-stopping time, such that for P-almost surely ω, $j_t \omega = (t, j_{\omega(t)}(\theta_t \omega)) \in S$ for every $t \in [0, \tau(\omega))$.

This definition does not look like the traditional definition of a stochastic differential equation, but we can see the relationship between the two by using coordinates. Since S is a embedded submanifold of $\mathbb{R} \times \mathcal{T}^S M$, it admits a *local defining function* in a neighborhood of each of its points (Lee 2013, Proposition 5.16). That is, for a coordinate chart $(\mathbb{R} \times U^{(1)}, (t, x^{(1)}))$ of the point $(0, j_q X) \in S_0$, there is a function $\Theta : \mathbb{R} \times U^{(1)} \to \mathbb{R}^K$ where $K = \dim \mathcal{T}^S M - \dim S$, such that $S \cap (\mathbb{R} \times U^{(1)}) = \Theta^{-1}(0)$ and 0 is a regular value of Θ. Then, the condition $j_t X = (t, j_{X(t)}(\theta_t X)) \in S$ before $X(t)$ leaves the neighborhood $U = \tau_M^S(U^{(1)})$ reads in local coordinates as

$$\Theta(t, x, Dx, Qx)(j_t X) = \Theta(t, X(t), DX(t), QX(t)) = 0, \qquad (3.8)$$

which defines a general MDE (in terms of mean derivatives). The use of a submanifold S is therefore a way to distinguish the definition of the equation from a definition of its solutions.

As an example, the system of MDEs (2.22) can be rewritten to the form (3.8) by setting the defining function

$$\Theta(t, x, Dx, Qx) = \left(Dx - \mathfrak{b}(t, x), Qx - (\sigma \circ \sigma^*)(t, x) \right). \qquad (3.9)$$

So far we have not done anything but reformulate the basic problem of finding solutions of systems of stochastic differential equations in a more geometrical form, ideally suited to our investigation into symmetry groups thereof.

4 Stochastic Symmetries

The symmetry group of a system of differential equations is the largest local group of transformations acting on the independent and dependent variables of the system with the property that it transform solutions of the system into other solutions (Olver 1998). In the stochastic case, we can proceed analogously.

All methods of this section work in the local case, that is, the vector fields are not necessarily complete and the bundle homomorphisms could be only locally defined.

4.1 Prolongations of Diffusions and Bundle Homomorphisms

Definition 4.1 (*Prolongations of diffusions*) Let X be an M-valued diffusion process defined on a stopping time interval $[t_0, \tau)$. The prolongation of X is a $\mathcal{T}^S M$-valued process jX defined by, for θ_t the shift operator,

$$jX(t) = j_{X(t)}(\theta_t X), \quad t \in [t_0, \tau).$$

Note that $j_t X = (t, j_{X(t)}(\theta_t X)) = (t, jX(t))$. Thus the graph of the prolongation process jX is nothing but the random section jX of the stochastic jet space $\mathcal{J}^S M$. It is easy to see that if X is an M-valued diffusion process, then jX is a $\mathcal{T}^S M$-valued diffusion process.

Given two smooth manifolds M and N, a bundle homomorphism F from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ is a projectable (or fiber-preserving) smooth map, which means it maps fibers of π to fibers of ρ. Hence, there exist two smooth maps $F^0 : \mathbb{R} \to \mathbb{R}$ and $\bar{F} : \mathbb{R} \times M \to N$ such that $F(t, q) = (F^0(t), \bar{F}(t, q))$. This leads to $\rho \circ F = F^0 \circ \pi$ which is the original definition of bundle homomorphisms. We denote $F = (F^0, \bar{F})$ and say that F projects to F^0.

The following lemma shows that a bundle homomorphisms has the property that it always transforms diffusions into diffusions. One can find a proof of it in Lemma 4.8 or Corollary A.5.

Lemma 4.2 *Given a bundle homomorphism $F = (F^0, \bar{F})$ from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$, where F^0 is a diffeomorphism, for every M-valued diffusion process $X = \{X(t)\}_{t \in [t_0, \tau)}$, the image of its graph (or its corresponding random local section) $\{(t, X(t)) : t \in [t_0, \tau)\}$ by F, i.e.,*

$$\{F(t, X(t)) : t \in [t_0, \tau)\}$$

is almost surely the graph of a well-defined N-valued diffusion process $\tilde{X}$ given by

$$\tilde{X}(s) = \bar{F}\left((F^0)^{-1}(s), X((F^0)^{-1}(s))\right), \quad s \in [F^0(t_0), F^0(\tau)). \tag{4.1}$$

As observed in Remark A.6, among all (deterministic) smooth maps from $\mathbb{R} \times M$ to $\mathbb{R} \times N$, the class of bundle homomorphisms is the only subclass that maps diffusions to diffusions.

Definition 4.3 (*Pushforwards of diffusions by bundle homomorphisms*) We call the diffusion $\tilde{X}$ of Lemma 4.2 the pushforward of X by F, and write $\tilde{X} = F \cdot X$. When $M = N$ and F is a bundle endomorphism on $(\mathbb{R} \times M, \pi, \mathbb{R})$, we also call $F \cdot X$ the transform of X by F.

We now introduce the idea of stochastic prolongation whereby a bundle homomorphism may be extended to act upon the model jet bundle.

Definition 4.4 (*Stochastic prolongations of bundle homomorphisms*) Let F be a bundle homomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to a diffeomorphism $F^0 : \mathbb{R} \to \mathbb{R}$. The stochastic prolongation of F is the map $jF : \mathbb{R} \times \mathcal{T}^S M \to \mathbb{R} \times \mathcal{T}^S N$ defined by

$$jF(j_{(t,q)}X) = j_{F(t,q)}(F \cdot X). \tag{4.2}$$

It is easy to see from (4.1) that if $j_{(t,q)}X = j_{(t,q)}Y$, then $j_{F(t,q)}(F \cdot X) = j_{F(t,q)}(F \cdot Y)$. Therefore, the map jF is well defined. By letting $F = (F^0, \bar{F})$, definition (4.2) can be rewritten in a more evident way:

$$jF(t, j_q(\theta_t X)) = \left(F^0(t), j_{\bar{F}(t,q)}\theta_{F^0(t)}(F \cdot X)\right). \tag{4.3}$$

The following properties are easy to check.

Corollary 4.5 (i) *The map $jF : \pi_1 \to \rho_1$ is a bundle homomorphism projecting to F^0.*
(ii) *The map $jF : \pi_{1,0} \to \rho_{1,0}$ is a bundle homomorphism projecting to F.*
(iii) *$j(\mathbf{Id}_{\mathbb{R} \times M}) = \mathbf{Id}_{\mathbb{R} \times \mathcal{T}^S M}$. Let F and G be two bundle endomorphisms on $(\mathbb{R} \times M, \pi, \mathbb{R})$ projecting to diffeomorphisms. Then, $j(F \circ G) = jF \circ jG$.*

By virtue of (4.3) and Corollary 4.5.(i), we may write $jF = (F^0, \overline{jF})$, where $\overline{jF} : \mathbb{R} \times \mathcal{T}^S M \to \mathcal{T}^S N$ is the smooth map given by

$$\overline{jF}(t, j_q(\theta_t X)) = j_{\bar{F}(t,q)}\theta_{F^0(t)}(F \cdot X). \tag{4.4}$$

We can also consider the pushforward of the $\mathcal{T}^S M$-valued process jX by the bundle homomorphism jF.

Corollary 4.6 *Given a bundle homomorphism $F : (\mathbb{R} \times M, \pi, \mathbb{R}) \to (\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to a diffeomorphism on $\mathbb{R}$, and an M-valued diffusion process X, we have*

$$jF \cdot jX = j(F \cdot X).$$

Proof It follows from (4.1), (4.4) and Definition 4.1 that

$$
\begin{aligned}
jF \cdot jX(s) &= \overline{jF}\left((F^0)^{-1}(s), jX((F^0)^{-1}(s))\right) \\
&= \overline{jF}\left((F^0)^{-1}(s), j_{X((F^0)^{-1}(s))}(\theta_{(F^0)^{-1}(s)}X)\right) \\
&= j_{\tilde{X}(s)}(\theta_s \tilde{X}) = j\tilde{X}(s).
\end{aligned}
$$

The result follows. □

Now we need to investigate the coordinate representation of jF, in stochastic analysis terms. Before that, we introduce the stochastic version of the notion of total derivatives.

Definition 4.7 (*Total mean derivatives*) Let f be a smooth real-valued function on $\mathbb{R} \times M$. The total mean derivative and total quadratic mean derivative of f are the unique smooth functions $\mathbf{D}_t f$ and $\mathbf{Q}_t f$ defined on $\mathbb{R} \times T^S M$, with the property that if $X \in I_{(t_0,q)}(M)$ is a representative diffusion process of $j_{(t_0,q)} X$, then

$$(\mathbf{D}_t f)(j_{(t_0,q)} X) = D[f(t_0, X(t_0))],$$
$$(\mathbf{Q}_t f)(j_{(t_0,q)} X) = Q[f(t_0, X(t_0))].$$

There is an abuse of notations in the above definition. Indeed, the left-hand sides (LHSs) of the above two equations both involve subscripts t, but their RHS's do not depend on t. Those two equations need to be understood as that functions $\mathbf{D}_t f$, $\mathbf{Q}_t f$ taking their values on the point $j_{(t_0,q)} X \in \mathbb{R} \times T^S M$ equal to the RHS's.

It is easy to check that the definitions of total mean derivatives are independent of the choice of representative diffusions. By Itô's formula, we have the following coordinate representation for total mean derivatives in the local chart $(\mathbb{R} \times U^{(1)}, (t, x^{(1)}))$ on $\mathbb{R} \times T^S M$,

$$\mathbf{D}_t f = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} D^i x + \frac{1}{2} \frac{\partial^2 f}{\partial x^j \partial x^k} Q^{jk} x,$$
$$\mathbf{Q}_t f = \frac{\partial f}{\partial x^j} \frac{\partial f}{\partial x^k} Q^{jk} x. \tag{4.5}$$

If a linear connection ∇ is specified, we can use (3.4) to rewrite $\mathbf{D}_t$ as follows:

$$\mathbf{D}_t = \partial_t + D_\nabla^i x \partial_i + \tfrac{1}{2} Q^{jk} x \nabla_{\partial_j, \partial_k}^2. \tag{4.6}$$

Lemma 4.8 *Let us be given a bundle homomorphism $F = (F^0, \bar{F})$ from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to a diffeomorphism F^0 and an M-valued diffusion process $X = \{X(t)\}_{t \in [t_0, \tau)}$. If $\tilde{X} = F \cdot X$, then in local coordinates (t, x^i) around (t_0, q) and (s, y^j) around $F(t_0, q)$,*

$$(D\tilde{X})^j (F^0(t)) = (\mathbf{D}_t \bar{F}^j) \left(j_{(t, X(t))} X \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s} (F^0(t)),$$
$$(Q\tilde{X})^{kl} (F^0(t)) = \left(\frac{\partial \bar{F}^k}{\partial x^i} \frac{\partial \bar{F}^l}{\partial x^j} \right) (t, X(t)) (QX)^{ij} (t) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s} (F^0(t)).$$

Proof Assume that the diffusion X can be represented in local coordinates by

$$dX^i(t) = \mathfrak{b}^i(t, X(t))dt + \sigma_r^i(t, X(t))dW^r(t), \quad X^i(t_0) = x^i(q).$$

where W is an N-dimensional Brownian motion, so that

$$j_t X = (DX(t), QX(t)) = (\mathfrak{b}, \sigma \circ \sigma^*)(t, X(t)).$$

Let $(s_0, \tilde{q}) = F(t_0, q) = (F^0(t_0), \bar{F}(t_0, q))$. Then,

$$X^i((F^0)^{-1}(s)) = x^i(q) + \int_{(F^0)^{-1}(s_0)}^{(F^0)^{-1}(s)} \mathfrak{b}^i(u, X(u))\mathrm{d}u$$
$$+ \int_{(F^0)^{-1}(s_0)}^{(F^0)^{-1}(s)} \sigma_r^i(u, X(u))\mathrm{d}W^r(u).$$

Define

$$B(s) = \int_0^{(F^0)^{-1}(s)} \sqrt{(F^0)'(u)}\mathrm{d}W(u).$$

Then, (Øksendal 2010, Theorem 8.5.7) says that B is an N-dimensional $\{\mathcal{F}_{(F^0)^{-1}(s)}\}$-Brownian motion, as by a change of variable $u = (F^0)^{-1}(v)$, we have

$$\int_{(F^0)^{-1}(s_0)}^{(F^0)^{-1}(s)} \sigma_r^i(u, X(u))\mathrm{d}W^r(u) = \int_{s_0}^{s} \sigma_r^i((F^0)^{-1}(v), X((F^0)^{-1}(v)))$$
$$\left(\frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(v)\right)^{\frac{1}{2}} \mathrm{d}B^r(v).$$

Therefore,

$$X^i((F^0)^{-1}(s)) = x^i(q) + \int_{s_0}^{s} \mathfrak{b}^i((F^0)^{-1}(v), X((F^0)^{-1}(v)))\mathrm{d}(F^0)^{-1}(v)$$
$$+ \int_{s_0}^{s} \sigma_r^i((F^0)^{-1}(v), X((F^0)^{-1}(v))) \left(\frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(v)\right)^{\frac{1}{2}} \mathrm{d}B^r(v).$$

Recall that $\tilde{X}(s) = \bar{F}\left((F^0)^{-1}(s), X((F^0)^{-1}(s))\right)$. Using Itô's formula, we have

$$\tilde{X}^j(s) = y^j(\tilde{q}) + \int_{s_0}^{s} \frac{\partial \bar{F}^j}{\partial t}\left((F^0)^{-1}(v), X((F^0)^{-1}(v))\right)\mathrm{d}(F^0)^{-1}(v)$$
$$+ \int_{s_0}^{s} \frac{\partial \bar{F}^j}{\partial x^i}\left((F^0)^{-1}(v), X((F^0)^{-1}(v))\right)\mathrm{d}X^i((F^0)^{-1}(v))$$
$$+ \frac{1}{2}\int_{s_0}^{s} \frac{\partial^2 \bar{F}^j}{\partial x^k \partial x^l}\left((F^0)^{-1}(v), X((F^0)^{-1}(v))\right)\mathrm{d}\langle X^k \circ (F^0)^{-1},$$

$$X^l \circ (F^0)^{-1} \rangle (v) = y^j(q) + \int_{s_0}^s \left[\frac{\partial \bar{F}^j}{\partial t} + \frac{\partial \bar{F}^j}{\partial x^i} \mathfrak{b}^i + \frac{1}{2} \frac{\partial^2 \bar{F}^j}{\partial x^k \partial x^l} \sigma_r^k \sigma_r^l \right]$$

$$\left((F^0)^{-1}(v), X((F^0)^{-1}(v)) \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(v) \mathrm{d}v$$

$$+ \int_{s_0}^s \left(\frac{\partial \bar{F}^j}{\partial x^i} \sigma_r^i \right) \left((F^0)^{-1}(v), X((F^0)^{-1}(v)) \right)$$

$$\left(\frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(v) \right)^{\frac{1}{2}} \mathrm{d}B^r(v).$$

It follows that

$$(D\tilde{X})^j(s) = \left[\frac{\partial \bar{F}^j}{\partial t} + \frac{\partial \bar{F}^j}{\partial x^i} \mathfrak{b}^i + \frac{1}{2} \frac{\partial^2 \bar{F}^j}{\partial x^k \partial x^l} \sigma_r^k \sigma_r^l \right]$$

$$\left((F^0)^{-1}(v), X((F^0)^{-1}(v)) \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(v)$$

$$= (\mathbf{D}_t \bar{F}^j) \left(j_{((F^0)^{-1}(s), X((F^0)^{-1}(s)))} X \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(s),$$

$$(Q\tilde{X})^{kl}(s) = \left(\frac{\partial \bar{F}^k}{\partial x^i} \sigma_r^i \frac{\partial \bar{F}^l}{\partial x^j} \sigma_r^j \right) \left((F^0)^{-1}(s), X((F^0)^{-1}(s)) \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(s)$$

$$= \left(\frac{\partial \bar{F}^k}{\partial x^i} \frac{\partial \bar{F}^l}{\partial x^j} \right) \left((F^0)^{-1}(s), X((F^0)^{-1}(s)) \right) (QX)^{ij}$$

$$\left((F^0)^{-1}(s) \right) \frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s}(s).$$

This completes the proof.   □

We denote the induced local coordinates on $T^S N$ by $(y^j, D^j y, Q^{kl} y)$. Then, clearly, $y^j \circ jF = y^j \circ \overline{jF} = y^j \circ F = \bar{F}^j$. Now take $j_{(t,q)} X \in \mathbb{R} \times T^S M$. Then,

$$D^j y \circ jF \left(j_{(t,q)} X \right) = D^j y \left(j_{F(t,q)} \tilde{X} \right) = (D\tilde{X})^j (F^0(t))$$

$$= (\mathbf{D}_t \bar{F}^j) \left(j_{(t,q)} X \right) \left(\frac{\mathrm{d}F^0}{\mathrm{d}t}(t) \right)^{-1}, \tag{4.7}$$

$$Q^{kl} y \circ jF \left(j_{(t,q)} X \right) = Q^{kl} y \left(j_{F(t,q)} \tilde{X} \right) = (Q\tilde{X})^{kl} (F^0(t))$$

$$= \left(\frac{\partial \bar{F}^k}{\partial x^i} \frac{\partial \bar{F}^l}{\partial x^j} \right) (t, X(t)) (QX)^{ij}(t) \left(\frac{\mathrm{d}F^0}{\mathrm{d}t}(t) \right)^{-1}. \tag{4.8}$$

4.2 Symmetries of SDEs

As an important application of the prolongations of diffusions and bundle homomorphisms, we now study the symmetries of stochastic differential equations. As in

classical Lie's theory of symmetries of ODEs, a symmetry of a stochastic differential equation is a space–time transformation that maps solutions to solutions. But this is not sufficient for the stochastic case. As we have mentioned in Sect. 4.1, the only smooth transformation on $\mathbb{R} \times M$ mapping diffusions to diffusions are bundle endomorphisms. Moreover, a solution of stochastic differential equation is always accompanied by a filtration, which will also be altered under space–time transformations. Thus, we have the following definition:

Definition 4.9 (*Symmetries*) Given a stochastic differential equation $S \subset \mathbb{R} \times \mathcal{T}^S M$, a symmetry of S is a bundle automorphism F on $(\mathbb{R} \times M, \pi, \mathbb{R})$ projecting to F^0 such that if $(X, \{\mathcal{P}_t\})$ is a solution of S, then so is $(F \cdot X, \{\mathcal{P}_{(F^0)^{-1}(s)}\})$.

Using the definitions of stochastic differential equations and pushforwards, we have the following equivalent characterization of symmetries.

Lemma 4.10 *Let S be a stochastic differential equation on M. A bundle automorphism F on $(\mathbb{R} \times M, \pi, \mathbb{R})$ is a symmetry of S, if and only if, whenever $j_{(t,q)}X \in S$ we have $jF(j_{(t,q)}X) \in S$, or equivalently, $jF(S) \subset S$.*

Recall that the infinitesimal version of bundle homomorphisms are the so called projectable or fiber-preserving vector fields. More precisely, a vector field V on $\mathbb{R} \times M$ is called π-projectable, if the (local) flow (or one-parameter group action) generated by V consists of (local) bundle endomorphisms on $(\mathbb{R} \times M, \pi, \mathbb{R})$ (cf. Olver 1998, Example 2.22 or Saunders 1989, Proposition 3.2.15). For such a vector field, we define its prolongation to be the infinitesimal generator of the prolongated flow.

Definition 4.11 (*Stochastic prolongations of projectable vector fields*) Let V be a π-projectable vector field on $\mathbb{R} \times M$, with corresponding (local) flow $\psi = \{\psi_\epsilon\}_{\epsilon \in (-\varepsilon, \varepsilon)}$. Then, the stochastic prolongation of V, denoted by jV, will be a vector field on the model jet bundle $\mathbb{R} \times \mathcal{T}^S M$, defined as the infinitesimal generator of the corresponding prolonged flow $\{j\psi_\epsilon\}_{\epsilon \in (-\varepsilon, \varepsilon)}$. In other words, jV is a vector field on $\mathbb{R} \times \mathcal{T}^S M$ defined by

$$jV\big|_{j_{(t,q)}X} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} (j\psi_\epsilon)(j_{(t,q)}X),$$

for any $j_{(t,q)}X \in \mathbb{R} \times \mathcal{T}^S M$.

Now we can define infinitesimal versions of symmetries.

Definition 4.12 (*Infinitesimal symmetries*) Let S be a stochastic differential equation on M. An infinitesimal symmetry of S is a π-projectable vector field V on $\mathbb{R} \times M$ whose stochastic prolongation jV is tangent to S.

The following properties follow straightforwardly from definitions.

Lemma 4.13 *Given a stochastic differential equation S on M, let V be a complete π-projectable vector field on $\mathbb{R} \times M$ and $\psi = \{\psi_\epsilon\}_{\epsilon \in \mathbb{R}}$ be its flow. Then,*

(i) *V is an infinitesimal symmetry of S if and only if $jV(\Theta) = 0$ for every local defining function Θ of S;*

(ii) *V is an infinitesimal symmetry of S if and only if for each $\epsilon \in \mathbb{R}$, ψ_ϵ is a symmetry of S.*

4.3 Stochastic Prolongation Formulae

We consider a coordinate chart $(\mathbb{R} \times U^{(1)}, (t, x^{(1)}))$ on the model jet bundle $\mathbb{R} \times \mathcal{T}^S M$, which is induced by the coordinate chart $(U, (x^i))$ on M. A π-projectable vector field V on $\mathbb{R} \times M$ has the following local coordinate representation

$$V_{(t,q)} = V^0(t)\frac{\partial}{\partial t}\Big|_t + V^i(t, q)\frac{\partial}{\partial x^i}\Big|_q. \tag{4.9}$$

Its prolongation jV is a vector field $\mathbb{R} \times \mathcal{T}^S M$ of the form

$$jV\big|_{j_{(t,q)}X} = V^0(t)\frac{\partial}{\partial t}\Big|_t + V^i(t, q)\frac{\partial}{\partial x^i}\Big|_{j_{(t,q)}X} + V_1^i(j_{(t,q)}X)\frac{\partial}{\partial D^i x}\Big|_{j_{(t,q)}X}$$

$$+ V_2^{jk}(j_{(t,q)}X)\frac{\partial}{\partial Q^{jk}x}\Big|_{j_{(t,q)}X}.$$

Now we use Lemma 4.8 to compute the coefficients V_1^i's and V_2^{jk}'s.

Theorem 4.14 *Suppose V is complete and π-projectable and has the local representation (4.9). Then, in the canonical coordinates $(t, x^{(1)})$, the coefficient functions of its prolongation jV are given by the following formulae:*

$$V_1^i\big(t, x^{(1)}\big) = \big(\mathbf{D}_t V^i\big)\big(t, x^{(1)}\big) - \dot{V}^0(t)D^i x, \tag{4.10}$$

$$V_2^{jk}\big(t, x^{(1)}\big) = \frac{\partial V^j}{\partial x^i}(t, x)Q^{ik}x + \frac{\partial V^k}{\partial x^i}(t, x)Q^{ij}x - \dot{V}^0(t)Q^{jk}x. \tag{4.11}$$

Proof Let $\psi = \{\psi_\epsilon\}_{\epsilon \in \mathbb{R}}$ be the flow generated by V. Since V is complete and π-projectable, each ψ_ϵ is a bundle endomorphism on $\mathbb{R} \times M$ projecting to a diffeomorphism on $\mathbb{R}$. Let $\psi_\epsilon(t, q) = (\psi_\epsilon^0(t), \bar{\psi}_\epsilon(t, q))$. Note that $\psi_0^0(t) = t$, $\bar{\psi}_0(t, q) = q$ and

$$V^0(t) = \frac{d}{d\epsilon}\Big|_{\epsilon=0}\psi_\epsilon^0(t), \quad V^i(t, q) = \frac{d}{d\epsilon}\Big|_{\epsilon=0}\bar{\psi}_\epsilon^i(t, q).$$

Let $X = \{X(t)\}_{t \in [t_0, \tau)}$ be a representative diffusion of $j_{(t_0,q)}X \in U^{(1)}$. Then, by Lemma 4.2 and Definition 4.4, a representative diffusion of $j\psi_\epsilon(j_{(t,q)}X)$ is

$$\tilde{X}_\epsilon(s) = \psi_\epsilon \cdot X(s) = \bar{\psi}_\epsilon\left(\big(\psi_\epsilon^0\big)^{-1}(s), X\big(\big(\psi_\epsilon^0\big)^{-1}(s)\big)\right), \quad s \in \big[\psi_\epsilon^0(t_0), \psi_\epsilon^0(\tau)\big).$$

Now we apply Lemma 4.8 and take derivatives with respect to ϵ. Since $\frac{d}{d\epsilon}$ commutes with the total mean derivative $\mathbf{D}_t$ as is clear from the coordinate representation, we have

$$
\begin{aligned}
V_1^i(j_{(t,q)}X) &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} (D\tilde{X}_\epsilon)^i \left(\psi_\epsilon^0(t)\right) \\
&= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left[\left(\mathbf{D}_t \bar{\psi}_\epsilon^i\right)\left(j_{(t,X(t))}X\right) \frac{d\left(\psi_\epsilon^0\right)^{-1}}{ds}\left(\psi_\epsilon^0(t)\right) \right] \\
&= \mathbf{D}_t V^i(j_{(t,q)}X) - (DX)^i(t)\dot{V}^0(t).
\end{aligned}
$$

Also,

$$
\begin{aligned}
V_2^{kl}(j_{(t,q)}X) &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} (Q\tilde{X}_\epsilon)^{kl}\left(\psi_\epsilon^0(t)\right) \\
&= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left[\left(\frac{\partial \bar{\psi}_\epsilon^k}{\partial x^i}\frac{\partial \bar{\psi}_\epsilon^l}{\partial x^j}\right)(t,X(t))\,(QX)^{ij}(t)\,\frac{d\left(\psi_\epsilon^0\right)^{-1}}{ds}\left(\psi_\epsilon^0(t)\right) \right] \\
&= \left(\frac{\partial V^k}{\partial x^i}\delta_j^l + \delta_i^k \frac{\partial V^l}{\partial x^j}\right)(t,X(t))(QX)^{ij}(t) - \delta_i^k \delta_j^l (QX)^{ij}(t)\dot{V}^0(t) \\
&= \frac{\partial V^k}{\partial x^i}(t,q)(QX)^{il}(t) + \frac{\partial V^l}{\partial x^j}(t,q)(QX)^{jk}(t) - (QX)^{kl}(t)\dot{V}^0(t).
\end{aligned}
$$

In the induced coordinate system $(t, x^{(1)}) = (t, x^i, D^i x, D_2^{jk}x)$, the last two formulae read as (4.10) and (4.11), respectively. ☐

Stochastic analogs of contact structure on $\mathbb{R} \times \mathcal{T}^S M$ and Cartan symmetries will be discussed in "Appendix B." It turns out that the infinitesimal symmetry of the mixed-order Cartan distribution is equivalent to stochastic prolongation formulae of Theorem 4.14.

Applying Theorem 4.14 to the system of mean differential equations (2.22), we have

Corollary 4.15 *The complete and π-projectable vector field V in (4.9) is an infinitesimal symmetry of MDEs (2.22) if and only if the coefficients V^0 and V^i's satisfy the following "determining equations":*

$$
\begin{aligned}
V^0 \frac{\partial \mathfrak{b}^i}{\partial t} + V^j \frac{\partial \mathfrak{b}^i}{\partial x^j} &= \frac{\partial V^i}{\partial t} + \frac{\partial V^i}{\partial x^j}\mathfrak{b}^j + \frac{1}{2}\frac{\partial^2 V^i}{\partial x^j \partial x^k}\sigma_r^j \sigma_r^k - \dot{V}^0 \mathfrak{b}^i, \\
V^0 \frac{\partial \left(\sigma_r^j \sigma_r^k\right)}{\partial t} + V^i \frac{\partial \left(\sigma_r^j \sigma_r^k\right)}{\partial x^i} &= \frac{\partial V^j}{\partial x^i}\sigma_r^i \sigma_r^k + \frac{\partial V^k}{\partial x^i}\sigma_r^i \sigma_r^j - \dot{V}^0 \sigma_r^j \sigma_r^k.
\end{aligned}
\tag{4.12}
$$

Proof We apply Lemma 4.13.(i) to (3.9) and then use Theorem 4.14, to get

$$
V^0 \frac{\partial \mathfrak{b}^i}{\partial t} + V^j \frac{\partial \mathfrak{b}^i}{\partial x^j} = \mathbf{D}_t V^i - \dot{V}^0 D^i x,
$$

$$V^0 \frac{\partial (\sigma_r^j \sigma_r^k)}{\partial t} + V^i \frac{\partial (\sigma_r^j \sigma_r^k)}{\partial x^i} = \frac{\partial V^j}{\partial x^i} Q^{ik} x + \frac{\partial V^k}{\partial x^i} Q^{ij} x - \dot{V}^0 Q^{jk} x.$$

Then, we use the coordinate representation (4.5) for the total mean derivative $\mathbf{D}_t$ and plug Eq. (3.9) in; the results follow. □

Remark 4.16 In Gaeta and Quintero (1999), the author proved a result similar to Corollary 4.15, with the following equation instead of Eq. (4.12):

$$V^0 \frac{\partial \sigma_r^j}{\partial t} + V^i \frac{\partial \sigma_r^j}{\partial x^i} = \frac{\partial V^j}{\partial x^i} \sigma_r^i - \frac{1}{2} \dot{V}^0 \sigma_r^j. \qquad (4.13)$$

By multiplying both sides of (4.13) with σ_r^k, and using the symmetry for index j, k, one gets easily (4.12). So our determining equations for infinitesimal symmetries are more general than those of Gaeta and Quintero (1999). Basically, the paper (Gaeta and Quintero 1999) concerns symmetries of the Itô equation $(\mathfrak{b}, \sigma)$, while we consider symmetries of the diffusion with generator $(\mathfrak{b}, \sigma \circ \sigma^*)$, or equivalently, a weak formulation of SDE. The former symmetries belong to the latter obviously, but not vice versa.

Now given a linear connection ∇ on M, we define the ∇-dependent versions of Definitions 4.1, 4.4 and 4.11. More precisely, for a diffusion X on M, we define its ∇-prolongation to be a TM-valued diffusion $j^\nabla X$ given by $j^\nabla X(t) = j^\nabla_{X(t)}(\theta_t X)$. For a bundle homomorphism from $F : (\mathbb{R} \times M, \pi, \mathbb{R}) \to (\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to a diffeomorphism $F^0 : \mathbb{R} \to \mathbb{R}$, the ∇-prolongation of F is the map $j^\nabla F : \mathbb{R} \times TM \to \mathbb{R} \times TN$ defined by $j^\nabla F(j^\nabla_{(t,q)} X) = j^\nabla_{F(t,q)}(F \cdot X)$. The ∇-prolongation of V, denoted by $j^\nabla V$, is defined to be the infinitesimal generator of the corresponding prolonged flow $\{j^\nabla \psi_\epsilon\}_{\epsilon \in (-\varepsilon, \varepsilon)}$, so that $j^\nabla V$ is a vector field on $\mathbb{R} \times TM$ and has the form

$$j^\nabla V \big|_{j^\nabla_{(t,q)} X} = V^0(t) \frac{\partial}{\partial t} \bigg|_t + V^i(t, q) \frac{\partial}{\partial x^i} \bigg|_{j^\nabla_{(t,q)} X} + V^i_\nabla \left(j^\nabla_{(t,q)} X \right) \frac{\partial}{\partial \dot{x}^i} \bigg|_{j^\nabla_{(t,q)} X},$$

for V of the form (4.9). If we denote $\bar{V} = V^i \frac{\partial}{\partial x^i}$ so that $V = V^0 + \bar{V}$, we have

Corollary 4.17 *Under the canonical coordinates $(t, x, \dot{x})$, the coefficient V^i_∇ of the ∇-prolongation $j^\nabla V$ are given by:*

$$V^i_\nabla(t, x, \dot{x}) = \left(\partial_t + \dot{x}^j \partial_j \right) V^i(t, x) + \frac{1}{2} Q^{jk} x \left[\nabla^2_{\partial_j, \partial_k} \bar{V} + R(\bar{V}, \partial_j) \partial_k \right]^i (t, x) - \dot{V}^0(t) \dot{x}^i,$$

where R is the curvature tensor.

Proof By (4.10) and (4.11), we have

$$
\begin{aligned}
V_\nabla^i(j_{(t,q)}X) &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} (D_\nabla \tilde{X}_\epsilon)^i\big(\psi_\epsilon^0(t)\big) \\
&= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left[(D\tilde{X}_\epsilon)^i\big(\psi_\epsilon^0(t)\big) + \frac{1}{2}\Gamma_{jk}^i(\tilde{X}_\epsilon(t))(Q\tilde{X}_\epsilon)^{jk}\big(\psi_\epsilon^0(t)\big) \right] \\
&= V_1^i(j_{(t,q)}X) + \frac{1}{2}\Gamma_{jk}^i(X(t))V_2^{jk}(j_{(t,q)}X) \\
&\quad + \frac{1}{2}\frac{\partial\Gamma_{jk}^i}{\partial x^l}(X(t))(QX)^{jk}(t)V^l(X(t)) \\
&= \left[\frac{\partial}{\partial t} + \left((D_\nabla X)^l(t) - \frac{1}{2}\Gamma_{jk}^l(X(t))(QX)^{jk}(t) \right)\frac{\partial}{\partial x^l} \right. \\
&\quad \left. + \frac{1}{2}(QX)^{jk}(t)\frac{\partial^2}{\partial x^j \partial x^k} \right] V^i(t,X(t)) - (DX)^i(t)\dot{V}^0(t) \\
&\quad + \frac{1}{2}\Gamma_{jk}^i(X(t)) \left[\frac{\partial V^j}{\partial x^l}(t,(X(t))(QX)^{kl}(t) \right. \\
&\quad \left. + \frac{\partial V^k}{\partial x^m}(t,(X(t))(QX)^{jm}(t) - (QX)^{jk}(t)\dot{V}^0(t) \right] \\
&\quad + \frac{1}{2}\frac{\partial\Gamma_{jk}^i}{\partial x^l}(X(t))(QX)^{jk}(t)V^l(t,X(t)) \\
&= \left[\frac{\partial}{\partial t} + (D_\nabla X)^l(t)\frac{\partial}{\partial x^l} \right] V^i(t,X(t)) \\
&\quad + \frac{1}{2}(Q_\nabla X)^{jk}(t)\left[\nabla_{\partial_j,\partial_k}^2 \bar{V} + R(\bar{V},\partial_j)\partial_k \right]^i(t,X(t)) \\
&\quad - (D_\nabla X)^i(t)\dot{V}^0(t).
\end{aligned}
$$

The proof is complete. □

5 The Second-Order Cotangent Bundle

5.1 Second-Order Covectors

Definition 5.1 (*Second-order cotangent space*) The second-order cotangent space at $q \in M$ is the dual vector space of $\mathcal{T}_q^O M$, denoted by $\mathcal{T}_q^{S*}M$. The pairing of $\alpha \in \mathcal{T}_q^{S*}M$ and $A \in \mathcal{T}_q^O M$ is denoted by $\langle \alpha, A \rangle$ or $\alpha(A)$. Elements of $\mathcal{T}_q^{S*}M$ are called second-order covectors at q. The disjoint union $\mathcal{T}^{S*}M := \amalg_{q \in M}\mathcal{T}_q^{S*}M$ is called the stochastic cotangent bundle of M. The natural projection map from $\mathcal{T}^{S*}M$ to M is denoted by τ_M^{S*}. A (local or global) smooth section of $\mathcal{T}^{S*}M$ is called a second-order covector field or a second-order form.

Dual to the left action (2.11) of G_I^d on fibers of $\mathcal{T}^S M$, G_I^d will act on those of $\mathcal{T}^{S*} M$ from the right.

Lemma 5.2 *The stochastic cotangent bundle* $(\mathcal{T}^{S*} M, \tau_M^{S*}, M)$ *is the fiber bundle dual to* $(\mathcal{T}^S M, \tau_M^S, M)$, *with structure group* G_I^d *acting on the typical fiber* $(\mathbb{R}^d \times \mathrm{Sym}^2(\mathbb{R}^d))^*$ *from the right by*

$$(p, o) \cdot (g, \kappa) = (g^* p, \kappa^* p + (g^* \otimes g^*) o),$$

for all $(g, \kappa) \in G_I^d$, $p \in (\mathbb{R}^d)^*$, $o \in (\mathrm{Sym}^2(\mathbb{R}^d))^*$.

The notion of second-order forms should not be confused with the classical one of 2-forms. There are two basic examples of second-order forms, say, $d^2 f$ and $df \cdot dg$, where f and g are given smooth functions on M. They are defined as follows: for $A \in \mathcal{T}^S M$,

$$\langle d^2 f, A \rangle := Af, \qquad \langle df \cdot dg, A \rangle := \Gamma_A(f, g) = A(fg) - fAg - gAf, \quad (5.1)$$

where Γ_A is the squared field operator defined in (2.8). These notations go back to Schwartz (1984) and Meyer (1981a) (see also Emery 1989, Chapters VI and VII), where the term $d^2 f$ is called the *second differential* of f, and the term $df \cdot dg$ is called the *symmetric product* of df and dg. Note that in these original references, there is a factor $\frac{1}{2}$ at the RHS of the definition of $df \cdot dg$. Here we drop this factor. Obviously, when restricted to TM, the second differential $d^2 f$ is just the differential df but the symmetric product $df \cdot dg$ vanishes.

The definition of the symmetric product $df \cdot dg$ yields two properties: $df \cdot dg$ is symmetric in f and g; and $(df \cdot dg)_q = 0$ if one of df_q and dg_q vanishes. These lead to a more general definition for symmetric products of two 1-forms. More precisely, let $\omega, \eta \in \mathcal{T}_q^* M$, then there exist smooth functions f and g on M such that $\omega = df_q$ and $\eta = dg_q$. By the preceding property, the second-order covector $(df \cdot dg)_q$ does not depend on the choice of f and g, and we will denote it by $\omega \cdot \eta$. Now if ω, η are second-order forms, then their symmetric product is defined pointwisely through $(\omega \cdot \eta)_q = \omega_q \cdot \eta_q$. More formally, we have

Definition 5.3 (*Symmetric product*, Emery 1989, Chapter VI) There exists a unique fiber-linear bundle homomorphism $\bullet$ from $T^* M \otimes T^* M$ to $\mathcal{T}^{S*} M$, which is called the symmetric product, such that for all $\omega, \eta \in T^* M$, $\bullet(\omega \otimes \eta) = \omega \cdot \eta$.

It is easy to verify from (5.1) that the local frame, dual to (2.12), for $(\mathcal{T}^{S*} M, \tau_M^{S*}, M)$ over the local chart $(U, (x^i))$ is given by (see also Emery 1989, Chapter VI)

$$\left\{ d^2 x^i, \tfrac{1}{2} dx^i \cdot dx^i, dx^j \cdot dx^k : 1 \le i \le d, 1 \le j < k \le d \right\}.$$

We adopt the convention that $dx^k \cdot dx^j = dx^j \cdot dx^k$ for all $1 \le j < k \le d$. Under this frame, a second-order covector $\alpha \in \mathcal{T}_q^{S*} M$ has a local expression

$$\alpha = \alpha_i d^2 x^i \big|_q + \tfrac{1}{2} \alpha_{jk} dx^j \cdot dx^k \big|_q, \tag{5.2}$$

where α^{jk} is symmetric in j, k. The coordinates (x^i) induce a canonical coordinate system on $\mathcal{T}^{S*}M$, denoted by (x^i, p_i, o_{jk}) and defined by

$$x^i(\alpha) = x^i(q), \quad p_i(\alpha) = \alpha_i, \quad o_{jk}(\alpha) = \alpha_{jk}. \tag{5.3}$$

for α in (5.2). Since the coefficients (α_i) do transform like a covector, as indicated in Lemma 5.2, it will cause no ambiguity to retain (x^i, p_i) as canonical coordinates on T^*M. As in classical geometric mechanics (Abraham and Marsden 1978; Holm et al. 2009), we still call the coordinates (p_i) the conjugate momenta. And we shall call the second-order coordinates (o_{jk}) the *conjugate diffusivities*.

The pairing of α and the second-order vector field A in (2.7) is then

$$\langle \alpha, A \rangle = \alpha_i A^i + \alpha_{jk} A^{jk}.$$

It follows from (5.1) and (2.8) that for smooths functions f and g on M,

$$d^2 f = \frac{\partial f}{\partial x^i} d^2 x^i + \frac{1}{2} \frac{\partial^2 f}{\partial x^j \partial x^k} dx^j \cdot dx^k, \qquad df \cdot dg = \frac{\partial f}{\partial x^i} \frac{\partial g}{\partial x^j} dx^i \cdot dx^j.$$

More generally, for 1-forms ω and η with local expressions $\omega = \omega_i dx^i$ and $\eta = \eta_i dx^i$, the symmetric product $\omega \cdot \eta$ has local expression

$$\omega \cdot \eta = \omega_i \eta_j dx^i \cdot dx^j. \tag{5.4}$$

Dual to the tangent case, there is indeed a canonical bundle epimorphism $\hat{\varrho}^*$: $(\mathcal{T}^{S*}M, \tau_M^{S*}, M) \to (T^*M, \tau_M^*, M)$, given by

$$\hat{\varrho}^*(\alpha) = \alpha|_{TM}.$$

In particular $\hat{\varrho}^*(d^2 f) = df$. In local coordinates, $\hat{\varrho}^*$ reads as

$$\hat{\varrho}^* \left(\alpha_i d^2 x^i|_q + \tfrac{1}{2} \alpha_{jk} dx^j \cdot dx^k|_q \right) = \alpha_i dx^i|_q,$$

The map $\hat{\varrho}^*$ is well defined since $\alpha|_{TM}$ is a covector. Clearly, $\hat{\varrho}^*$ is also a surjective submersion, so that $\mathcal{T}^{S*}M$ is a fiber bundle over T^*M. Occasionally, we will use the notation $\hat{\varrho}_M^*$ to indicate the base manifold M.

However, there is no canonical bundle monomorphism from T^*M to $\mathcal{T}^{S*}M$ which is a left inverse of $\hat{\varrho}^*$ and linear in fiber. We call such a bundle epimorphism a *fiber-linear bundle injection* from T^*M to $\mathcal{T}^{S*}M$. Similarly to Proposition 2.11, we also have a connection correspondence property. Namely, if we are given a linear connection ∇ on M, then it induces a fiber-linear bundle injection from T^*M to $\mathcal{T}^{S*}M$ by

$$\hat{\iota}_\nabla^* : T^*M \to \mathcal{T}^{S*}M, \quad dx^i|_q \mapsto d^2 x^i|_q + \tfrac{1}{2} \Gamma_{jk}^i(q) dx^j \cdot dx^k|_q =: d^\nabla x^i|_q, \tag{5.5}$$

or in local coordinates $\hat{\iota}_\nabla^*(x, p) = (x, p, (\Gamma_{jk}^i(x)p_i))$. Any fiber-linear bundle injection from T^*M to $\mathcal{T}^{S*}M$ induces a torsion-free linear connection on M.

Denote by $\text{Sym}^2(T^*M)$ the subbundle of $T^*M \otimes T^*M$ consisting of all $(0, 2)$-tensors on M. Then, the symmetric product $\bullet$, when restricting to $\text{Sym}^2(T^*M)$, is a bundle monomorphism whose image is the kernel of $\hat{\varrho}^*$. Conversely, still by the connection correspondence, a linear connection ∇ induces a fiber-linear bundle epimorphism from $\mathcal{T}^{S*}M$ to $\text{Sym}^2(T^*M)$ which is a right inverse of $\bullet$ and is given by

$$\varrho_\nabla^* : \mathcal{T}^{S*}M \to \text{Sym}^2(T^*M), \quad \alpha_i d^2 x^i|_q + \tfrac{1}{2}\alpha_{jk}dx^j \cdot dx^k|_q$$
$$\mapsto \left(\alpha_{jk} - \alpha_i \Gamma_{jk}^i(q)\right) dx^j \otimes dx^k|_q.$$

We introduce the ∇-dependent coordinates (o_{jk}^∇) by $o_{jk}^\nabla(\alpha) = \alpha_{jk} - \alpha_i\Gamma_{jk}^i(q)$ for α in (5.2), i.e.,

$$o_{jk}^\nabla = o_{jk} - p_i\left(\Gamma_{jk}^i \circ x\right). \tag{5.6}$$

Then, $\varrho_\nabla^*(\alpha) = o_{jk}^\nabla(\alpha)dx^j \otimes dx^k|_q$ and in particular

$$\varrho_\nabla^*(d^2 f) = \left(\frac{\partial^2 f}{\partial x^j \partial x^k} - \Gamma_{jk}^i \frac{\partial f}{\partial x^i}\right) dx^j \otimes dx^k = \nabla^2 f.$$

The coordinates $(x^i, p_i, o_{jk}^\nabla)$ form a coordinate system on $\mathcal{T}^{S*}M$, which we call the ∇-canonical coordinate system. The coordinates (x^i, o_{jk}^∇) also form a coordinate system on $\text{Sym}^2(T^*M)$ when restricted to it. We will call the coordinates (o_{jk}^∇) the *tensorial conjugate diffusivities*.

To sum up, we have the following short exact sequence which is split when a linear connection is provided:

$$0 \longrightarrow \text{Sym}^2(T^*M) \xrightarrow{\ \bullet\ } \mathcal{T}^{S*}M \xrightarrow{\ \hat{\varrho}^*\ } T^*M \longrightarrow 0. \tag{5.7}$$

It is easy to check that the bundle homomorphisms $\hat{\varrho}^*$, $\hat{\iota}_\nabla^*$, $\bullet$ and ϱ_∇^* are dual to ι, ϱ_∇, $\hat{\varrho}$ and $\hat{\iota}_\nabla$ in (2.13), (2.14), (2.15) and (2.16), respectively, so that the short exact sequence (5.7) is dual to (2.17). Similarly to (2.18), we have the following decomposition if a linear connection ∇ is given,

$$\mathcal{T}^{S*}M = \hat{\iota}_\nabla^*(T^*M) \oplus \bullet\left(\text{Sym}^2(T^*M)\right) \cong T^*M \oplus \text{Sym}^2(T^*M),$$

with fiber-wise isomorphism $\cong$ and first direct sum $\oplus$, which is given by

$$\alpha = \alpha_i d^\nabla x^i|_q + \tfrac{1}{2}\left(\alpha_{jk} - \alpha_i\Gamma_{jk}^i(q)\right) dx^j \cdot dx^k|_q$$
$$\mapsto \left(\alpha_i dx^i|_q, \left(\alpha_{jk} - \alpha_i\Gamma_{jk}^i(q)\right) dx^j \otimes dx^k|_q\right).$$

In particular,

$$d^2 f = \partial_i f \, d^\nabla x^i + \tfrac{1}{2} \nabla^2_{\partial_j, \partial_k} f \, dx^j \cdot dx^k \mapsto (df, \nabla^2 f). \tag{5.8}$$

Similarly to the classical cotangent space, the second-order cotangent space may be defined via germs. To be precise, we denote by $C^\infty_q(M)$ the set of all germs of smooth functions at $q \in M$, and define a equivalence relation between germs: $[f]_q, [g]_q \in C^\infty_q(M)$ are equivalent if and only if they have the same Taylor expansion at q higher than order zero and up to order two. Then, one can easily check that there is a one-to-one correspondence between $\mathcal{T}^{S*}_q M$ and the quotient space of $C^\infty_q(M)$ by this equivalence relation. Along this way, we can also observe the following diffeomorphism:

$$\mathcal{T}^{S*} M \times \mathbb{R} \cong J^2 \hat{\pi}, \tag{5.9}$$

by mapping $(d^2 f_q, f(q))$ to $j^2_q f$, where $J^2 \hat{\pi}$ is the classical second-order jet bundle of $(M \times \mathbb{R}, \hat{\pi}, M)$. This is similar to $T^*M \times \mathbb{R}$ is diffeomorphic to the first-order jet bundle $J^1 \hat{\pi}$ (e.g., Geiges 2008, Example 2.5.11 or Saunders 1989, Example 4.1.15). We denote the natural projection maps from $\mathcal{T}^{S*} M \times \mathbb{R}$ to $\mathbb{R}$ and from $T^*M \times \mathbb{R}$ to $\mathbb{R}$ by $\hat{\pi}^2_{0,1}$ and $\hat{\pi}^1_{0,1}$, respectively.

The relations and projection maps are integrated into the following commutative diagram:
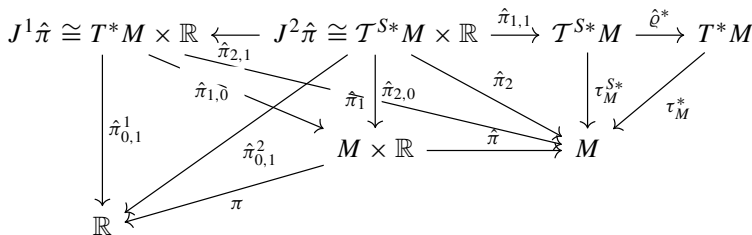
Remark 5.4 (i) As in Remark 3.5, given a linear connection ∇, we can obtain a one-to-one correspondence between $(T^*M \oplus \mathrm{Sym}^2(T^*M)) \times \mathbb{R}$ and $J^2 \hat{\pi}$ by mapping $(df_q, \nabla^2 f_q, f(q))$ to $j^2_q f$. One can find in Dahlqvist et al. (2019) an application of the jet-like structure on $T^*M \oplus \mathrm{Sym}^2(T^*M)$ and higher-order bundles to Martin Hairer's theory of regularity structures (Hairer 2014).

(ii) As we have seen, the product $\mathbb{R} \times \mathcal{T}^S M$ is the model bundle of the stochastic jet space $\mathcal{J}^S M$, while the product $\mathcal{T}^{S*} M \times \mathbb{R}$ is diffeomorphic to the second-order jet bundle $J^2 \hat{\pi}$. So, in a way, we can say that the "stochastic" and the "second-order" are dual to each other. This stochastic–second-order duality is somehow analogous to the particle–wave duality in quantum mechanics.

5.2 Second-Order Tangent and Cotangent Maps

Definition 5.5 (*Second-order tangent and cotangent maps*, Emery 1989, Chapter VI) Let M and N be two smooth manifolds, $F : M \to N$ be a smooth map. The second-order tangent map of F at $q \in M$ is a linear map $d^2 F_q : \mathcal{T}^S_q M \to \mathcal{T}^S_{F(q)} N$ defined

by

$$d^2 F_q(A)f = A(f \circ F), \quad \text{for } A \in \mathcal{T}_q^S M, f \in C^\infty(N).$$

The second-order cotangent map of F at $q \in M$ is a linear map $d^2 F_q^* : \mathcal{T}_{F(q)}^{S*} N \to \mathcal{T}_q^{S*} M$ dual to $d^2 F_q$, that is,

$$d^2 F_q^*(\alpha)(A) = \alpha(d^2 F_q(A)), \quad \text{for } A \in \mathcal{T}_q^S M, \alpha \in \mathcal{T}_{F(q)}^{S*} N.$$

The restrictions of $d^2 F_q$ to $T_q M$ coincide with the classical tangent map $d F_q$. But this is not the case for $d^2 F_q^*$ when restricting to $T_{F(q)}^* N$, since for $\alpha \in T_{F(q)}^* N$, $d^2 F_q^*(\alpha)$ is still a linear map on $\mathcal{T}_q^S M$. A manifestation of these phenomena may be seen through local coordinates in the following lemma.

Lemma 5.6 *Let* $(U, (x^i))$ *and* $(V, (y^j))$ *be local coordinate charts around q and $F(q)$, respectively. If*

$$A = A^i \frac{\partial}{\partial x^i}\Big|_q + A^{ij} \frac{\partial^2}{\partial x^i \partial x^j}\Big|_q \quad \text{and} \quad \alpha = \alpha_i d^2 y^i|_{F(q)} + \alpha_{ij} dy^i \cdot dy^j|_{F(q)}.$$

Then,

$$d^2 F_q(A) = (A F^i) \frac{\partial}{\partial y^i}\Big|_{F(q)} + \Gamma_A(F^i, F^j) \frac{\partial^2}{\partial y^i \partial y^j}\Big|_{F(q)},$$
$$d^2 F_q^*(\alpha) = \alpha_i d^2 F^i|_q + \alpha_{ij} d F^i \cdot d F^j|_q.$$

Now if $A \in \mathcal{T}_q M$, then all (A^{ij}) vanish and thereby so do $\Gamma_A(F^i, F^j)$'s. Thus, $d^2 F_q(A) = (A F^i) \frac{\partial}{\partial y^i}|_{F(q)} = d F_q(A)$. This makes clear that $d^2 F_q|_{\mathcal{T}_q M} = d F_q$. But if $\alpha \in \mathcal{T}_{F(q)}^* N$, then α^{ij}'s vanish and

$$d^2 F_q^*(\alpha) = \alpha_i d^2 F^i|_q = \alpha_i \frac{\partial F^i}{\partial x^j}(q) d^2 x^j|_q + \alpha_i \frac{\partial^2 F^i}{\partial x^j \partial x^k}(q) dx^j \cdot dx^k|_q,$$

while $d F_q^*(\alpha) = \alpha_i d F^i|_q = \alpha_i \frac{\partial F^i}{\partial x^j}(q) d^2 x^j|_q$. Hence, $d^2 F_q^*|_{\mathcal{T}_{F(q)}^* N} \neq d F_q^*$.

Definition 5.7 (*Second-order pushforwards and pullbacks*) Let $F : M \to N$ be smooth map. The second-order pushforward by F is a bundle homomorphism $F_*^S : (\mathcal{T}^S M, \tau_M^S, M) \to (\mathcal{T}^S N, \tau_N^S, N)$ defined by

$$F_*^S|_{\mathcal{T}_q^S M} = d^2 F_q.$$

Given a second-order form α on N, the second-order pullback of α by F is a second-order form $F^{S*}\alpha$ on M defined by

$$\left(F^{S*}\alpha\right)_q = d^2 F_q^* \left(\alpha_{F(q)}\right), \quad q \in M.$$

Let F be a diffeomorphism. The second-order pullback by F is a bundle isomorphism $F^{S*} : (\mathcal{T}^{S*}N, \tau_N^{S*}, N) \to (\mathcal{T}^{S*}M, \tau_M^{S*}, M)$ defined by

$$F^{S*}|_{\mathcal{T}_{q'}^{S*}N} = d^2 F_{F^{-1}(q')}^*.$$

Given a second-order vector field A on M, the second-order pushforward of A by F is a second-order vector field $F_*^S A$ on N defined by

$$(F_*^S A)_{q'} = d^2 F_{F^{-1}(q')} \left(A_{F^{-1}(q')} \right), \quad q' \in N.$$

Clearly, $F_*^S|_{TM} = F_*$ is the usual pushforward, but $F^{S*}|_{T^*N} \neq F^*$. The following properties are straightforward.

Lemma 5.8 *Let $F : M \to N$, $G : N \to K$ be two smooth maps. Let A be a second-order vector field on M and f, g be two smooth functions on N.*

(i) $G_*^S \circ F_*^S = (G \circ F)_*^S$.
(ii) *If F is a diffeomorphism, then $((F_*^S A)f) \circ F = A(f \circ F)$.*
(iii) $F^{S*}(d^2 f) = d^2(f \circ F)$, $F^{S*}(df \cdot dg) = d(f \circ F) \cdot d(g \circ F)$.

5.3 Mixed-Order Tangent and Cotangent Bundles

In this subsection, we will extend the notions of the previous two subsections to the product manifold $\mathbb{R} \times M$.

Definition 5.9 The mixed-order tangent bundle of $\mathbb{R} \times M$ is the product bundle (Saunders 1989, Definition 1.4.1) $(T\mathbb{R} \times \mathcal{T}^S M, \tau_\mathbb{R} \times \tau_M^S, \mathbb{R} \times M)$. The mixed-order cotangent bundle of $\mathbb{R} \times M$ is the product bundle $(T^*\mathbb{R} \times \mathcal{T}^{S*}M, \tau_\mathbb{R}^* \times \tau_M^{S*}, \mathbb{R} \times M)$. A section of the mixed-order tangent or cotangent bundle is called a mixed-order vector field or mixed-order form, respectively.

The mixed-order tangent and cotangent bundles are dual to each other. The mixed-order tangent (or cotangent) bundle is the bundle that mixes the first-order tangent (or cotangent) bundle in time and the second-order one in space (this is why we use the terminology "mixed-order"). It also matches the fundamental principle of stochastic analysis, whose Itô's logo is $(dX(t))^2 \sim dt$.

For an M-valued diffusion X with (time-dependent) generator A^X, we call the operator $\frac{\partial}{\partial t} + A^X$ its extended generator. This extended generator is a mixed-order vector field on $\mathbb{R} \times M$. Also note that the extended generator $\frac{\partial}{\partial t} + A^X$ of $X \in I_{t_0}(M)$ can be characterized by the property that for every $f \in C^\infty(\mathbb{R} \times M)$, the process

$$f(t, X(t)) - f(t_0, X(t_0)) - \int_{t_0}^t \left(\frac{\partial}{\partial t} + A^X \right) f(s, X(s)) ds, \quad t \geq t_0,$$

is a real-valued continuous $\{\mathcal{P}_t\}$-martingale. In general, a mixed-order vector field A has the following local expression:

$$A = A^0 \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + A^{jk} \frac{\partial^2}{\partial x^j \partial x^k}.$$

To give an example of mixed-order forms, we consider a smooth function f on $\mathbb{R} \times M$ and define in local coordinates

$$d^\circ f := \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x^i} d^2 x^i + \frac{1}{2} \frac{\partial^2 f}{\partial x^j \partial x^k} dx^j \cdot dx^k.$$

Then, $d^\circ f$ is a mixed-order form, and we call it the *mixed differential* of f. Clearly, the pairing of the mixed differential $d^\circ f$ and a mixed-order vector field A is $\langle d^\circ f, A \rangle = Af$.

Given a bundle homomorphism from $F : (\mathbb{R} \times M, \pi, \mathbb{R}) \to (\mathbb{R} \times N, \rho, \mathbb{R})$, we define its mixed-order tangent map at $(t, q) \in \mathbb{R} \times M$ by

$$d^\circ F_{(t,q)} = d^2 F_{(t,q)}|_{T_t \mathbb{R} \times \mathcal{T}_q^S M} : T\mathbb{R} \times \mathcal{T}^S M|_{(t,q)} \to T\mathbb{R} \times \mathcal{T}^S N|_{F(t,q)}.$$

Its mixed-order cotangent map at $(t, q) \in \mathbb{R} \times M$ is defined as the linear map $d^\circ F^*_{(t,q)} :$ $T^*\mathbb{R} \times \mathcal{T}^{S*} N|_{F(t,q)} \to T^*\mathbb{R} \times \mathcal{T}^{S*} M|_{(t,q)}$ dual to $d^\circ F_{(t,q)}$. If, moreover, F is a bundle isomorphism, its mixed-order pushforward and pullback, denoted by F_*^R and F^{R*}, respectively, can be defined in a similar manner to Definition 5.7. We leave their detailed but cumbersome definitions and properties to "Appendix A.1."

6 Stochastic Hamiltonian Mechanics

6.1 Horizontal Diffusions

In this subsection, we consider a general fiber bundle (E, π_M, M) over a manifold M, with fiber dimension n. We first introduce a special class of diffusions on this fiber bundle, which we call horizontal diffusions. They are defined in a similar fashion as the horizontal subspaces in Definition 3.7. Roughly speaking, a horizontal diffusion process on E is a diffusion that is random only "horizontally," but not on fibers.

Definition 6.1 (*Horizontal diffusions on fiber bundles*) Let (E, π_M, M) be a fiber bundle. A E-valued diffusion process $\mathbf{X}$ is said to be horizontal, if there exists an M-valued diffusion process X and a smoothly time-dependent section $\phi = (\phi_t)$ of π_M, such that a.s. $\mathbf{X}(t) = \phi(t, X(t))$ for all t.

The process X in the above definition is just the projection of $\mathbf{X}$, for $\pi_M(\mathbf{X}(t)) = \pi_M(\phi(t, X(t))) = X(t)$ a.s.. Since the projection map π_M is smooth, X is still a diffusion process.

Now we are going to define a subclass of "integral processes" for second-order vector fields on E by making use of horizontal diffusions. We use (x^i, u^μ) for an

adapted coordinate system on E (see Saunders 1989, Definition 1.1.5), where we use Greek alphabet to label the coordinates of fibers.

Given a second-order vector field with local expression

$$A = A^i \frac{\partial}{\partial x^i} + A^\mu \frac{\partial}{\partial u^\mu} + A^{jk} \frac{\partial^2}{\partial x^j \partial x^k} + A^{j\mu} \frac{\partial^2}{\partial x^j \partial u^\mu} + A^{\mu\nu} \frac{\partial^2}{\partial u^\mu \partial u^\nu}, \quad (6.1)$$

where A^i, A^μ, A^{jk}, $A^{j\mu}$, $A^{\mu\nu}$ are smooth functions in the local chart of E, by a *horizontal integral process* of A in (6.1) we mean an E-valued horizontal diffusion process **X** such that **X** is an integral process of A in the sense of (2.22), that is, it is determined by the system

$$\begin{cases} (D(x \circ \mathbf{X}))^i(t) = A^i(\mathbf{X}(t)), \\ (Q(x \circ \mathbf{X}))^{jk}(t) = 2A^{jk}(\mathbf{X}(t)), \\ (D(u \circ \mathbf{X}))^\mu(t) = A^\mu(\mathbf{X}(t)), \\ (Q(x \circ \mathbf{X}, u \circ \mathbf{X}))^{j\nu}(t) = A^{j\mu}(\mathbf{X}(t)), \\ (Q(u \circ \mathbf{X}))^{\mu\nu}(t) = 2A^{\mu\nu}(\mathbf{X}(t)), \end{cases} \quad (6.2)$$

where the expression $x \circ \mathbf{X}$ means that the family of coordinate functions (x^i) acts on **X**, and so on. Set $\mathbf{X}(t) = \phi(t, X(t))$ for some time-dependent section ϕ of π_M and M-valued diffusion X. Denote $\phi^\mu = u^\mu \circ \phi$. By Itô's formula, the system (6.2) can be written as

$$\begin{cases} (DX)^i(t) = A^i(\phi(t, X(t))), \\ (QX)^{jk}(t) = 2A^{jk}(\phi(t, X(t))), \\ \left(\frac{\partial}{\partial t} + A^i(\phi(t, X(t))) \frac{\partial}{\partial x^i} + A^{jk}(\phi(t, X(t))) \frac{\partial^2}{\partial x^j \partial x^k} \right) \phi^\mu(t, X(t)) \\ \qquad = A^\mu(\phi(t, X(t))) \\ 2A^{jk}(\phi(t, X(t))) \frac{\partial \phi^\mu}{\partial x^k}(t, X(t)) = A^{j\mu}(\phi(t, X(t))) \\ A^{jk}(\phi(t, X(t))) \frac{\partial \phi^\mu}{\partial x^j} \frac{\partial \phi^\nu}{\partial x^k}(t, X(t)) = A^{\mu\nu}(\phi(t, X(t))). \end{cases} \quad (6.3)$$

If $X(t)$ has full support for all t, then the last three equations in (6.3) translate into a system of (possibly degenerate) parabolic equations on E,

$$\begin{cases} \left(\frac{\partial}{\partial t} + A^i(\phi(t, q)) \frac{\partial}{\partial x^i} + A^{jk}(\phi(t, q)) \frac{\partial^2}{\partial x^j \partial x^k} \right) \phi^\mu(t, q) = A^\mu(\phi(t, q)), \\ 2A^{jk}(\phi(t, q)) \frac{\partial \phi^\mu}{\partial x^k}(t, q) = A^{j\mu}(\phi(t, q)) \\ A^{jk}(\phi(t, q)) \frac{\partial \phi^\mu}{\partial x^j} \frac{\partial \phi^\nu}{\partial x^k}(t, q) = A^{\mu\nu}(\phi(t, q)). \end{cases} \quad (6.4)$$

Therefore, under suitable assumptions for the coefficients A^i, A^μ, A^{jk}, $A^{j\mu}$, $A^{\mu\nu}$, Eq. (6.4) is solvable, at least locally, by some time-dependent local section $\phi = (\phi_t)$ over a time interval $[0, T]$. Then, plugging $\phi(t)$ into the first two equations of (6.3), we can find X and hence $\mathbf{X}$. We call X an *projective integral process* of A.

6.2 The Second-Order Symplectic Structure on $\mathcal{T}^{S*}M$ and Stochastic Hamilton's Equations

It is well known that the classical cotangent bundle T^*M has a natural symplectic structure, given by the canonical symplectic form $\omega_0 = dx^i \wedge dp_i$, where (x^i, p_i) are the canonical local coordinates on T^*M induced by local coordinates (x^i) on M. Clearly ω_0 is closed, because it is exact as $\omega_0 = -d\theta_0$, where $\theta_0 = p_i dx^i$ is called the Poincaré (or tautological) 1-form.

Now we need to define a similar structure on the second-order cotangent bundle $\mathcal{T}^{S*}M$, which is a second-order counterpart of the symplectic structure. Firstly, we adapt the coordinate-free definition of the tautological 1-form to the second-order case.

Definition 6.2 The second-order tautological form θ is a second-order form on $\mathcal{T}^{S*}M$ defined by

$$\theta_\alpha = d^2 \left(\tau_M^{S*}\right)_\alpha^* (\alpha), \quad \alpha \in \mathcal{T}_q^{S*}M.$$

Under the induced coordinate system on $\mathcal{T}^{S*}M$ defined in (5.3), the second-order tautological form θ has the following coordinate representation

$$\theta = p_i d^2 x^i + \tfrac{1}{2} o_{jk} dx^j \cdot dx^k. \tag{6.5}$$

We introduce the canonical second-order symplectic form ω on $\mathcal{T}^{S*}M$ by writing $\omega = -d^2\theta$. Although we do not define the exterior differential for second-order forms, we can still take d^2 formally on both sides of (6.5), using Leibniz's rule and the composition rule $d \circ d = d^2$ (cf. Meyer 1981b, Section 6.(e)), and forcing $d^3 = 0$ and $(d^2-) \cdot (d-) = (d-) \cdot (d^2-) = 0$. Then, we get

$$\begin{aligned} \omega &= d \left(d^2 x^i \wedge dp_i + \tfrac{1}{2} dx^j \cdot dx^k \wedge do_{jk} - p_i d^3 x^i + o_{jk} d^2 x^j \wedge dx^k \right) \\ &= d^2 x^i \wedge d^2 p_i + \tfrac{1}{2} dx^j \cdot dx^k \wedge d^2 o_{jk}. \end{aligned} \tag{6.6}$$

We call the pair $(\mathcal{T}^{S*}M, \omega)$ a second-order symplectic manifold. The complete axiom system for a second-order differential system $(d, d^2, \wedge, \cdot)$ is beyond the scope of this paper.

Remark 6.3 In the formal expression $(d \circ d) f = d^2 f$, $f \in C^\infty(M)$, the two differential operators d at LHS are *different*. The second d is still de Rham's exterior differential on M, while the first needs to be understood as the exterior differential on TM by

regarding the first differential df as a function on TM. Thus, the complete expression should be $d_{TM} \circ d_M = d^2$. Along this way, the differential operator d_{TM} can be extended to a linear transform that maps 1-forms to second-order forms and satisfies Leibniz's rule, see Emery (1989, Theorem 7.1). We shall denote the linear operator extended from d_{TM} by $\boldsymbol{d}$ in order to distinguish. In local coordinates, it acts on a 1-form $\eta = \eta_i dx^i$ by $\boldsymbol{d}\eta = \eta_i d^2 x^i + \frac{1}{2} \frac{\partial \eta_i}{\partial x^j} dx^i \cdot dx^j$, so that $\hat{\varrho}^*(\boldsymbol{d}\eta) = \eta$ and $d^2 = \boldsymbol{d} \circ d$. When a linear connection ∇ is specified, $\boldsymbol{d}\eta = \eta_i d^\nabla x^i + \frac{1}{2} \nabla \eta(\partial_i, \partial_j) dx^i \cdot dx^j$ which covers (5.8).

As in the classical case, we have the following property for the second-order tautological form.

Lemma 6.4 *The second-order tautological form θ is the unique second-order form on $\mathcal{T}^{S*}M$ with the property that, for every second-order form α on M, $\alpha^{S*}\theta = \alpha$.*

Proof From Lemma 5.8, we have, for any second-order vector $A \in \mathcal{T}_q^S M$,

$$\langle (\alpha^{S*}\theta)_q, A \rangle = \langle \theta_{\alpha_q}, d^2\alpha_q(A) \rangle = \langle d^2(\tau_M^{S*})^*_{\alpha_q}(\alpha_q), d^2\alpha_q(A) \rangle$$
$$= \langle \alpha_q, d^2(\tau_M^{S*})_{\alpha_q} \circ d^2\alpha_q(A) \rangle = \langle \alpha_q, A \rangle,$$

since $\tau_M^{S*} \circ \alpha = \mathbf{Id}_M$. $\qquad\qquad\square$

Recall that, in Definition 5.7, we have defined the second-order pullbacks of second-order forms. Now, given a smooth map $\mathbf{F} : \mathcal{T}^{S*}M \to \mathcal{T}^{S*}N$ and a second-order 2-form η on $\mathcal{T}^{S*}N$, we may also define the second-order pullback $\mathbf{F}^{S*}\eta$ of η by $\mathbf{F}$ by allowing $\mathbf{F}^{S*}$ to be exchangeable with the symmetric product $\cdot$ as well as the wedge product $\wedge$. Then, as a corollary of Lemma 6.4, we have

$$\alpha^{S*}\omega = -d^2\alpha.$$

Definition 6.5 Let ω and η be the canonical second-order symplectic forms on $\mathcal{T}^{S*}M$ and $\mathcal{T}^{S*}N$, respectively. A bundle homomorphism $\mathbf{F} : (\mathcal{T}^{S*}M, \hat{\varrho}_M^*, T^*M) \to (\mathcal{T}^{S*}N, \hat{\varrho}_N^*, T^*N)$ is called second-order symplectic or a second-order symplectomorphism if $\mathbf{F}^{S*}\eta = \omega$.

Theorem 6.6 *Let $F : N \to M$ be a diffeomorphism. The second-order pullback $F^{S*} : \mathcal{T}^{S*}M \to \mathcal{T}^{S*}N$ by F is second-order symplectic; in fact $(F^{S*})^{S*}\vartheta = \theta$, where ϑ is the second-order tautological form on $\mathcal{T}^{S*}N$.*

Proof For $q \in M$, $\alpha_q \in \mathcal{T}_q^{S*}M$ and $A \in \mathcal{T}_{\alpha_q}^S T^{S*}M$,

$$
\begin{aligned}
\langle (F^{S*})^{S*} \vartheta, A \rangle &= \langle \vartheta, d^2 (F^{S*})_{\alpha_q} A \rangle = \langle d^2 (\tau_N^{S*})_{F^{S*}(\alpha_q)}^* (F^{S*}(\alpha_q)), d^2 (F^{S*})_{\alpha_q} A \rangle \\
&= \langle F^{S*}(\alpha_q), d^2 (\tau_N^{S*})_{F^{S*}(\alpha_q)} \circ d^2 (F^{S*})_{\alpha_q} A \rangle \\
&= \langle \alpha_q, d^2 F_{F^{-1}(q)} \circ d^2 (\tau_N^{S*})_{F^{S*}(\alpha_q)} \circ d^2 (F^{S*})_{\alpha_q} A \rangle \\
&= \langle \alpha_q, d^2 (\tau_M^{S*})_{\alpha_q} A \rangle \\
&= \langle d^2 (\tau_M^{S*})_{\alpha_q}^* (\alpha_q), A \rangle \\
&= \langle \theta_{\alpha_q}, A \rangle,
\end{aligned}
$$

where we used the fact that $F \circ \tau_N^{S*} \circ F^{S*} = \tau_M^{S*}$ in the fourth line. $\qquad\square$

Clearly, the counterparts of Hamiltonian vector fields on T^*M are now second-order vector fields on $\mathcal{T}^{S*}M$. Remark that for a second-order vector field A on $\mathcal{T}^{S*}M$, the form $A \lrcorner \omega$ take values in the cotangent bundle $\mathcal{T}^{S*}\mathcal{T}^{S*}M$.

Definition 6.7 Let $H : \mathcal{T}^{S*}M \to \mathbb{R}$ be a given smooth function. A second-order vector field A_H on $\mathcal{T}^{S*}M$ satisfying

$$
A_H \lrcorner \omega = d^2 H \tag{6.7}
$$

is called a second-order Hamiltonian vector field of H. We call the triple $(\mathcal{T}^{S*}M, \omega, H)$ a second-order Hamiltonian system. The function H is called the second-order Hamiltonian of the system.

According to (6.7), the second-order vector field A_H satisfies

$$
A_H H = d^2 H (A_H) = \omega(A_H, A_H) = 0. \tag{6.8}
$$

The condition (6.7) cannot uniquely determine A_H. It is easy to verify that A_H is of the general form

$$
\begin{aligned}
A_H = {} & \frac{\partial H}{\partial p_i} \frac{\partial}{\partial x^i} - \frac{\partial H}{\partial x^i} \frac{\partial}{\partial p_i} + \frac{\partial H}{\partial o_{jk}} \frac{\partial^2}{\partial x^j \partial x^k} - \left(\frac{\partial^2 H}{\partial x^j \partial x^k} + C_{jk} \right) \frac{\partial}{\partial o_{jk}} \\
& + A_{jk} \frac{\partial^2}{\partial p_j \partial p_k} + A_{ijkl} \frac{\partial^2}{\partial o_{ij} \partial o_{kl}} + A_k^j \frac{\partial^2}{\partial x^j \partial p_k} \\
& + A_{kl}^j \frac{\partial^2}{\partial x^j \partial o_{kl}} + A_{jkl} \frac{\partial^2}{\partial p_j \partial o_{kl}},
\end{aligned} \tag{6.9}
$$

where the coefficients C_{jk}, A_{jk}, A_{ijkl}, A_k^j, A_{kl}^j, A_{jkl} are smooth functions on local chart satisfying

$$C_{jk}\frac{\partial H}{\partial o_{jk}} = A_{jk}\frac{\partial^2 H}{\partial p_j \partial p_k} + A_{ijkl}\frac{\partial^2 H}{\partial o_{ij}\partial o_{kl}} + A_k^j\frac{\partial^2 H}{\partial x^j \partial p_k}$$

$$+A_{kl}^j\frac{\partial^2 H}{\partial x^j \partial o_{kl}} + A_{jkl}\frac{\partial^2 H}{\partial p_j \partial o_{kl}},$$

such that the local expression (6.9) is invariant under the canonical change of coordinates on $\mathcal{T}^{S*}M$ induced by a change of coordinates on M, governed by the structure group in Lemma 5.2.

Given such a second-order Hamiltonian vector field of H, its horizontal integral process is a $\mathcal{T}^{S*}M$-valued horizontal diffusion $\mathbf{X}$ determined by the following MDEs on $\mathcal{T}^{S*}M$,

$$
\begin{cases}
(D(x \circ \mathbf{X}))^i(t) = \dfrac{\partial H}{\partial p_i}(\mathbf{X}(t)), \\[2mm]
(Q(x \circ \mathbf{X}))^{jk}(t) = 2\dfrac{\partial H}{\partial o_{jk}}(\mathbf{X}(t)), \\[2mm]
(D(p \circ \mathbf{X}))_i(t) = -\dfrac{\partial H}{\partial x^i}(\mathbf{X}(t)), \\[2mm]
(D(o \circ \mathbf{X}))_{jk}(t) = -\left(\dfrac{\partial^2 H}{\partial x^j \partial x^k} + C_{jk}\right)(\mathbf{X}(t)), \\[2mm]
\left(C_{ij}\dfrac{\partial H}{\partial o_{ij}}\right)(\mathbf{X}(t)) = \dfrac{1}{2}(Q(p \circ \mathbf{X}))_{jk}(t)\dfrac{\partial^2 H}{\partial p_j \partial p_k}(\mathbf{X}(t)) \\[2mm]
\qquad\qquad + \dfrac{1}{2}(Q(o \circ \mathbf{X}))_{ijkl}(t)\dfrac{\partial^2 H}{\partial o_{ij}\partial o_{kl}}(\mathbf{X}(t)) \\[2mm]
\qquad\qquad + (Q(x \circ \mathbf{X}, p \circ \mathbf{X}))_k^j\dfrac{\partial^2 H}{\partial x^j \partial p_k}(\mathbf{X}(t)) \\[2mm]
\qquad\qquad + (Q(x \circ \mathbf{X}, o \circ \mathbf{X}))_{kl}^j\dfrac{\partial^2 H}{\partial x^j \partial o_{kl}}(\mathbf{X}(t)) \\[2mm]
\qquad\qquad + (Q(p \circ \mathbf{X}, o \circ \mathbf{X}))_{jkl}\dfrac{\partial^2 H}{\partial p_j \partial o_{kl}}(\mathbf{X}(t)),
\end{cases} \tag{6.10}
$$

or, in coordinates,

$$
\begin{cases}
D^i x = \dfrac{\partial H}{\partial p_i}, \\[2mm]
Q^{jk} x = 2\dfrac{\partial H}{\partial o_{jk}}, \\[2mm]
D_i p = -\dfrac{\partial H}{\partial x^i}, \\[2mm]
D_{jk} o = -\left(\dfrac{\partial^2 H}{\partial x^j \partial x^k} + C_{jk} \right), \\[2mm]
C_{ij} \dfrac{\partial H}{\partial o_{ij}} = \dfrac{1}{2} Q_{jk} p \dfrac{\partial^2 H}{\partial p_j \partial p_k} + \dfrac{1}{2} Q_{ijkl} o \dfrac{\partial^2 H}{\partial o_{ij} \partial o_{kl}} + Q_k^j(x, p) \dfrac{\partial^2 H}{\partial x^j \partial p_k} \\[2mm]
\qquad\qquad + Q_{kl}^j(x, o) \dfrac{\partial^2 H}{\partial x^j \partial o_{kl}} + Q_{jkl}(p, o) \dfrac{\partial^2 H}{\partial p_j \partial o_{kl}},
\end{cases}
$$

where $\left(x^i, p_i, o_{jk}, D^i x, D_i p, D_{jk} o, Q^{jk} x, Q_{jk} p, Q_{ijkl} o, Q_k^j(x, p), Q_{kl}^j(x, o), Q_{jkl}(p, o) \right)$ are canonical coordinates on $\mathcal{T}^S \mathcal{T}^{S*} M$. The first and third equations has been conjectured in Zambrini (2015) as stochastic Hamilton's equations in the Euclidean space, since they have the same form as classical Hamilton's equations (e.g., Abraham and Marsden 1978, Proposition 3.3.2) except that mean derivative D replaces classical time derivative.

At first glance, one may think that the system (6.10) is underdetermined, as there are fewer equations than unknowns (the number of unknowns is equal to the fiber dimension of $\mathcal{T}^S \mathcal{T}^{S*} M$). Besides, we haven not yet given (6.10) initial or terminal data. These will become clear after we make the following observations. Firstly, the first two equations of (6.10) constitute MDEs that are equivalent to an Itô SDE for $x(\mathbf{X})$ in weak sense, as we have seen in Sect. 2.4. So $x(\mathbf{X})$ should be assigned an initial value, say,

$$
\text{Law}((x \circ \mathbf{X})(0)) = \mu_0, \tag{6.11}
$$

where μ_0 is a given probability measure on M. Secondly, in the third and fourth equations of (6.10), only the "drift" information of $p(\mathbf{X})$ and $o(\mathbf{X})$ is clear. To overcome the lack of information, we need to assign $p(\mathbf{X})$ and $o(\mathbf{X})$ terminal values, say,

$$
\begin{cases}
(p \circ \mathbf{X})(T) = p^*(x \circ \mathbf{X}(T)), \\
(o \circ \mathbf{X})(T) = o^*(x \circ \mathbf{X}(T)),
\end{cases} \tag{6.12}
$$

where (p^*, o^*) is a given second-order form. Therefore, the third and fourth equations are understood as backward SDEs, whose drifts rely on diffusion coefficients via the last equation. The system (6.10) together with boundary values (6.11) and (6.12) could be understood as a (coupled) forward–backward system of SDEs (Yong and Zhou 1999) (where "backward" is taken in a different sense from ours in Sect. 2).

Notice that those forward–backward SDEs are not necessarily solvable (see Yong and Zhou 1999, Proposition 7.5.2 for an example). In order to solve (6.10)–(6.12), we have to take the horizontal condition into consideration, and make some compatibility assumption. More precisely, we set $X = \tau_M^{S*}(\mathbf{X})$ and

$$\mathbf{X}(t) = \alpha(t, X(t)), \tag{6.13}$$

for some time-dependent second-order form α on M, and denote $p_i(t, x) = p_i(\alpha(t, x))$ and $o_{jk}(t, x) = o_{jk}(\alpha(t, x))$, so that $\alpha(t, x) = (p(t, x), o(t, x))$. Assume that for each $t \in (0, T)$, $X(t)$ has full support. Then, by applying Itô's formula, in the same way as in (6.4), the system (6.10) reduces to

$$\begin{cases} \left(\dfrac{\partial}{\partial t} + \dfrac{\partial H}{\partial p_j} \dfrac{\partial}{\partial x^j} + \dfrac{\partial H}{\partial o_{jk}} \dfrac{\partial^2}{\partial x^j \partial x^k} \right) p_i = -\dfrac{\partial H}{\partial x^i}, \\[2mm] \left(\dfrac{\partial}{\partial t} + \dfrac{\partial H}{\partial p_k} \dfrac{\partial}{\partial x^k} + \dfrac{\partial H}{\partial o_{kl}} \dfrac{\partial^2}{\partial x^k \partial x^l} \right) o_{ij} = -\left(\dfrac{\partial^2 H}{\partial x^i \partial x^j} + C_{ij} \right), \\[2mm] C_{ij} \dfrac{\partial H}{\partial o_{ij}} = \dfrac{\partial H}{\partial o_{ij}} \left(\dfrac{\partial p_k}{\partial x^i} \dfrac{\partial p_l}{\partial x^j} \dfrac{\partial^2 H}{\partial p_k \partial p_l} + \dfrac{\partial o_{kl}}{\partial x^i} \dfrac{\partial o_{mn}}{\partial x^j} \dfrac{\partial^2 H}{\partial o_{kl} \partial o_{mn}} \right. \\[2mm] \qquad\quad \left. + 2\dfrac{\partial p_k}{\partial x^i} \dfrac{\partial^2 H}{\partial x^j \partial p_k} + 2\dfrac{\partial o_{kl}}{\partial x^i} \dfrac{\partial^2 H}{\partial x^j \partial o_{kl}} + 2\dfrac{\partial p_k}{\partial x^i} \dfrac{\partial o_{lm}}{\partial x^j} \dfrac{\partial^2 H}{\partial p_k \partial o_{lm}} \right). \end{cases} \tag{6.14}$$

Next, by taking partial derivative $\frac{\partial}{\partial x^j}$ on both sides of the first equation of (6.14) and comparing with the next two, we find the following sufficient condition for the last two equations of (6.14):

$$o_{ij}(t, x) = \frac{\partial p_i}{\partial x^j}(t, x) = \frac{\partial p_j}{\partial x^i}(t, x), \tag{6.15}$$

or equivalent, for the terminal value (p^*, o^*),

$$o_{ij}^*(x) = \frac{\partial p_i^*}{\partial x^j}(x) = \frac{\partial p_j^*}{\partial x^i}(x). \tag{6.16}$$

Equation (6.15) implies that α in (6.13) is "exact," in the sense that $\alpha = d\eta$ for the time-dependent 1-form $\eta = p_i dx^i$, where d is the extended differential operator defined in Remark 6.3. Similarly, Eq. (6.16) implies that $(p^*, o^*) = d\eta^*$ for 1-form $\eta^* = p_i^* dx^i$. The second equality of (6.15) [or (6.16)], called *Onsager reciprocity* or *Maxwell relations* (Abraham and Marsden 1978, Section 5.3), implies that the 1-form η (or η^*) is closed. We will refer to Eq. (6.15) or (6.16) as second-order Maxwell relations.

Under the second-order Maxwell relations, the original stochastic Hamilton's system (6.10) turns to the following MDE-PDE coupled system.

$$
\begin{cases}
(DX)^i(t) = \dfrac{\partial H}{\partial p_i}(X(t), p(t, X(t)), o(t, X(t))), \\[2mm]
(QX)^{jk}(t) = 2\dfrac{\partial H}{\partial o_{jk}}(X(t), p(t, X(t)), o(t, X(t))), \\[2mm]
\left(\dfrac{\partial}{\partial t} + \dfrac{\partial H}{\partial p_j}(x, p(t,x), o(t,x)) \dfrac{\partial}{\partial x^j} + \dfrac{\partial H}{\partial o_{jk}}(x, p(t,x), o(t,x)) \dfrac{\partial^2}{\partial x^j \partial x^k} \right) p_i(t, x) \\[2mm]
\qquad = -\dfrac{\partial H}{\partial x^i}(x, p(t,x), o(t,x)), \\[2mm]
o_{ij}(t, x) = \dfrac{\partial p_i}{\partial x^j}(t, x),
\end{cases}
\tag{6.17}
$$

The boundary values in (6.11) and (6.12) now read

$$
\text{Law}(X(0)) = \mu_0, \quad (p, o)(T) = \boldsymbol{d}\eta^*.
\tag{6.18}
$$

We first use the terminal value in (6.18), which satisfies (6.16), to solve the last two PDEs in (6.17). This gives (p, o) and hence the second-order form α. Then, we plug p and o into the first two MDEs and solve them with initial distribution in (6.18). This yields in law the M-valued diffusion $X = \tau_M^{S*}(\mathbf{X})$ as a projective integral process of A_H.

We call system (6.10) or (6.17) the *stochastic Hamilton's equations* (S-H equations in short). The second-order Maxwell relations are sufficient for the component o of α in (6.13) to solve the last two equations of (6.10), so we refer to it as an *integrability condition* of (6.10). When restricting settings to Riemannian manifolds, the S-H equations (6.10) can be simplified to a global Hamiltonian-type system on T^*M, as we will see in Sect. 7.4.2.

Lemma 6.8 *Let $H : \mathcal{T}^{S*}M \times \mathbb{R} \to \mathbb{R}$ be a time-dependent second-order Hamiltonian, and $\mathbf{X}$ be a horizontal integral process of A_H. Then, the total mean derivative of H along $\mathbf{X}$ is*

$$
\mathbf{D}_t H = \frac{\partial H}{\partial t}.
$$

Proof We use (6.10) and local coordinates to derive

$$
\begin{aligned}
\mathbf{D}_t H &= D[H(\mathbf{X}(t), t)] \\
&= \frac{\partial H}{\partial t} + D^i x \frac{\partial H}{\partial x^i} + D_i p \frac{\partial H}{\partial p_i} + D_{jk} o \frac{\partial H}{\partial o_{jk}} + \frac{1}{2} Q^{jk} x \frac{\partial^2 H}{\partial x^j \partial x^k} \\
&\quad + \frac{1}{2} Q_{jk} p \frac{\partial^2 H}{\partial p_j \partial p_k} + \frac{1}{2} Q_{ijkl} o \frac{\partial^2 H}{\partial o_{ij} \partial o_{kl}} + Q_k^j(x, p) \frac{\partial^2 H}{\partial x^j \partial p_k} \\
&\quad + Q_{kl}^j(x, o) \frac{\partial^2 H}{\partial x^j \partial o_{kl}} + Q_{jkl}(p, o) \frac{\partial^2 H}{\partial p_j \partial o_{kl}} \\
&= \frac{\partial H}{\partial t} + D^i x \frac{\partial H}{\partial x^i} + D_i p \frac{\partial H}{\partial p_i} + D_{jk} o \frac{\partial H}{\partial o_{jk}} + \frac{1}{2} Q^{jk} x \frac{\partial^2 H}{\partial x^j \partial x^k} + C_{ij} \frac{\partial H}{\partial o_{ij}} \\
&= \frac{\partial H}{\partial t}.
\end{aligned}
$$

The result follows. $\qquad\square$

In particular, when H is time-independent, we have

$$
\mathbf{D}_t H = 0, \tag{6.19}
$$

which is also a consequence of (6.8). Equivalently, H is harmonic with respect to the horizontal integral process $\mathbf{X}$. In this case, we can say that H is stochastically conserved, or is a stochastic conserved quantity. In particular, the expectation $\mathbf{E}[H(\mathbf{X})]$ is a constant.

6.3 Two Inspirational Examples

Let M be a Riemannian manifold with Riemannian metric g. Assume for simplicity that M is compact. Let ∇ be the Levi–Civita connection on TM with Christoffel symbols (Γ_{ij}^k). In this subsection, we will consider two types of processes on M, to provide some intuition of our stochastic Hamiltonian formalism.

6.3.1 Diffusion Processes on Riemannian Manifolds

Consider a second-order Hamiltonian H on $\mathcal{T}^{S*} M$ with the following coordinate expression:

$$
H(x, p, o) = b^i(x) p_i - \frac{1}{2} g^{ij}(x) \Gamma_{ij}^k(x) p_k + \frac{1}{2} g^{ij}(x) o_{ij} + F(x), \tag{6.20}
$$

where b is a given smooth vector field on M and F a smooth function on M. One can easily verify that the expression at RHS of (6.20) is indeed invariant under changes of coordinates. We consider the S-H equations (6.17) subject to boundary conditions

Law$(X(0)) = \mu_0$ and $(p, o)(T) = d^2 S_T$, where μ_0 is a given probability distribution and S_T a given smooth function on M.

By the first two equations of system (6.17), the projection diffusion X satisfies the following MDEs,

$$
\begin{cases}
(DX)^i(t) = b^i(X(t)) - \dfrac{1}{2} g^{jk}(X(t))\Gamma^i_{jk}(X(t)), \\
(QX)^{jk}(t) = g^{jk}(X(t)),
\end{cases}
\tag{6.21}
$$

subject to the initial distribution Law$(X(0)) = \mu_0$; or equivalently (according to the end of Sect. 2.4), it can be rewritten as the following Itô SDE in weak sense,

$$
dX^i(t) = \left[b^i(X(t)) - \frac{1}{2} g^{jk}(X(t))\Gamma^i_{jk}(X(t)) \right] dt + \sigma^i_r(X(t))dW^r(t),
$$
$$
\text{Law}(X(0)) = \mu_0,
\tag{6.22}
$$

where σ is the positive-definite square root $(1, 1)$-tensor of g, i.e., $\sum_{r=1}^d \sigma^i_r \sigma^j_r = g^{ij}$, W denotes an $\mathbb{R}^d$-valued standard Brownian motion. Note that the Eq. (6.21) are independent of coordinates (p, o), so they form a closed system on the base manifold M and can be solved independently. Indeed, the solution X is a diffusion on M with generator $A^X = (b^i - \frac{1}{2} g^{jk}\Gamma^i_{jk})\partial_i + \frac{1}{2} g^{jk}\partial_j \partial_k = \nabla_b + \frac{1}{2}\Delta$.

Now we consider the last two equations of (6.17). The LHS of the third equation reads

$$
\left[\frac{\partial}{\partial t} + \left(b^j - \frac{1}{2} g^{kl}\Gamma^j_{kl} \right) \frac{\partial}{\partial x^j} + \frac{1}{2} g^{jk} \frac{\partial^2}{\partial x^j \partial x^k} \right] p_i = \left(\frac{\partial}{\partial t} + \langle b, \nabla \rangle + \frac{1}{2}\Delta \right) p_i,
$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing of vectors and covectors, Δ is the Laplace–Beltrami operator and ∇ the gradient, with respect to g. In order to find the solution of the third equation of (6.17), we consider the following linear backward parabolic equation (where "backward" has a meaning different from that in Sect. 2.2)

$$
\frac{\partial S}{\partial t} + \langle b, \nabla S \rangle + \frac{1}{2}\Delta S + F = 0, \quad t \in [0, T),
\tag{6.23}
$$

with terminal value $S(T, x) = S_T(x)$. We let

$$
p_i = \frac{\partial S}{\partial x^i},
\tag{6.24}
$$

and use (6.23) and (6.15) to derive

$$
\begin{aligned}
-\frac{\partial F}{\partial x^i} &= \frac{\partial}{\partial x^i}\left(\frac{\partial S}{\partial t} + \langle b, \nabla S\rangle + \frac{1}{2}\Delta S\right)\\
&= \left(\frac{\partial}{\partial t} + \langle b, \nabla\rangle + \frac{1}{2}\Delta\right) p_i\\
&\quad + \left(\frac{\partial b^j}{\partial x^i}p_j - \frac{1}{2}\frac{\partial g^{kl}}{\partial x^i}\Gamma^j_{kl}p_j - \frac{1}{2}g^{kl}\frac{\partial \Gamma^j_{kl}}{\partial x^i}p_j + \frac{1}{2}\frac{\partial g^{jk}}{\partial x^i}o_{jk}\right)\\
&= \left(\frac{\partial}{\partial t} + \langle b, \nabla\rangle + \frac{1}{2}\Delta\right) p_i + \frac{\partial}{\partial x^i}(H - F),
\end{aligned}
\tag{6.25}
$$

which agree with the third equation of (6.17).

Finally, we combine (6.24) with (6.15) to conclude that the horizontal integral process **X** is

$$
\mathbf{X}(t) = (p, o)(t, X(t)) = \left(\frac{\partial S}{\partial x^i}, \frac{\partial^2 S}{\partial x^j \partial x^k}\right)(t, X(t)) = d^2 S(t, X(t)).
$$

Example 6.9 (Brownian motions) When $b \equiv 0$ and $F \equiv 0$, the second-order Hamiltonian is $H(x, p, o) = \frac{1}{2}g^{ij}(x)(o_{ij} - \Gamma^k_{ij}(x)p_k)$, the solution process X is a standard Brownian motion on M with initial distribution μ_0. Such second-order Hamiltonian H can be regarded as a "stochastic deformation" of the trivial classical Hamiltonian $H_0 = 0$. Indeed, H is the g-canonical lift of H_0 that will be defined in forthcoming Sect. 6.6. Therefore, we may regard Brownian motions as "stochastization" or "stochastic deformation" of trivially constant curves on the base manifold M.

We are going to describe in the next example a dynamical approach to diffusions, elaborated afterward (Sect. 7.3), inspired by Schrödinger.

6.3.2 Reciprocal Processes and Diffusion Bridges on Riemannian Manifolds

With the same coefficients b, F and boundary data μ_0, S_T in Sect. 6.3.1, we consider the S-H system (6.17) with the following second-order Hamiltonian H on $\mathcal{T}^{S*}M$:

$$
H(x, p, o) = \frac{1}{2}g^{ij}(x)p_i p_j + b^i(x)p_i - \frac{1}{2}g^{ij}(x)\Gamma^k_{ij}(x)p_k + \frac{1}{2}g^{ij}(x)o_{ij} + F(x),
\tag{6.26}
$$

subject to boundary conditions $\mathrm{Law}(X(0)) = \mu_0$ and $(p, o)(T) = d^2 S_T$. Here, b and F are called, respectively, vector and scalar potentials in classical mechanics. Again, it is easy to verify that the expression at RHS of (6.26) is indeed invariant under changes of coordinates.

The LHS of the third equation in (6.17) now reads

$$
\left[\frac{\partial}{\partial t} + \left(g^{jk}p_k + b^j - \frac{1}{2}g^{kl}\Gamma^j_{kl}\right)\frac{\partial}{\partial x^j} + \frac{1}{2}g^{jk}\frac{\partial^2}{\partial x^j \partial x^k}\right]p_i
$$
$$
= \left(\frac{\partial}{\partial t} + p \cdot \nabla + \langle b, \nabla\rangle + \frac{1}{2}\Delta\right)p_i,
$$

In order to find the solution of the third equation of (6.17), we first consider the positive solution of following backward parabolic equation on M

$$
\frac{\partial u}{\partial t} + \langle b, \nabla u\rangle + \frac{1}{2}\Delta u + Fu = 0, \quad t \in [0, T), \tag{6.27}
$$

with terminal value $u(T, x) = e^{S_T(x)}$, where $\langle \cdot, \cdot\rangle$ denotes the Riemannian inner product with respect to g. If we let $S = \ln u$, then it is easy to verify that S satisfies the following Hamilton–Jacobi–Bellman (HJB) equation

$$
\frac{\partial S}{\partial t} + \langle b, \nabla S\rangle + \frac{1}{2}|\nabla S|^2 + \frac{1}{2}\Delta S + F = 0, \quad t \in [0, T), \tag{6.28}
$$

with terminal value $S(T, x) = S_T(x)$, where $|\cdot|$ denotes the Riemannian norm with respect to g. Now we let

$$
p_i = \frac{\partial S}{\partial x^i} = \frac{\partial \ln u}{\partial x^i}, \tag{6.29}
$$

and use (6.28) and (6.15) to derive, in a way similar to (6.25),

$$
-\frac{\partial F}{\partial x^i} = \frac{\partial}{\partial x^i}\left(\frac{\partial S}{\partial t} + \langle b, \nabla S\rangle + \frac{1}{2}|\nabla S|^2 + \frac{1}{2}\Delta S\right)
$$
$$
= \left(\frac{\partial}{\partial t} + p \cdot \nabla + \langle b, \nabla\rangle + \frac{1}{2}\Delta\right)p_i + \frac{\partial}{\partial x^i}(H - F),
$$

which agree with the third equation of (6.17). Therefore, the projection diffusion X of the system (6.17) satisfies the following MDEs,

$$
\begin{cases}
(DX)^i(t) = g^{ij}(X(t))\dfrac{\partial \ln u}{\partial x^j}(t, X(t)) + b^i(X(t)) - \dfrac{1}{2}g^{jk}(X(t))\Gamma^i_{jk}(X(t)), \\
(QX)^{jk}(t) = g^{jk}(X(t)),
\end{cases} \tag{6.30}
$$

subject to the initial distribution $\mathrm{Law}(X(0)) = \mu_0$; or equivalently (according to the end of Sect. 2.4), it can be rewritten as the following Itô SDE in weak sense,

$$
\begin{cases}
dX^i(t) = \left[g^{ij}(X(t)) \dfrac{\partial \ln u}{\partial x^j}(t, X(t)) + b^i(X(t)) - \dfrac{1}{2} g^{jk}(X(t)) \Gamma^i_{jk}(X(t)) \right] dt \\
\qquad\qquad + \sigma^i_r(X(t)) dW^r(t), \\
\mathrm{Law}(X(0)) = \mu_0,
\end{cases}
\tag{6.31}
$$

where σ is the positive-definite square root $(1, 1)$-tensor of g, i.e., $\sum_{r=1}^d \sigma^i_r \sigma^j_r = g^{ij}$, W denotes an $\mathbb{R}^d$-valued standard Brownian motion.

The solution process X of (6.31) is called a *Bernstein process* (Bernstein 1932; Cruzeiro et al. 2000) [or the *reciprocal process* derived from the M-valued diffusion in (6.22) Jamison (1975)]. The time marginal distribution μ_t of X satisfies a Born-type formula $\mu_t(dx) = u(t, x)v(t, x)dx$ (see, e.g., Zambrini 1986, Corollary 3.3.1 or Cruzeiro and Zambrini 1991, Equations (2.9), (4.6) and (4.8)), where v satisfies the adjoint equation of (6.27). The terminal law of X can be determined in the following way: we first solve (6.27) to get $u(0, x)$, and then find out the initial value for v via $\mu_0(dx) = u(0, x)v(0, x)dx$ and solve the equation for v to get $v(T, x)$, finally the terminal law of X is given by $\mu_T(dx) = u(T, x)v(T, x)dx$. In particular, when $\mu_0 = \delta_{q_1}$ and $\mu_T = \delta_{q_2}$ for $q_1, q_2 \in M$, the solution X of (6.31) is the Markovian bridge of the diffusion Y conditioning on ending point q_2 Çetin and Danilova (2016).

Again, we combine (6.29) with (6.15) to conclude that the horizontal integral process **X** is

$$
\mathbf{X}(t) = (p, o)(t, X(t)) = \left(\frac{\partial S}{\partial x^i}, \frac{\partial^2 S}{\partial x^j \partial x^k} \right)(t, X(t)) = d^2 S(t, X(t)). \tag{6.32}
$$

Remark 6.10 (i) The derivation of the reciprocal process (6.31) from the diffusion (6.22) was the way chosen by Jamison (1975), inspired by Schrödinger's original problem (Schrödinger 1932). No geometry or dynamical equations like HJB equation (6.28) was involved by him. Like here, Jamison's construction was involving only the past (nondecreasing) filtration. The dynamical content dates back to Zambrini (1986), Cruzeiro and Zambrini (1991), Chung and Zambrini (2003), where a reciprocal process was constructed from the only data of a Hamiltonian operator as required by Schrödinger's original problem, and the future (nonincreasing) filtration was also used to study the time-reversed dynamics. Cf. also Example 6.12 and Sect. 7.3.

(ii) Equations (6.30) suggest that the transformation from coordinates (x, p, o) to coordinates (x, Dx, Qx) is *not* invertible. More precisely, the coordinates $(D^i x)$ are transformed from (x, p) but the coordinates $(Q^{jk} x)$ are only related to (x^i). Besides, these two equations have nothing to do with the coordinates (o_{jk}). However, if we look at the ∇-canonical coordinates $(D^i_\nabla x)$ for (6.30), then

$$
(D_\nabla X)^i(t) = g^{ij}(X(t)) p_j(t, X(t)) + b^i(X(t)),
$$

which indicates that the transform from (x, p) to $(x, D_\nabla x)$ is invertible. These will help us establish stochastic Lagrangian mechanics and second-order Legendre transforms, in forthcoming Sect. 7.

(iii) As observed in Sect. 2.2, every result presented here has a backward version (in the sense of backward mean derivatives with respect to the future filtration $\{\mathcal{F}_t\}$). Indeed, two forward–backward SDE systems for Bernstein diffusions on Euclidean space were derived in Cruzeiro and Vuillermot (2015): one is under the past filtration and coincides with ours, whereas the other one is under the future filtration.

There are some special cases which are of independent interests and have been considered in the literature.

Example 6.11 (Brownian (free) reciprocal processes and Brownian bridges) Consider the case where $b \equiv 0$, $F \equiv 0$. In this case, Y is a Brownian motion on M, so we call X a Brownian reciprocal process. In particular, the Brownian bridge from q_1 to q_2 of time length $T > 0$ is driven by the Itô SDE (6.31) where $X(0) = q_1$, $b \equiv 0$ and u satisfies the backward heat equation (6.27) with $F \equiv 0$ and final value $u(T, x) = \delta_{q_2}(x)$. See also Hsu (2002, Theorem 5.4.4). Thus, Brownian bridges are understood as stochastic Hamiltonian flows of the second-order Hamiltonian $H(x, p, o) = \frac{1}{2} g^{ij}(x) p_i p_j - \frac{1}{2} g^{ij}(x) \Gamma_{ij}^k(x) p_k + \frac{1}{2} g^{ij}(x) o_{ij}$, compared with geodesics as Hamiltonian flows of the classical Hamiltonian $H_0(x, p) = \frac{1}{2} g^{ij}(x) p_i p_j$ (cf. Abraham and Marsden 1978, Theorem 3.7.1). Here, the second-order Hamiltonian H is the g-canonical lift of H_0. We can also say that Brownian bridges are "stochastization" or "stochastic deformation" of geodesics, cf. Example 6.9. Relations between geodesics and Brownian motions have attracted many studies. For example, one can find various interpolation relations between geodesics and Brownian motions in Angst et al. (2015) and Li (2016).

Example 6.12 (Euclidean quantum mechanics Chung and Zambrini 2003; Albeverio et al. 1989, 2006) It is insightful to consider the case $M = \mathbb{R}^d$ and $b \equiv 0$. The Riemannian metric under consideration is the flat Euclidean one. To catch sight of the analogy with quantum mechanics, we involve the reduced Planck constant $\hbar$ into the second-order Hamiltonian H of (6.26), so that

$$H_\hbar(x, p, o) = \frac{1}{2}|p|^2 + \frac{\hbar}{2}\operatorname{tr} o + F(x).$$

The system (6.10) then reads as

$$\begin{cases} (DX)^i(t) = p_i(t, X(t)), \\ (QX)^{jk}(t) = \hbar \delta^{jk}, \\ D[p_i(t, X(t))] = -\dfrac{\partial F}{\partial x^i}(X(t)), \\ o_{ik}(t, x) = \dfrac{\partial p_k}{\partial x^i}(t, x). \end{cases}$$

Note that the first three equations form a sub-system and can be solved separately, as they are independent of the coordinates o_{ij}'s. Equation (6.27) and its adjoint now reduce to the following $\hbar$-dependent backward and forward heat equations, respectively,

$$\hbar \frac{\partial u}{\partial t} + \frac{\hbar^2}{2} \Delta u + Fu = 0, \qquad -\hbar \frac{\partial v}{\partial t} + \frac{\hbar^2}{2} \Delta v + Fv = 0,$$

which together with the Born-type formula $\mu_t(dx) = u(t, x)v(t, x)dx$ display the strong analogy to quantum mechanics Zambrini (1986).

The function $S = \hbar \ln u$ solves the following $\hbar$-dependent HJB equation:

$$\frac{\partial S}{\partial t} + \frac{1}{2} |\nabla S|^2 + \frac{\hbar}{2} \Delta S + F = 0.$$

The first three equations then can be solved by letting $p = \nabla S$. The first and third equations imply a Newton-type equation

$$DDX(t) = -\nabla F(X(t)). \tag{6.33}$$

This is indeed the equation of motion of the Euclidean version of quantum mechanics, which was the original motivation of Schrödinger in his well-known problem to be discussed in Sect. 7.3. See Chung and Zambrini (2003, p. 158) and Zambrini (2015, Eq. (4.17)) for more. Note that Chung and Zambrini (2003) and Zambrini (2015) used $V = -F$ to denote the physical scalar potential and used the relation $S = -\hbar \ln u$ and $p = -\nabla S$ to formulate the HJB equation from backward heat equation in the case of nondecreasing (past) filtration.

There are two special cases of which more will be studied later.

(i) When $d = 1$ and $F(x) = \frac{1}{2}x^2$, i.e., $H(x, p, o) = \frac{1}{2}(p^2 + x^2) + \frac{1}{2}o$, we call its projective integral process X the (forward) *stochastic harmonic oscillator*. It is a stochastization of the classical harmonic oscillator with Hamiltonian $H_0(x, p) = \frac{1}{2}(p^2 + x^2)$ (Abraham and Marsden 1978, Example 5.2.3). Likewise, here H is the canonical lift H_0, see Sect. 6.6.

(ii) When $d = 1$ and $F(x) = -\frac{1}{2}x^2$, i.e., $H(x, p, o) = \frac{1}{2}(p^2 - x^2) + \frac{1}{2}o$, we call it the (forward) *Euclidean harmonic oscillator*.

6.4 The Mixed-Order Contact Structure on $\mathcal{T}^{S*}M \times \mathbb{R}$

In the later subsections we will investigate time-dependent systems. The proper space for consideration is now $\mathcal{T}^{S*}M \times \mathbb{R}$. Recall in (5.9) that $\mathcal{T}^{S*}M \times \mathbb{R} = J^2\hat{\pi}$, where the latter is the second-order jet bundle of $(M \times \mathbb{R}, \hat{\pi}, M)$.

In classical differential geometry, the first-order jet bundle $J^1\hat{\pi} = T^*M \times \mathbb{R}$ can be equipped with an exact contact structure in several ways (Abraham and Marsden 1978, Section 5.1). Among others, the canonical symplectic form ω_0 on T^*M corresponds to a contact structure on $J^1\hat{\pi}$ via $\tilde{\omega}_0 = \hat{\pi}^*\omega_0$, which is indeed exact as $\tilde{\omega}_0 = -d\tilde{\theta}_0$

for $\tilde{\theta}_0 = dt + \hat{\pi}^* \theta_0$. Another commonly used contact structure is the Poincaré–Cartan form $\omega_{H_0}^0 = \tilde{\omega}_0 + dH_0 \wedge dt$ for a given function $H_0 \in C^\infty(J^1\hat{\pi})$. It is also exact as $\omega_{H_0}^0 = -d\theta_{H_0}^0$ where $\theta_{H_0}^0 = \hat{\pi}^* \theta_0 - H_0 dt$. The advantage of the Poincaré–Cartan form, compared with the contact form ω_0, is that it can be related to the (time-dependent) Hamiltonian vector field V_{H_0} on T^*M of H_0. More precisely, the vector field $\tilde{V}_{H_0} = \frac{\partial}{\partial t} + V_{H_0}$, treated as a vector field on $J^1\hat{\pi}$ and called the characteristic vector field of $\omega_{H_0}^0$, is the unique vector field satisfying $\tilde{V}_{H_0} \lrcorner \omega_{H_0}^0 = 0$ and $\tilde{V}_{H_0} \lrcorner dt = 1$.

Now we proceed in a similar way for the second-order jet bundle $J^2\hat{\pi}$. Define

$$\tilde{\omega} = \hat{\pi}^{S*}\omega \quad \text{and} \quad \tilde{\theta} = dt + \hat{\pi}^{S*}\theta.$$

Then, $\tilde{\omega} = -d\tilde{\theta}$. We call the pair $(J^2\hat{\pi}, \tilde{\omega})$ a second-order contact manifold and the pair $(J^2\hat{\pi}, \tilde{\theta})$ a mixed-order exact contact manifold. In local coordinates, $\tilde{\omega}$ has the same expression as ω in (6.6), but we stress that it is now a second-order form on $\mathcal{T}^{S*}M \times \mathbb{R}$. The form $\tilde{\theta}$ has the local expression

$$\tilde{\theta} = dt + p_i d^2 x^i + \tfrac{1}{2} o_{jk} dx^j \cdot dx^k.$$

This makes clear that $\tilde{\theta}$ is a mixed-order form on $\mathcal{T}^{S*}M \times \mathbb{R}$.

A time-dependent second-order Hamiltonian H is a smooth function on $J^2\hat{\pi} \,\hat{\cong}\, \mathcal{T}^{S*}M \times \mathbb{R}$. The second-order Hamiltonian vector field A_H of H is now a time-dependent second-order vector field on $\mathcal{T}^{S*}M$, its horizontal integral process share the same equations as (6.10) or (6.17), only with H explicitly depending on time. Define a mixed-order vector field $\tilde{A}_H$ on $\mathcal{T}^{S*}M \times \mathbb{R}$ by

$$\tilde{A}_H := A_H + \frac{\partial}{\partial t},$$

where A_H is a second-order Hamiltonian vector field of the form (6.9). We call $\tilde{A}_H$ the extended second-order Hamiltonian vector field of H.

We define the second-order counterpart of Poincaré–Cartan form by

$$\omega_H := \tilde{\omega} + d^\circ H \wedge dt = d^2 x^i \wedge d^2 p_i + \tfrac{1}{2} dx^j \cdot dx^k \wedge d^2 o_{jk} + d^2 H \wedge dt,$$

and call it the mixed-order Poincaré–Cartan form on $\mathcal{T}^{S*}M \times \mathbb{R}$. It is exact in the sense that $\omega_H = -d^\circ \theta_H$, where $\theta_H = \hat{\pi}^{S*}\theta - H dt = p_i d^2 x^i + \tfrac{1}{2} o_{jk} dx^j \cdot dx^k - H dt$.

The following lemma gives the relations between ω_H and $\tilde{A}_H$.

Lemma 6.13 *The class of extended second-order Hamiltonian vector fields $\tilde{A}_H$ is the unique class of mixed-order vector fields on $\mathcal{T}^{S*}M \times \mathbb{R}$ satisfying*

$$\tilde{A}_H \lrcorner \omega_H = 0 \quad \text{and} \quad \tilde{A}_H \lrcorner dt = 1.$$

Proof Firstly, we show that $\tilde{A}_H$ satisfies the two equalities. The second equality is trivial. For the first one, we pick a mixed-order vector field B on $\mathcal{T}^{S*}M \times \mathbb{R}$; then,

$$
\begin{aligned}
\omega_H(\tilde{A}_H, B) &= \tilde{\omega}(\tilde{A}_H, B) + d^\circ H(\tilde{A}_H)dt(B) - dt(\tilde{A}_H)d^\circ H(B) \\
&= \omega\big(A_H, \hat{\pi}_*^S(B)\big) + \big[d^\circ H(A_H) + d^\circ H\big(\tfrac{\partial}{\partial t}\big)\big]dt(B) - d^\circ H(B) \\
&= d^2 H(\hat{\pi}_*^S(B)) + \tfrac{\partial H}{\partial t}dt(B) - d^\circ H(B) \\
&= 0.
\end{aligned}
$$

To prove the uniqueness, it suffices to show that any mixed-order vector field A on $\mathcal{T}^{S*}M \times \mathbb{R}$ satisfying $A \lrcorner \omega_H = 0$ is a multiplier of $\tilde{A}_H$. Suppose that A has the local expression

$$
\begin{aligned}
A = {}& A^0 \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + A_i \frac{\partial}{\partial p_i} + A^{jk} \frac{\partial^2}{\partial x^j \partial x^k} + A^2_{jk} \frac{\partial}{\partial o_{jk}} \\
& + A^{11}_{jk} \frac{\partial^2}{\partial p_j \partial p_k} + A_{ijkl} \frac{\partial^2}{\partial o_{ij} \partial o_{kl}} + A^j_k \frac{\partial^2}{\partial x^j \partial p_k} + A^j_{kl} \frac{\partial^2}{\partial x^j \partial o_{kl}} + A_{jkl} \frac{\partial^2}{\partial p_j \partial o_{kl}}.
\end{aligned}
$$

Then, it follows that

$$
\begin{aligned}
0 = A \lrcorner \omega_H = {}& A^i d^2 p_i - A_i d^2 x^i + A^{jk} d^2 o_{jk} - \tfrac{1}{2} A^2_{jk} dx^j \cdot dx^k \\
& + \text{terms}\left(A^{11}_{jk}, A_{ijkl}, A^j_k, A^j_{kl}, A_{jkl}\right) \\
& - A^0 \left(\frac{\partial H}{\partial x^i} d^2 x^i + \frac{\partial H}{\partial p_i} d^2 p_i + \frac{\partial H}{\partial o_{jk}} d^2 o_{jk} + \frac{1}{2} \frac{\partial^2 H}{\partial x^j \partial x^k} dx^j \cdot dx^k + \cdots\right) \\
& + \left(A^i \frac{\partial H}{\partial x^i} + A_i \frac{\partial H}{\partial p_i} + {+} A^{jk} \frac{\partial^2 H}{\partial x^j \partial x^k} + A^2_{jk} \frac{\partial H}{\partial o_{jk}} + \cdots\right) dt.
\end{aligned}
$$

The vanishing of each coefficient gives

$$
A^i = A^0 \frac{\partial H}{\partial p_i}, \quad A_i = -A^0 \frac{\partial H}{\partial x^i}, \quad A^{jk} = A^0 \frac{\partial H}{\partial o_{jk}}, \quad A^2_{jk} = -A^0 \left(\frac{\partial^2 H}{\partial x^j \partial x^k} + \cdots\right), \quad \cdots.
$$

Therefore, $A = A^0 \tilde{A}_H$. $\qquad\square$

6.5 Canonical Transformations and Hamilton–Jacobi–Bellman Equations

Let us study the second-order analogs of canonical transformations and their generating functions. To do so, we need to find a change of coordinates from (x^i, p_i, o_{jk}, t) to (y^i, P_i, O_{jk}, s) that preserves the form of stochastic Hamilton's equations (6.10) (with time-dependent second-order Hamiltonian). More precisely, we have the following definition of canonical transformations between mixed-order contact structures, which is adapted from those between classical contact structures in Asorey et al. (1983).

Definition 6.14 Let $(\mathcal{T}^{S*}M \times \mathbb{R}, \tilde{\omega})$ and $(\mathcal{T}^{S*}N \times \mathbb{R}, \tilde{\eta})$ be two second-order contact manifolds corresponding to second-order tautological forms θ and ϑ. A bundle isomorphism $\mathbf{F} : (\mathcal{T}^{S*}M \times \mathbb{R}, \hat{\pi}_{2,1}, T^*M \times \mathbb{R}) \to (\mathcal{T}^{S*}N \times \mathbb{R}, \hat{\rho}_{2,1}, T^*N \times \mathbb{R})$ is called a canonical transformation if its projection $\mathbb{F}$ is a bundle isomorphism from $(T^*M \times \mathbb{R}, \hat{\pi}_{0,1}^1, \mathbb{R})$ to $(T^*N \times \mathbb{R}, \hat{\rho}_{0,1}^1, \mathbb{R})$ projecting to $F^0 : \mathbb{R} \to \mathbb{R}$, and there is a function $H_{\mathbf{F}} \in C^\infty(\mathcal{T}^{S*}M \times \mathbb{R})$ such that

$$\mathbf{F}^{R*}\tilde{\eta} = \omega_{H_{\mathbf{F}}}, \tag{6.34}$$

where $\omega_{H_{\mathbf{F}}} = \tilde{\omega} + d^\circ H_{\mathbf{F}} \wedge dF^0$.

The map $\mathbf{F}$ in the definition is also a bundle isomorphism from $(\mathcal{T}^{S*}M \times \mathbb{R}, \hat{\pi}_{0,1}^2, \mathbb{R})$ to $(\mathcal{T}^{S*}N \times \mathbb{R}, \hat{\rho}_{0,1}^2, \mathbb{R})$ projecting to F^0. Hence, we may assume $\mathbf{F}(\alpha_q, t) = (\bar{\mathbf{F}}(\alpha_q, t), F^0(t))$ for all $(\alpha_q, t) \in \mathcal{T}^{S*}M \times \mathbb{R}$, where $\bar{\mathbf{F}}$ is a smooth map from $\mathcal{T}^{S*}M \times \mathbb{R}$ to $\mathcal{T}^{S*}N$. For each $t \in \mathbb{R}$, we define a map $\bar{\mathbf{F}}_t : \mathcal{T}^{S*}M \to \mathcal{T}^{S*}N$ by $\bar{\mathbf{F}}_t(\alpha_q) = \bar{\mathbf{F}}(\alpha_q, t)$. We also introduce an injection $J_t : \mathcal{T}^{S*}M \to \mathcal{T}^{S*}M \times \mathbb{R}$ by $J_t(\alpha_q) = (\alpha_q, t)$. Then, we have $\bar{\mathbf{F}}_t = \hat{\rho}_{1,1} \circ \mathbf{F} \circ J_t$.

Lemma 6.15 *The map $\bar{\mathbf{F}}_t$ is second-order symplectic for each $t \in \mathbb{R}$ if and only if there is a mixed-order form α on $\mathcal{T}^{S*}M \times \mathbb{R}$ such that*

$$\mathbf{F}^{R*}\tilde{\eta} = \tilde{\omega} + \alpha \wedge dt.$$

In particular, condition (6.34) *implies that each $\bar{\mathbf{F}}_t$ is a second-order symplectomorphism.*

Proof The sufficiency follows from

$$(\bar{\mathbf{F}}_t)^{S*}\eta = (J_t)^{R*} \circ \mathbf{F}^{R*} \circ (\hat{\rho}_{1,1})^{S*}\eta = (J_t)^{R*} \circ \mathbf{F}^{R*}\tilde{\eta}$$
$$= (J_t)^{R*}\tilde{\omega} + (J_t)^{R*}\alpha \wedge (J_t)^{R*}dt = \omega + (J_t)^{R*}\alpha \wedge 0 = \omega.$$

For the necessity, we observe that

$$(J_t)^{R*}(\mathbf{F}^{R*}\tilde{\eta} - \tilde{\omega}) = (\bar{\mathbf{F}}_t)^{S*}\eta - \omega = 0.$$

So we can write $\mathbf{F}^{R*}\tilde{\eta} - \tilde{\omega} = \alpha \wedge dt + \gamma$, where γ is a mixed-order form which does not involve dt. This leads to $\gamma = (\hat{\pi}_{1,1})^{R*} \circ (J_t)^{R*}\gamma = (\hat{\pi}_{1,1})^{R*} \circ (J_t)^{R*}(\mathbf{F}^{R*}\tilde{\eta} - \tilde{\omega} - \alpha \wedge dt) = 0$. The result follows. $\square$

The following lemma gives some equivalent statements to the condition (6.34).

Lemma 6.16 *Condition* (6.34) *is equivalent to the following:*

(i) $\mathbf{F}^{R*}\tilde{\vartheta} - \tilde{\theta} + H_{\mathbf{F}}dF^0$ *is mixed-order closed;*
(ii) *for all $K \in C^\infty(\mathcal{T}^{S*}N \times \mathbb{R})$, $\mathbf{F}^{R*}\eta_K = \omega_H$;*
(iii) *for all $K \in C^\infty(\mathcal{T}^{S*}N \times \mathbb{R})$, $\mathbf{F}_*^R \tilde{A}_H = \tilde{A}_K$;*

where $H = (K \circ \mathbf{F} + H_{\mathbf{F}})\dot{F}^0$.

Proof The equivalence between (6.34) and (i) is clear. For (6.34) $\Rightarrow$ (ii), since $\mathbf{F}$ projects to F^0,

$$\mathbf{F}^{R*}\eta_K = \mathbf{F}^{R*}\tilde{\eta} + d^\circ(K \circ \mathbf{F}) \wedge d(t \circ \mathbf{F}) = \tilde{\omega} + d^\circ H_{\mathbf{F}} \wedge dF^0 + d^\circ(K \circ \mathbf{F}) \wedge dF^0$$
$$= \tilde{\omega} + d^\circ H \wedge dt = \omega_H.$$

The converse (ii) $\Rightarrow$ (6.34) is straightforward by letting $K \equiv 0$. To show (ii) $\Rightarrow$ (iii), by applying Lemma 6.13, it suffices to prove that

$$\mathbf{F}^R_* \tilde{A}_H \lrcorner \eta_K = 0 \quad \text{and} \quad \mathbf{F}^R_* \tilde{A}_H \lrcorner dt = 1,$$

while

$$\mathbf{F}^R_* \tilde{A}_H \lrcorner \eta_K = (\mathbf{F}^{R*})^{-1}(\tilde{A}_H \lrcorner \mathbf{F}^{R*}\eta_K) = (\mathbf{F}^{R*})^{-1}(\tilde{A}_H \lrcorner \omega_H) = 0,$$

and

$$\mathbf{F}^R_* \tilde{A}_H \lrcorner dt = (\mathbf{F}^{R*})^{-1}(\tilde{A}_H \lrcorner \mathbf{F}^{R*}dt) = (\mathbf{F}^{R*})^{-1}(\dot{F}^0 \tilde{A}_H \lrcorner dt) = (\mathbf{F}^{R*})^{-1}(\dot{F}^0) = 1.$$

(iii) $\Rightarrow$ (ii) is similar. $\qquad\square$

Definition 6.17 Let $\mathbf{F} : \mathcal{T}^{S*}M \times \mathbb{R} \to \mathcal{T}^{S*}N \times \mathbb{R}$ be canonical. If we can locally write

$$\mathbf{F}^{R*}\tilde{\vartheta} - \tilde{\theta} + H_{\mathbf{F}}dF^0 = -d^\circ G \qquad (6.35)$$

for $G \in C^\infty(M \times \mathbb{R})$, then we call G a generating function for the canonical transformation $\mathbf{F}$.

We use (x, p, o, t) for local coordinates on $\mathcal{T}^{S*}M \times \mathbb{R}$ and (y, P, O, s) for those on $\mathcal{T}^{S*}N \times \mathbb{R}$. Recall that $\mathbf{F}(\alpha_q, t) = (\bar{\mathbf{F}}(\alpha_q, t), F^0(t))$. Then, using (A.4), the relation (6.35) reads in coordinates as

$$\left[\dot{F}^0 + (P_i \circ \mathbf{F})\frac{\partial \bar{\mathbf{F}}^i}{\partial t} \right] dt + (P_i \circ \mathbf{F})\frac{\partial \bar{\mathbf{F}}^i}{\partial x^j}d^2x^j$$
$$+ \frac{1}{2}\left[(P_i \circ \mathbf{F})\frac{\partial^2 \bar{\mathbf{F}}^i}{\partial x^k \partial x^l} + (O_{ij} \circ \mathbf{F})\frac{\partial \bar{\mathbf{F}}^i}{\partial x^k}\frac{d\bar{\mathbf{F}}^j}{dx^l} \right] dx^k \cdot dx^l$$
$$- \left(dt + p_i d^2x^i + \frac{1}{2}o_{jk}dx^j \cdot dx^k \right) + H_{\mathbf{F}}dF^0$$
$$+ \frac{\partial G}{\partial t}dt + \frac{\partial G}{\partial x^i}d^2x^i + \frac{1}{2}\frac{\partial^2 G}{\partial x^j \partial x^k}dx^j \cdot dx^k = 0.$$

Balancing the coefficient of dt, we get

$$\frac{\partial G}{\partial t} + H_{\mathbf{F}} + (P_i \circ \mathbf{F})\frac{\partial \bar{\mathbf{F}}^i}{\partial t} + \dot{F}^0 - 1 = 0.$$

By Lemma 6.16, the new Hamiltonian function K after transformation $\mathbf{F}$ is related to the old Hamiltonian H by $(H - K \circ \mathbf{F})\dot{F}^0 = H_{\mathbf{F}}$. Let us further assume that we can choose coordinates in which (y^i) and (x^i) are independent, so that the independent variables in (6.35) are (x, y, t). Then, relation (6.35) means

$$
\left(P_i d^2 y^i + \tfrac{1}{2} O_{jk} dy^j \cdot dy^k + dF^0 \right) - \left(p_i d^2 x^i + \tfrac{1}{2} o_{jk} dx^j \cdot dx^k + dt \right)
$$
$$
+ (H dt - K dF^0) = -d^\circ G, \tag{6.36}
$$

which implies that the generating function of the canonical transformation $G(x, y, t)$ satisfies

$$
\begin{cases}
p_i = \dfrac{\partial G}{\partial x^i}, \quad o_{jk} \dfrac{\partial x^k}{\partial y^l} = \dfrac{\partial^2 G}{\partial x^j \partial x^k} \dfrac{\partial x^k}{\partial y^l} + \dfrac{\partial^2 G}{\partial x^j \partial y^l}, \\[3mm]
P_i = -\dfrac{\partial G}{\partial y^i}, \; O_{jk} = -\dfrac{\partial^2 G}{\partial y^j \partial y^k} - \dfrac{\partial^2 G}{\partial y^j \partial x^l} \dfrac{\partial x^l}{\partial y^k}, \\[3mm]
(K - 1)\dot{F}^0 - H + 1 = \dfrac{\partial G}{\partial t}.
\end{cases} \tag{6.37}
$$

The expressions for (o_{jk}) and (O_{jk}) are due to the mixed differential term in $d^\circ G$ and correspond to the relation (6.15).

Remark 6.18 Unlike the canonical transformations of classical Hamiltonian systems which have four types of generating functions related via classical Legendre transform (see Goldstein et al. 2002, Section 9.1), here we can only have the type using (x, y, t) as independent variables but not others. This can be attributed to the ill-behaveness of the second-order analog of Legendre transform, as indicated in Remark 6.10.(iii). However, if the configuration space M is a Riemannian manifold, stochastic Hamiltonian mechanics can be simplified to share the same phase space T^*M as classical Hamiltonian mechanics, so that we can also have four types of generating functions. See Sect. 7.4.2 for details and examples of canonical transformations.

The Hamilton–Jacobi–Bellman (HJB) equation can be introduced as a special case of a time-dependent canonical transformation (6.37). In the case where $F^0 = \mathbf{Id}_{\mathbb{R}}$ and the new Hamiltonian K vanishes formally, we denote by S the corresponding generating function G. It follows from (6.37) that S solves the Hamilton–Jacobi–Bellman equation,

$$
\frac{\partial S}{\partial t} + H\left(x^i, \frac{\partial S}{\partial x^i}, \frac{\partial^2 S}{\partial x^j \partial x^k}, t \right) = 0. \tag{6.38}
$$

We will refer to Eq. (6.38) as the HJB equation *associated with* second-order Hamiltonian H, and a solution S of (6.38) as a *second-order Hamilton's principal function* of H.

More generally, we have

Theorem 6.19 *Let A_H be a second-order Hamiltonian vector field on $(\mathcal{T}^{S*}M, \omega)$ and let $S \in C^\infty(M \times \mathbb{R})$. Then, the following statements are equivalent:*

(i) *for every M-valued diffusion X satisfying*

$$(DX(t), QX(t)) = d^2(\tau_M^*)_{d^2 S(t, X(t))} A_H,$$

   the $\mathcal{T}^{S*}M$-valued process $d^2 S \circ X$ is a horizontal integral process of A_H;

(ii) *S satisfies the Hamilton–Jacobi–Bellman equation*

$$\frac{\partial S}{\partial t} + H(d^2 S, t) = f(t), \tag{6.39}$$

   for some function f depending only on t.

Proof Let $\mathbf{X} = d^2 S \circ X$ and set $x^i = x^i \circ d^2 S$, $p_i = p_i \circ d^2 S$, $o_{jk} = o_{jk} \circ d^2 S$. Then,

$$p_i(t, x) = \frac{\partial S}{\partial x^i}(t, x), \quad o_{jk}(t, x) = \frac{\partial^2 S}{\partial x^j \partial x^k}(t, x). \tag{6.40}$$

These imply that the last equation of the system (6.17) holds. Since

$$d^2(\tau_M^*)_{\mathbf{X}(t)} A_H = \frac{\partial H}{\partial p_i}(\mathbf{X}(t)) \frac{\partial}{\partial x^i} + \frac{\partial H}{\partial o_{jk}}(\mathbf{X}(t)) \frac{\partial^2}{\partial x^j \partial x^k},$$

the first two equations in (6.10) or (6.17) hold. Hence, to turn the process $\mathbf{X} = d^2 S \circ X$ into a horizontal integral process of A_H, it is sufficient and necessary to make sure that the third equation in (6.17) holds. Plugging the first equation of (6.40) into the third equation, it reads as

$$\left(\frac{\partial}{\partial t} + \frac{\partial H}{\partial p_j} \frac{\partial}{\partial x^j} + \frac{\partial H}{\partial o_{jk}} \frac{\partial^2}{\partial x^j \partial x^k} \right) \frac{\partial S}{\partial x^i} = -\frac{\partial H}{\partial x^i}.$$

A straightforward reinterpretation yields

$$\frac{\partial}{\partial x^i} \left[\frac{\partial S}{\partial t} + H \left(x^j, \frac{\partial S}{\partial x^j}, \frac{\partial^2 S}{\partial x^j \partial x^k}, t \right) \right] = 0.$$

The result follows.   □

Remark 6.20 If S solves the HJB equation (6.39), then $\tilde{S} = S - \tilde{f}$ solve (6.38) with $\tilde{f}$ a primitive function of f. As a matter of fact, one can always integrate the time-dependent function f into the second-order Hamiltonian function H such that the HJB equation (6.39) has the same form as (6.38). More precisely, if we let $\tilde{H} = H - f$, then Theorem 6.19 also holds with $\tilde{H}$ and zero function in place of H and f, respectively. A similar argument holds for S-H equations (6.10). Indeed, adding a

function f depending only on time to a second-order Hamiltonian does not change its S-H equations.

Example 6.21 The function $S = \ln u$ considered in Sect. 6.3 satisfies the Hamilton–Jacobi–Bellman equation (6.28), which is exactly $\frac{\partial S}{\partial t} + H(d^2 S) = 0$ with the second-order Hamiltonian H given in (6.26). Hence, this theorem yields that the process $d^2 S \circ X$ is a horizontal integral process of A_H, which coincides with (6.32). The Euclidean case for such argument has been discovered in Chung and Zambrini (2003, p. 180) or Zambrini (2015, Eq. (4.20)).

By (6.38) and (6.40), the total mean derivative of a second-order Hamilton's principal function S is given by

$$\begin{aligned}
\mathbf{D}_t S &= \frac{\partial S}{\partial t} + D^i x \frac{\partial S}{\partial x^i} + \frac{1}{2} Q^{jk} x \frac{\partial^2 S}{\partial x^j \partial x^k} \\
&= p_i D^i x + \frac{1}{2} o_{jk} Q^{jk} x - H(x, p, o, t).
\end{aligned} \tag{6.41}$$

where $(p(t, x), o(t, x)) = d^2 S(t, x)$ as in (6.40).

6.6 Second-Order Hamiltonian Functions from Classical Ones

In the presence of a linear connection ∇ on M, we are able to reduce (or produce) second-order Hamiltonian functions to (from) classical ones.

Let be given a second-order Hamiltonian function $H : \mathcal{T}^{S*}M \times \mathbb{R} \to \mathbb{R}$. We make use of the fiber-linear bundle injection $\hat{\iota}_\nabla^* : T^*M \to \mathcal{T}^{S*}M$ in (5.5) to define a classical Hamiltonian by

$$H_0 = H \circ \left(\hat{\iota}_\nabla^* \times \mathbf{Id}_\mathbb{R} \right) : T^*M \times \mathbb{R} \to \mathbb{R}. \tag{6.42}$$

In canonical coordinates, it maps as $H_0(x, p, t) = H(x, p, (\Gamma^i_{jk}(x) p_i), t)$. If we introduce a family of auxiliary variables by

$$\hat{o}_{jk} = \hat{o}_{jk}(x, p) := \Gamma^i_{jk}(x) p_i. \tag{6.43}$$

Then, we can write

$$H_0(x, p, t) = H(x, p, \hat{o}(x, p), t).$$

We say H *reduces* to H_0 under the connection ∇, or H_0 is the ∇-*reduction* of H.

Clearly, the way to lift from a classical Hamiltonian $H_0 : T^*M \times \mathbb{R} \to \mathbb{R}$ to a second-order Hamiltonian function that reduces to H_0 under ∇ is *not* unique. But there is a canonical reduction when we are provided with a symmetric $(2, 0)$-tensor field g (not necessarily Riemannian), given by

$$\overline{H}_0^g(x, p, o, t) := H_0(x, p, t) + \tfrac{1}{2} g^{jk}(x) \left(o_{jk} - \Gamma_{jk}^i(x) p_i \right)$$

$$= H_0(x, p, t) + \tfrac{1}{2} g^{jk}(x) o_{jk}^\nabla. \tag{6.44}$$

Then, H_0 is the ∇-reduction of $\overline{H}_0^g$, and

$$\tfrac{1}{2} o_{jk} g^{jk} - \overline{H}_0^g(x, p, o, t) = \tfrac{1}{2} \hat{o}_{jk} g^{jk} - \overline{H}_0^g(x, p, \hat{o}, t). \tag{6.45}$$

We call $\overline{H}_0^g$ the (g, ∇)-*canonical lift* of H_0. If g is a Riemannian metric and ∇ is the associated Levi–Civita connection, then we simply call $\overline{H}_0^g$ the *g-canonical lift* of H_0. If there is a classical Hamiltonian H_0 such that the second-order Hamiltonian H is the (g, ∇)- (or g-) canonical lift of H_0, we say H is (g, ∇)- (or g-) canonical.

As an example, the second-order Hamiltonian H of (6.26) is g-canonical and reduces to $H_0(x, p) = \tfrac{1}{2} g^{ij}(x) p_i p_j + b^i(x) p_i + F(x)$.

Furthermore, for the canonical transformation $\mathbf{F} : \mathcal{T}^{S*}M \to \mathcal{T}^{S*}N$ in Definition 6.14, we can reduce its associated function $H_{\mathbf{F}} \in C^\infty(\mathcal{T}^{S*}M \times \mathbb{R})$ to a classical function $H_{\mathbf{F}}^0 \in C^\infty(T^*M \times \mathbb{R})$ via (6.42). As a consequence of (6.34), the projection of $\mathbf{F}$, i.e., the map $\mathbb{F} : T^*M \times \mathbb{R} \to T^*N \times \mathbb{R}$ satisfies $\mathbb{F}^* \tilde{\eta}_0 = \omega_{H_{\mathbf{F}}^0}^0$ where $\omega_{H_{\mathbf{F}}^0}^0 = \tilde{\omega}_0 + dH_{\mathbf{F}}^0 \wedge dF^0$. It follows that $\mathbb{F}$ is a classical canonical transformation (Abraham and Marsden 1978, Definition 5.2.6).

We will go back to this issue in Sect. 7.4 where the second-order Legendre transform will be developed. In particular, we will show there that for the canonical second-order Hamiltonian in (6.44), the corresponding second-order Hamilton's equations (6.17) can be rewritten on the cotangent bundle T^*M in a global fashion, see Theorem 7.22.

7 Stochastic Lagrangian Mechanics

In this section, we specify a Riemannian metric g for the manifold M, and a g-compatible linear connection ∇. Note that such g and ∇ always exist but are not unique in general.

We will denote by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ the Riemannian norm and inner product, respectively. Also, denote by $\check{g}$ the inverse metric tensor of g, and (Γ_{jk}^i) the Christoffel symbols of ∇. We observe that $\check{g}$ is a $(2, 0)$-tensor field. Denote by R the Riemann curvature tensor and Ric the Ricci $(1, 1)$-tensor.

7.1 Mean Covariant Derivatives

Definition 7.1 (*Vector fields and 1-forms along diffusions*) Let X be diffusion on M. By a vector field along X, we mean a TM-valued process V, such that $\tau_M(V(t)) = X(t)$ for all t. Similarly, by a 1-form along X, we mean a T^*M-valued process η, such that $\tau_M^*(\eta(t)) = X(t)$ for all t.

Clearly, for a time-dependent vector field V on M, the restriction of V on X, i.e., $\{V_{(t, X(t))}\}$, is a vector field along X. In this case, we call $\{V_{(t, X(t))}\}$ a vector field

restricted on X. In this way, vector fields restricted on X are just TM-valued horizontal diffusions projecting to X. Similarly for 1-forms.

Definition 7.2 (*Parallelisms along diffusions*) Let $X \in I_{t_0}(M)$. A vector field V along X is said to be parallel along X if the following Stratonovich SDE in local coordinates holds,

$$dV^i(t) + \Gamma^i_{jk}(X(t))V^j(t) \circ dX^k(t) = 0. \tag{7.1}$$

A 1-form η along X is said to be parallel along X if

$$d\eta_j(t) - \Gamma^i_{jk}(X(t))\eta_i(t) \circ dX^k(t) = 0.$$

Definition 7.3 (*Stochastic parallel displacements*) Given a diffusion $X \in I_{t_0}(M)$ and a (random) vector $v \in T_{X(t_0)}M$, the stochastic parallel displacement of v along X is the extension of v to a parallel vector field V along X, that is, V satisfies the SDE (7.1) with initial condition $V(t_0) = v$. We denote $\Gamma(X)^t_{t_0} v := V(t)$ and $\Gamma(X)^{t_0}_t V(t) := v$. The stochastic parallel displacement of a (random) covector $\eta \in T^*_{X(t_0)}M$ along X is defined in a similar fashion.

Definition 7.4 (*Damped parallel displacements*) Let $X \in I_{t_0}(M)$. Given a (random) vector $v \in T_{X(t_0)}M$ and covector $\eta_0 \in T^*_{X(t_0)}M$, the damped parallel displacement of v along X is the extension of v to a vector field V along X that satisfies the SDE

$$dV^i(t) + \Gamma^i_{jk}(X(t))V^j(t) \circ dX^k(t) + \frac{1}{2}R^i_{kjl}(X(t))V^j(t)(QX)^{kl}(t)dt = 0,$$
$$V(t_0) = v. \tag{7.2}$$

The damped parallel displacement of η_0 along X is the extension of η to a vector field η along X that satisfies the SDE

$$d\eta_j(t) - \Gamma^i_{jk}(X(t))\eta_i(t) \circ dX^k(t) - \frac{1}{2}R^i_{kjl}(X(t))\eta_i(t)(QX)^{kl}(t)dt = 0,$$
$$\eta(t_0) = \eta_0. \tag{7.3}$$

We denote $\overline{\Gamma}(X)^t_{t_0} v := V(t), \overline{\Gamma}(X)^t_{t_0}\eta_0 := \eta(t)$, and $\overline{\Gamma}(X)^{t_0}_t V(t) := v, \overline{\Gamma}(X)^{t_0}_t \eta(t) := \eta_0$.

If V and η are restrictions on X, that is, $V(t) = V_{(t,X(t))}$ and $\eta(t) = \eta_{(t,X(t))}$, then equations (7.2) and (7.3) can be rewritten, respectively, as

$$\frac{\partial V}{\partial t}dt + \nabla_{\circ dX}V + \frac{1}{2}R(V, \circ dX) \circ dX = 0,$$
$$\frac{\partial \eta}{\partial t}dt + \nabla_{\circ dX}\eta - \frac{1}{2}R(\eta, \circ dX) \circ dX = 0,$$

where we mean by $R(\eta, V)W$ the 1-form $[R(\eta^\sharp, V)W]^\flat$. The Stratonovich stochastic differentials can be transformed into Itô ones. For example, (7.3) is equivalent to

$$
d\eta_j(t) = \Gamma^i_{jk}(X(t))\eta_i(t)dX^k(t) + \frac{1}{2}(QX)^{kl}(t)\left(\frac{\partial \Gamma^i_{jk}}{\partial x^l} + \Gamma^m_{jk}\Gamma^i_{ml}\right)(X(t))\eta_i(t)dt
$$

$$
+ \frac{1}{2}R^i_{kjl}(X(t))\eta_i(t)(QX)^{kl}(t)dt. \tag{7.4}
$$

Remark 7.5 The notion of stochastic parallel displacements was introduced by Itô (1975) and Dynkin (1968). The notion of damped parallel displacement is due to Malliavin (1997). It was originally introduced by Dohrn and Guerra (1979), where they call it geodesic correction to the stochastic parallel displacement.

Lemma 7.6 *Let $X \in I_{t_0}(M)$.*

(i) *Let η be a 1-form on M parallel along X. If V is a vector field on M which is also parallel along X, then $\eta(V)(t) = \eta(V)(t_0)$ for all $t \geq t_0$; if $v \in T_{X(t_0)}M$, then $\eta(\Gamma(X)^t_{t_0}v)(t) = \eta(v)(t_0)$ for all $t \geq t_0$.*
(ii) *Let η be a 1-form on along X satisfying the SDE (7.3). If V is a vector field along X satisfying the SDE (7.2), then $\eta(V)(t) = \eta(V)(t_0)$ for all $t \geq t_0$; if $v \in T_{X(t_0)}M$, then $\eta(\overline{\Gamma}(X)^t_{t_0}v)(t) = \eta(v)(t_0)$ for all $t \geq t_0$.*

Proof We only prove Assertion (ii), as (i) is similar. Since Stratonovich stochastic differentials obey Leibniz's rule, we have

$$
\begin{aligned}
d[\eta(V)] &= \eta_i \circ dV^i + V^j \circ d\eta_j \\
&= -\eta_i \Gamma^i_{jk}V^j \circ dX^k - \frac{1}{2}\eta_i R^i_{kjl}V^j(QX)^{kl}dt + V^j \Gamma^i_{jk}\eta_i \circ dX^k \\
&\quad + \frac{1}{2}V^j R^i_{kjl}\eta_i(QX)^{kl}dt \\
&= 0.
\end{aligned}
$$

This proves the first statement of (ii). The second statement of (ii) follows by letting $V(t) := \overline{\Gamma}(X)^t_{t_0}v$.   □

Definition 7.7 (*Mean covariant derivatives along diffusions*) Given a diffusion X on M. Let V and η be time-dependent vector field and 1-form along X, respectively. The (forward) mean covariant derivative of V with respect to X is a time-dependent vector field $\frac{\mathbf{D}V}{dt}$ along X, defined by

$$
\frac{\mathbf{D}V}{dt}(t) = \lim_{\epsilon \to 0^+} \mathbf{E}\left[\frac{\Gamma(X)^t_{t+\epsilon}V(t+\epsilon) - V(t)}{\epsilon}\bigg| \mathcal{P}_t\right]. \tag{7.5}
$$

The damped mean covariant derivative of V with respect to X is a time-dependent vector field $\frac{\overline{\mathbf{D}}V}{dt}$ along X with $\overline{\Gamma}$ instead of Γ in (7.5). Similarly, we can define $\frac{\mathbf{D}\eta}{dt}$ and $\frac{\overline{\mathbf{D}}\eta}{dt}$.

Lemma 7.8 (i) *Let V and η be vector field and 1-form along X. If η is parallel along X, then*

$$\mathbf{E}\left[\eta\left(\tfrac{\mathbf{D}V}{dt}\right)\right] = \mathbf{E}\left(D[\eta(V)]\right). \tag{7.6}$$

If η satisfies the SDE (7.3), then (7.6) holds true with $\frac{\overline{\mathbf{D}}}{dt}$ instead of $\frac{\mathbf{D}}{dt}$.

(ii) *Let V be a vector field restricted on X. Then,*

$$\frac{\overline{\mathbf{D}}V}{dt} = \frac{\mathbf{D}V}{dt} + \frac{1}{2}(QX)^{ij}R(V,\partial_i)\partial_j = \frac{\partial V}{\partial t} + \nabla_{D_\nabla X}V$$
$$+ \frac{1}{2}(QX)^{ij}\left(\nabla^2_{\partial_i,\partial_j}V + R(V,\partial_i)\partial_j\right).$$

(iii) *Let η be a 1-form restricted on X. Then,*

$$\frac{\overline{\mathbf{D}}\eta}{dt} = \frac{\mathbf{D}\eta}{dt} - \frac{1}{2}(QX)^{ij}R(\eta,\partial_j)\partial_i = \frac{\partial\eta}{\partial t} + \nabla_{D_\nabla X}\eta$$
$$+ \frac{1}{2}(QX)^{ij}\left(\nabla^2_{\partial_i,\partial_j}\eta - R(\eta,\partial_j)\partial_i\right).$$

(iv) *Let V and η be a vector field and a 1-form restricted on X. Then,*

$$\mathbf{D}_t[\eta(V)] = \eta\left(\frac{\mathbf{D}V}{dt}\right) + \frac{\mathbf{D}\eta}{dt}(V) + (QX)^{ij}(\nabla_{\partial_i}\eta)(\nabla_{\partial_j}V)$$
$$= \eta\left(\frac{\overline{\mathbf{D}}V}{dt}\right) + \frac{\overline{\mathbf{D}}\eta}{dt}(V) + (QX)^{ij}(\nabla_{\partial_i}\eta)(\nabla_{\partial_j}V).$$

Proof (i) By Lemma 7.6.(i), we have

$$\mathbf{E}\left[\eta\left(\frac{\mathbf{D}V}{dt}\right)(t)\right] = \lim_{\epsilon\to 0}\mathbf{E}\left[\frac{\eta(t)(\Gamma(X)^t_{t+\epsilon}V(t+\epsilon)) - \eta(t)(V(t))}{\epsilon}\right]$$
$$= \lim_{\epsilon\to 0}\mathbf{E}\left[\frac{\eta(V)(t+\epsilon) - \eta(V)(t)}{\epsilon}\right]$$
$$= \mathbf{E}\left(D[\eta(V)(t)]\right).$$

This proves the first statement of (i). The second statement of (i) follows by a similar argument with $\frac{\overline{\mathbf{D}}}{dt}$ in place of $\frac{\mathbf{D}}{dt}$ and $\overline{\Gamma}$ in place of Γ.

(ii) It suffices to derive the expression for $\frac{\overline{\mathbf{D}}V}{dt}$. Suppose that η is a 1-form satisfying the SDE (7.3) and the diffusion X satisfies $QX(t) = (\sigma \circ \sigma^*)(t, X(t))$. Then, we apply Itô's formula to $\eta(V)(X(t))$ and make use of (2.20) and (7.4). We get

$$
\begin{aligned}
d[\eta(V)] = d(\eta_i V^i) &= \eta_i \left(\frac{\partial V^i}{\partial t} dt + \frac{\partial V^i}{\partial x^j} dX^j + \frac{1}{2} \frac{\partial^2 V^i}{\partial x^j \partial x^k} d[X^j, X^k] \right) \\
&\quad + V^j d\eta_j + d[\eta_j, V^j] \\
&= \eta_i \left(\frac{\partial V^i}{\partial t} + \frac{\partial V^i}{\partial x^j} (DX)^j + \frac{1}{2} \frac{\partial^2 V^i}{\partial x^j \partial x^k} (QX)^{jk} \right) dt + \eta_i \frac{\partial V^i}{\partial x^j} \sigma_r^j dB^r \\
&\quad + V^j \left[\Gamma_{jk}^i (DX)^k + \frac{1}{2} (QX)^{kl} \left(\frac{\partial \Gamma_{jk}^i}{\partial x^l} + \Gamma_{jk}^m \Gamma_{ml}^i \right) + \frac{1}{2} R_{kjl}^i (QX)^{kl} \right] \eta_i dt \\
&\quad + V^j \Gamma_{jk}^i \eta_i \sigma_r^k dB^r + \Gamma_{jk}^i \eta_i \frac{\partial V^j}{\partial x^l} (QX)^{kl} dt \\
&= \eta_i \left[\frac{\partial V^i}{\partial t} + \left(\frac{\partial V^i}{\partial x^k} + V^j \Gamma_{jk}^i \right) (D_\nabla X)^k \right] dt \\
&\quad + \frac{1}{2} \eta_i (QX)^{kl} \left[-\frac{\partial V^i}{\partial x^j} \Gamma_{kl}^j + \frac{\partial^2 V^i}{\partial x^k \partial x^l} + V^j \left(-\Gamma_{jm}^i \Gamma_{kl}^m + \frac{\partial \Gamma_{jk}^i}{\partial x^l} + \Gamma_{jk}^m \Gamma_{ml}^i \right) \right. \\
&\quad\quad \left. + 2\Gamma_{jk}^i \frac{\partial V^j}{\partial x^l} \right] dt \\
&\quad + \frac{1}{2} \eta_i R_{kjl}^i (QX)^{kl} V^j dt + \eta_i \left(\frac{\partial V^i}{\partial x^k} + V^j \Gamma_{jk}^i \right) \sigma_r^k dB^r \\
&= \eta \left(\frac{\partial V}{\partial t} + \nabla_{D_\nabla X} V + \frac{1}{2} (QX)^{ij} \left(\nabla_{\partial_i, \partial_j}^2 V + R(V, \partial_i) \partial_j \right) \right) dt \\
&\quad + \eta \left(\nabla_{\sigma_r} V \right) dB^r .
\end{aligned}
$$

Hence, the result (i) implies

$$
\begin{aligned}
\mathbf{E} \left[\eta \left(\frac{\overline{\mathbf{D}}V}{dt} \right) \right] &= \mathbf{E} \left(D[\eta(V)(t)] \right) \\
&= \mathbf{E} \left[\eta \left(\frac{\partial V}{\partial t} + \nabla_{D_\nabla X} V + \frac{1}{2} (QX)^{ij} \left(\nabla_{\partial_i, \partial_j}^2 V + R(V, \partial_i) \partial_j \right) \right) \right].
\end{aligned}
$$

The arbitrariness of η yields (ii).

(iii) Similar to (ii). $\qquad\blacksquare$

(iv) We only prove the first equality as the second is similar. By (4.6),

$$
\begin{aligned}
\mathbf{D}_t[\eta(V)] &= \left(\frac{\partial}{\partial t} + (D_\nabla X)^i \partial_i + \frac{1}{2}(QX)^{ij} \nabla^2_{\partial_i, \partial_j} \right) [\eta(V)] \\
&= \left(\frac{\partial \eta}{\partial t} \right)(V) + \eta\left(\frac{\partial V}{\partial t} \right) + \left(\nabla_{D_\nabla X}\eta \right)(V) + \eta\left(\nabla_{D_\nabla X} V \right) \\
&\quad + \frac{1}{2}(QX)^{ij} \left[\left(\nabla^2_{\partial_i, \partial_j}\eta \right)(V) + \eta\left(\nabla^2_{\partial_i, \partial_j} V \right) \right. \\
&\quad \left. + \left(\nabla_{\partial_i}\eta \right)\left(\nabla_{\partial_j} V \right) + \left(\nabla_{\partial_j}\eta \right)\left(\nabla_{\partial_i} V \right) \right] \\
&= \eta\left(\frac{\mathbf{D}V}{dt} \right) + \frac{\mathbf{D}\eta}{dt}(V) + (QX)^{ij}(\nabla_{\partial_i}\eta)(\nabla_{\partial_j} V).
\end{aligned}
$$

The result follows.

$\square$

If $QX(t) = \check{g}(X(t))$, then

$$
\frac{\overline{\mathbf{D}}V}{dt} = \frac{\partial V}{\partial t} + \nabla_{D_\nabla X}V + \frac{1}{2}\Delta V + \frac{1}{2}\mathrm{Ric}(V),
$$

and similarly,

$$
\frac{\overline{\mathbf{D}}\eta}{dt} = \frac{\partial \eta}{\partial t} + \nabla_{D_\nabla X}\eta + \frac{1}{2}\Delta\eta - \frac{1}{2}\mathrm{Ric}(\eta) = \frac{\partial \eta}{\partial t} + \nabla_{D_\nabla X}\eta + \frac{1}{2}\Delta_{\mathrm{LD}}\eta, \quad (7.7)
$$

where Δ is the connection Laplacian, and $\Delta_{\mathrm{LD}} = -(dd^* + d^*d)$ is the Laplace–de Rham operator on forms. The relation $\Delta_{\mathrm{LD}} = \Delta - \mathrm{Ric}$ is due to the Weitzenböck identity (Petersen 2016, Theorem 9.4.1). We remark here that the operator $\Delta + \mathrm{Ric}$ acting on vector fields is also called Laplace–de Rham operator in Dohrn and Guerra (1979).

In the context of fluid dynamics, the operator $\frac{\partial}{\partial t} + \nabla_v$, with v a vector field, is often referred to as material derivative or hydrodynamic derivative. So the mean covariant derivative $\frac{\mathbf{D}}{dt}$ and its damped variant $\frac{\overline{\mathbf{D}}}{dt}$ can be regarded as stochastic deformations of material derivative.

7.2 A Stochastic Stationary-Action Principle

In this subsection, we will establish a type of stochastic stationary-action principle: the stochastic Hamilton's principle. Another version for systems with conserved energy, the stochastic Maupertuis's principle, can be found in "Appendix C."

In contrast to second-order Hamiltonians, not all real-valued functions on $T^S M$ can be used as second-order Lagrangians in stochastic Lagrangian mechanics. This has been hinted in Sect. 6.3, as we have mentioned in Remark 6.10. For this reason, we will produce a class of second-order Lagrangians from classical Lagrangians, via

the fiber-linear bundle projection ϱ_∇ in (3.3) and the ∇-canonical coordinates $(D_\nabla^i x)$ in (3.2).

Definition 7.9 By an admissible second-order Lagrangian, we mean a function $L : \mathbb{R} \times T^S M \to \mathbb{R}$ such that there exists a classical Lagrangian $L_0 : \mathbb{R} \times TM \to \mathbb{R}$ satisfying $L = L_0 \circ (\mathbf{Id}_\mathbb{R} \times \varrho_\nabla)$. We call L the ∇-lift of L_0.

In local coordinates, the ∇-lift L of L_0 is expressed as

$$L(t, x, Dx, Qx) = L_0 \circ \varrho_\nabla(t, x, Dx, Qx) = L_0(t, x, D_\nabla x). \tag{7.8}$$

Let $T > 0$. Our stochastic variational problem consists in finding the extrema (maxima or minima) of the stochastic action functional

$$\mathcal{S}[X; 0, T] := \mathbf{E} \int_0^T L\left(t, X(t), DX(t), QX(t)\right) \mathrm{d}t$$
$$= \mathbf{E} \int_0^T L_0\left(t, X(t), D_\nabla X(t)\right) \mathrm{d}t \tag{7.9}$$

over a suitable domain of diffusions X on M, where L is an admissible second-order Lagrangian lifted from L_0.

In order to formulate a well-posed stochastic variational problem in an economical way, we assume that the manifold M is compact and the metric g is geodesically complete (which will be used to characterize the variations of diffusions in Lemma 7.13), and that the connection ∇ is the associated Levi–Civita connection. The geodesic completeness can be ensured, for example, if M is connected (see, e.g., Lee 2013, p. 346). Whenever the metric g is given, the associated Levi–Civita connection is uniquely determined, due to the fundamental theorem of Riemannian geometry (Kobayashi and Nomizu 1963, Theorem IV.2.2). We will refer to such a geodesically complete Riemannian metric as a *reference metric tensor*.

For a fixed point $q \in M$ and a probability distribution $\mu \in \mathcal{P}(M)$ on M, we define an admissible class of diffusions by

$$\mathcal{A}_g([0, T]; q, \mu) = \left\{ X \in I_{(0,q)}^{(T,\mu)}(M) : QX(t) = \check{g}(X(t)), \forall t \in [0, T], \text{a.s.} \right\}, \tag{7.10}$$

where $I_{(0,q)}^{(T,\mu)}(M)$ denotes the set all M-valued diffusion processes starting from q at $t = 0$ and with final distribution μ, i.e., $\mathbf{P} \circ (X(T))^{-1} = \mu$. The action functional $\mathcal{S}$ is now defined on the set $\mathcal{A}_g([0, T]; q, \mu)$, that is, $\mathcal{S} : \mathcal{A}_g([0, T]; q, \mu) \to \mathbb{R}$.

Note that the admissible class $\mathcal{A}_g$ is similar to the Wiener space, so that a candidate for its "tangent space" is Cameron–Martin space. Denote by $\mathcal{H}([0, T]; q)$ the Hilbert space of absolutely continuous curves $v : [0, T] \to T_q M$ such that $\int_0^T |\dot{v}(t)|^2 \mathrm{d}t < \infty$. Let $\mathcal{H}_0([0, T]; q)$ be the subspace consisting of all $v \in \mathcal{H}([0, T]; q)$ satisfying $v(0) = v(T) = 0$.

Definition 7.10 Let $X \in \mathcal{A}_g([0, T]; q, \mu)$. For a curve $v \in \mathcal{H}_0([0, T]; q)$, the vector field along X given by $V(t) := \Gamma(X)_0^t v(t)$ is called a tangent vector to $\mathcal{A}_g([0, T]; q, \mu)$ at X. The tangent space to $\mathcal{A}_g([0, T]; q, \mu)$ at X is the set of all such tangent vectors, that is,

$$T_X \mathcal{A}_g([0, T]; q, \mu) := \left\{ \Gamma(X)_0^{\cdot} v(\cdot) : v \in \mathcal{H}_0([0, T]; q) \right\}.$$

Definition 7.11 By a variation (or deformation) of a diffusion $X \in \mathcal{A}_g([0, T]; q, \mu)$ along $v \in \mathcal{H}_0([0, T]; q)$, we mean a one-parameter family of diffusions $\{X_\epsilon^v\}_{\epsilon \in (-\varepsilon, \varepsilon)}$, where for each $t \in [0, T]$, $X_\epsilon^v(t)$ satisfies the following ODE

$$\frac{\partial}{\partial \epsilon} X_\epsilon^v(t) = \Gamma\left(X_\epsilon^v\right)_0^t v(t), \quad X_0^v(t) = X(t). \tag{7.11}$$

The diffusion $X \in \mathcal{A}_g([0, T]; q, \mu)$ is called a stationary (or critical) point of $\mathcal{S}$, if the first variation $\delta \mathcal{S}$ vanishes at X, i.e.,

$$\left. \frac{\mathrm{d}}{\mathrm{d}\epsilon} \right|_{\epsilon=0} \mathcal{S}\left[X_\epsilon^v; 0, T\right] = 0, \quad \text{for all } v \in \mathcal{H}_0([0, T]; q). \tag{7.12}$$

Remark 7.12 (i) The variations of diffusions on manifolds, via differential equation (7.11), is standard in stochastic analysis on path spaces of Riemannian manifolds. See for example Driver (1992, Eq. (2.3)) and Hsu (1995, Theorem 4.1), where it is shown that Wiener measure is quasi-invariant under such variations. This kind of variations has some equivalent constructions. For instance, the previous two references also provided an approach by lifting to the frame bundle and projecting to the Euclidean space (a stochastic analog of Cartan's development), while Fang and Malliavin (1993) provided an alternative perspective via Bismut connection.

(ii) The stochastic variational problem (7.9)–(7.12) in the Euclidean context has also been familiar in stochastic optimal transport/control. See Sects. 7.3 and 7.4.4 for connections to those areas.

(iii) Unlike the infinitesimal variation used in Definition 4.11 for studying symmetries of SDEs, the infinitesimal variation here in (7.11) needs to be a parallel vector field.

The following lemma is the key for establishing stochastic Hamilton's principle. The first statement shows that the variation X_ϵ^v is well defined on the path space $\mathcal{A}_g([0, T]; q, \mu)$. The second one describes the infinitesimal changes of $D_\nabla X_\epsilon^v$ with respect to the variation parameter ϵ. The proof of the latter is based on a geodesic approximation technique, which is originally due to Itô (1962).

Lemma 7.13 *Given $X \in \mathcal{A}_g([0, T]; q, \mu)$ and $v \in \mathcal{H}_0([0, T]; q)$. We have*

(i) *for each $\epsilon \in (-\varepsilon, \varepsilon)$, $X_\epsilon^v \in \mathcal{A}_g([0, T]; q, \mu)$; and*
(ii) *for all $t \in [0, T]$,*

$$\left. \frac{D}{d\epsilon} \right|_{\epsilon=0} D_\nabla X_\epsilon^v(t) = \Gamma(X)_0^t \dot{v}(t) + \frac{1}{2}(QX)^{ij}(t) R\left(\Gamma(X)_0^t v(t), \partial_i\right) \partial_j, \tag{7.13}$$

where $\dot{v}(t) = \frac{\mathrm{d}}{\mathrm{d}t}v(t) \in T_{v(t)}T_qM \cong T_qM$, $\frac{D}{d\epsilon}$ is the (classical) covariant derivative with respect to the parameter ϵ.

Proof (i) Let ξ and ξ_ϵ be the anti-development (Hsu 2002, Definition 2.3.1) of X and X_ϵ^v, respectively, with fixed initial frame $r(0) \in O_qM$. Equivalently, for example, ξ is an $\mathbb{R}^d$-valued diffusion related to X by the following SDEs (Hsu 2002, Section 2.3)

$$\begin{cases} dX^i(t) = r_j^i(t) \circ d\xi^j(t), \\ dr_j^i(t) = -\Gamma_{kl}^i(X(t))r_j^l(t)r_m^k(t) \circ d\xi^m(t). \end{cases}$$

Applying the fact that $\sum_{k=1}^d r_k^i r_k^j = g^{ij}$ (e.g., Kobayashi and Nomizu 1963, Proposition 1.5) and the condition $QX(t) = \check{g}(X(t))$, we have

$$r_k^i(t)r_l^j(t)\delta^{kl} = g^{ij}(X(t)) = (QX)^{ij}(t) = r_k^i(t)r_l^j(t)(Q\xi)^{kl}(t), \qquad (7.14)$$

and consequently, $Q\xi \equiv \mathbf{I}_d$. Meanwhile, it follows from Fang and Malliavin (1993, Section 3.5) (or Driver 1992, Theorem 5.1, Hsu 1995, Section 3) that

$$d\xi_\epsilon(t) = \exp\left(\epsilon\int_0^t \Omega\left(\left(r(0)^{-1}v\right)(s), \circ d\xi(s)\right)\right) \circ \mathrm{d}\xi(t) + \epsilon d\left(r(0)^{-1}v\right)(t),$$

where Ω is the curvature form on the orthogonal frame bundle OM, taking values in $\mathfrak{so}(d)$, and the frame $r(0)$ is viewed as an isomorphism from $\mathbb{R}^d$ to T_qM. It follows that $Q\xi_\epsilon = Q\xi \equiv \mathbf{I}_d$. For the reason similar to (7.14), we have $QX_\epsilon^v(t) = \check{g}(X_\epsilon^v(t))$. The result follows. See (Driver 1992, Theorem 8.3) for a more elaborate proof.

(ii) Fix $n, m \in \mathbb{N}_+$. Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a division of the time interval $[0, T]$, and let $-\varepsilon = \epsilon_{m-} < \cdots < \epsilon_{-1} < 0 = \epsilon_0 < \epsilon_1 < \cdots < \epsilon_m = \varepsilon$ be one of the variation parameter interval $(-\varepsilon, \varepsilon)$. Denote $\Delta t_i := t_i - t_{i-1}$. Consider the polygonal curve $x^n = \{x^n(t)\}_{t \in [0,T]}$, which is an approximation of X made of minimizing geodesic segments joining $X(t_{i-1})$ with $X(t_i)$ for all $1 \le i \le n$. This is attainable by the geodesic completeness. We will construct an approximation scheme for the variational processes X_ϵ^v's.

For $\epsilon \in [\epsilon_0, \epsilon_1]$, we construct the approximation x_ϵ^n of X_ϵ^v as follows. We extend each $X(t_i)$, $0 \le i \le n$, to a geodesic

$$\gamma_0^{(i)}(\epsilon) = \exp_{X(t_i)}\left(\epsilon\Gamma(x^n)_0^{t_i}v(t_i)\right), \quad \epsilon \in [\epsilon_0, \epsilon_1].$$

Let $x_\epsilon^n = \{x_\epsilon^n(t)\}_{t \in [0,T]}$ be the polygonal curve consisting of minimizing geodesic segments joining $\gamma_0^{(i-1)}(\epsilon)$ with $\gamma_0^{(i)}(\epsilon)$ for all $1 \le i \le n$.

Then, we construct x_ϵ^n for $\epsilon \in [\epsilon_j, \epsilon_{j+1}]$, $1 \le j \le m-1$, by induction. Suppose x_ϵ^n, $\epsilon \in [\epsilon_{j-1}, \epsilon_j]$, has been defined. Then, in particular, we have a curve $x_{\epsilon_j}^n$. Extend each $x_{\epsilon_j}^n(t_i)$, $0 \le i \le n$, to a geodesic by

$$\gamma_j^{(i)}(\epsilon) = \exp_{x_{\epsilon_j}^n(t_i)}\left((\epsilon - \epsilon_j)\Gamma\left(x_{\epsilon_j}^n\right)_0^{t_i}v(t_i)\right), \quad \epsilon \in [\epsilon_j, \epsilon_{j+1}].$$

Let x_ϵ^n be the polygonal curve consisting of minimizing geodesic segments joining $\gamma_j^{(i-1)}(\epsilon)$ with $\gamma_j^{(i)}(\epsilon)$ for all $1 \leq i \leq n$. In a similar way, we can define x_ϵ^n for $\epsilon \in [\epsilon_j, \epsilon_{j+1}]$, $-m \leq j \leq -1$.

Now we have a family of polygonal curves $\{x_\epsilon^n : \epsilon \in (-\varepsilon, \varepsilon)\}$, which satisfies $x_0^n = x^n$ and

$$\frac{\partial^{\operatorname{sign}(\epsilon)}}{\partial \epsilon}\bigg|_{\epsilon=\epsilon_j} x_\epsilon^n(t_i) = \Gamma\left(x_{\epsilon_j}^n\right)_0^{t_i} v(t_i).$$

As for each $\epsilon \in (-\varepsilon, \varepsilon)$ and $1 \leq i \leq n$, $\{x_\epsilon^n(t)\}_{t \in [t_{i-1}, t_i]}$ is a geodesic, the vector field

$$J(t) := \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} x_\epsilon^n(t), \quad t \in [t_{i-1}, t_i]$$

is a Jacobi field along $\{x^n(t)\}_{t \in [t_{i-1}, t_i]}$. This leads to the following Jacobi equation

$$\frac{D^2}{dt^2} J(t) + R\left(J(t), \dot{x}^n(t)\right) \dot{x}^n(t) = 0, \quad t \in [t_{i-1}, t_i], \tag{7.15}$$

with boundary values

$$J(t_{i-1}) = \Gamma(x^n)_0^{t_{i-1}} v(t_{i-1}), \quad J(t_i) = \Gamma(x^n)_0^{t_i} v(t_i). \tag{7.16}$$

Since the connection is torsion-free, we can exchange the covariant derivative and standard derivative to have

$$\frac{D}{dt} J(t_{i-1}) = \frac{D}{dt} \frac{\partial}{\partial \epsilon} x_\epsilon^n(t)\bigg|_{\epsilon=0, t=t_{i-1}} = \frac{D}{d\epsilon} \frac{\partial}{\partial t} x_\epsilon^n(t)\bigg|_{\epsilon=0, t=t_{i-1}}$$

$$= \frac{D}{d\epsilon}\bigg|_{\epsilon=0} \dot{x}_\epsilon^n(t_{i-1}), \tag{7.17}$$

On the other hand, Taylor's theorem yields

$$\Gamma(x^n)_{t_i}^{t_{i-1}} J(t_i) = J(t_{i-1}) + \frac{D}{dt} J(t_{i-1}) \Delta t_i$$

$$+ \frac{1}{2} \frac{D^2}{dt^2} J(t_{i-1}) (\Delta t_i)^2 + o\left((\Delta t_i)^2\right). \tag{7.18}$$

Combining (7.15)–(7.18), we have

$$\frac{D}{d\epsilon}\bigg|_{\epsilon=0} \dot{x}_\epsilon^n(t_{i-1}) = \Gamma(x^n)_0^{t_{i-1}} \frac{v(t_i) - v(t_{i-1})}{\Delta t_i}$$

$$+ \frac{1}{2} R\left(\Gamma(x^n)_0^{t_{i-1}} v(t_{i-1}), \dot{x}^n(t_{i-1})\right) \dot{x}^n(t_{i-1}) \Delta t_i + o\left(\Delta t_i\right).$$

A standard limit theorem yields the result (ii). □

Remark 7.14 (i) The constraint $QX(t) = \check{g}(X(t))$ in (7.10) looks strong. A possibly better viewpoint is to force all diffusions under consideration to have the same nondegenerate diffusion tensor a, i.e., $QX(t) = a(X(t))$. Then, the inverse of a defines a Riemannian metric g, cf. Ikeda and Watanabe (1989, Section V.4). As can be seen from the first part of the above proof, the constraint of fixing the diffusion tensor is a natural one in the literature of variational calculus on the path space. An intuitive reason for this constraint is to assure that the induced measures are equivalent, which is necessary for Eq. (7.11) to be well-posed, cf. Driver (1992). The assumption of Levi–Civita connection ∇ may be relaxed to that the connection ∇ is g-compatible and torsion skew symmetric (Driver 1992, Definition 8.1), in which case the second assertion of this lemma needs to add the effect of torsion.

(ii) One may expect from the limits of (7.15) and (7.16) that there is a "stochastic" Jacobi equation with two boundary values describing the difference between a diffusion and an "infinitesimally close" diffusion, cf. Arnaudon and Thalmaier (1998).

For a smooth function f on TM, we denote by $d_{\dot{x}} f$ the differential of f with respect to the coordinates $(\dot{x}^i)$. Since $T_{(x,\dot{x})} T_x M \cong T_x M$, $d_{\dot{x}} f$ is treated as a 1-form on $T_x M$ and

$$d_{\dot{x}} f = \frac{\partial f}{\partial \dot{x}^i} dx^i. \tag{7.19}$$

We call $d_{\dot{x}} f$ the *vertical differential* of f. Regarding the differential with respect to the coordinates (x^i), we introduce the *horizontal differential* which depends on the connection ∇, by

$$d_x f = \left(\frac{\partial f}{\partial x^i} - \Gamma_{ij}^k \dot{x}^j \frac{\partial f}{\partial \dot{x}^k} \right) dx^i. \tag{7.20}$$

It is easy to check that both definitions (7.19) and (7.20) are invariant under change of coordinates. In fact, by the classical theory (Saunders 1989, Section 3.5 and Example 4.6.7), we know that the connection ∇ can uniquely determine a TTM-valued 1-form on TM horizontal over M, which is given in local coordinates by

$$\Gamma = dx^i \otimes \left(\frac{\partial}{\partial x^i} - \Gamma_{ij}^k \dot{x}^j \frac{\partial}{\partial \dot{x}^k} \right).$$

Hence, the horizontal differential is $d_x f = \Gamma(df)$, where df is the total differential of f. Given a vector field V on M, $f \circ V : q \mapsto f(V_q)$ is a smooth function on V. Then, it is easy to check that

$$d(f \circ V) = d_x f \circ V + (d_{\dot{x}} f \circ V)(\nabla_{\partial_i} V) dx^i. \tag{7.21}$$

The following integration-by-parts formula will be used. Its proof is straightforward from definitions of stochastic integrals and mean derivatives, cf. Cruzeiro and Zambrini (1991, Lemma 4.4).

Lemma 7.15 *Let $X = \{X(t)\}_{t \in [0,T]}$ be a real-valued continuous semimartingale such that DX exists, let f be a real-valued continuous process on $[0, T]$, of finite variation. Then,*

$$\mathbf{E} \int_0^T X(t) \dot{f}(t) dt = E \left[f(T) X(T) - f(0) X(0) \right] - \mathbf{E} \int_0^T f(t) DX(t) dt.$$

Now we are in position to present the stochastic version of Hamilton's principle.

Theorem 7.16 (Stochastic Hamilton's principle) *Let L_0 be a regular Lagrangian on $\mathbb{R} \times TM$. A diffusion $X \in \mathcal{A}_g([0, T]; q, \mu)$ is a stationary point of $\mathcal{S}$, if and only if X satisfies the following stochastic Euler–Lagrange (S-EL) equation*

$$\frac{\overline{\mathbf{D}}}{dt} \left(d_{\dot{x}} L_0 \left(t, X(t), D_\nabla X(t) \right) \right) = d_x L_0 \left(t, X(t), D_\nabla X(t) \right), \qquad (7.22)$$

where $\frac{\overline{\mathbf{D}}}{dt}$ is the damped mean covariant derivative with respect to X.

We remark that since $QX(t) = \check{g}(X(t))$, the operator $\frac{\overline{\mathbf{D}}}{dt}$ in (7.22) is just the one of (7.7). The unknown in (7.22) is the process X, so the conditions $X(0) = q$ and $\mathbf{P} \circ (X(T))^{-1} = \mu$, indicated in the assumption $X \in \mathcal{A}_g([0, T]; q, \mu)$, can be regarded as boundary conditions of (7.22).

Proof Denote $V(t) = \Gamma(X)_0^t v(t)$. It follows from (7.13) and (7.21) that

$$\begin{aligned}
\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \mathcal{S}[X_\epsilon^v; 0, T] &= \mathbf{E} \int_0^T \frac{d}{d\epsilon}\bigg|_{\epsilon=0} L_0 \left(t, X_\epsilon^v(t), D_\nabla X_\epsilon^v(t) \right) dt \\
&= \mathbf{E} \int_0^T \left[d_x L_0 \left(\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} X_\epsilon^v(t) \right) + d_{\dot{x}} L_0 \left(\frac{D}{d\epsilon}\bigg|_{\epsilon=0} D_\nabla X_\epsilon^v(t) \right) \right] dt \\
&= \mathbf{E} \int_0^T \left[d_x L_0 \left(V(t) \right) + d_{\dot{x}} L_0 \left(\Gamma(X)_0^t \dot{v}(t) \right) \right. \\
&\quad \left. + \frac{1}{2} (QX)^{ij}(t) d_{\dot{x}} L_0 \left(R(V(t), \partial_i) \partial_j \right) \right] dt.
\end{aligned}$$
$$(7.23)$$

By Lemmas 7.6.(ii) and 7.15 and the fact that $v(0) = v(T) = 0$, we have

$$\begin{aligned}
\mathbf{E} \int_0^T & d_{\dot{x}} L_0 \left(\Gamma(X)_0^t \dot{v}(t) \right) dt \\
&= \mathbf{E} \int_0^T \Gamma(X)_t^0 (d_{\dot{x}} L_0) \left(\dot{v}(t) \right) dt \\
&= -\mathbf{E} \int_0^T \lim_{\epsilon \to 0} \mathbf{E} \left[\left(\frac{\Gamma(X)_{t+\epsilon}^0 (d_{\dot{x}} L_0) - \Gamma(X)_t^0 (d_{\dot{x}} L_0)}{\epsilon} \right) (v(t)) \,\bigg|\, \mathcal{P}_t \right] dt
\end{aligned}$$

$$= -\mathbf{E} \int_0^T \lim_{\epsilon \to 0} \mathbf{E} \left[\left(\frac{\Gamma(X)_{t+\epsilon}^t (d_{\dot{x}} L_0) - d_{\dot{x}} L_0}{\epsilon} \right) \left(\Gamma(X)_0^t v(t) \right) \Big| \mathcal{P}_t \right] \mathrm{d}t$$

$$= -\mathbf{E} \int_0^T \lim_{\epsilon \to 0} \mathbf{E} \left[\frac{\Gamma(X)_{t+\epsilon}^t (d_{\dot{x}} L_0) - d_{\dot{x}} L_0}{\epsilon} \Big| \mathcal{P}_t \right] \left(\Gamma(X)_0^t v(t) \right) \mathrm{d}t$$

$$= -\mathbf{E} \int_0^T \frac{\mathbf{D}}{dt} (d_{\dot{x}} L_0) \left(V(t) \right) \mathrm{d}t. \tag{7.24}$$

Thus, by Lemma 7.8.(iii),

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \Big|_{\epsilon=0} \mathcal{S}[X_\epsilon^v; 0, T] = \mathbf{E} \int_0^T \left[d_x L_0 \left(V(t) \right) - \frac{\mathbf{D}}{dt} (d_{\dot{x}} L_0) \left(V(t) \right) \right.$$
$$\left. + \frac{1}{2} (QX)^{ij}(t) R(d_{\dot{x}} L_0, \partial_j) \partial_i \left(V(t) \right) \right] \mathrm{d}t$$
$$= \mathbf{E} \int_0^T \left(d_x L_0 - \frac{\overline{\mathbf{D}}}{dt} (d_{\dot{x}} L_0) \right) \left(V(t) \right) \mathrm{d}t.$$

The arbitrariness of v yields the desired result.   □

Remark 7.17 (i) For a special class of Lagrangians in the Euclidean context, the stochastic Euler–Lagrange equation (7.22) has been established in Cruzeiro and Zambrini (1991, Subsection 5.1) where they called it stochastic Newton equation, see also Zambrini (2015). For general Lagrangians on Riemmannian manifolds, Eq. (7.22) is new (to the authors' best knowledge). See Sect. 7.3 for discussions of a special case.

(ii) The second author and his collaborator formulated a weak stochastic Euler–Lagrange equation in Lassalle and Zambrini (2016). They mean by "weak" that their stochastic Euler–Lagrange equation holds in the sense of stochastic integrals. The main differences between their formulation and ours is that we get rid of the stochastic integral (martingale) part in our equation since we use mean derivatives instead of stochastic differentials.

7.3 An Inspirational Example: Schrödinger's Problem

The inspirational example of stochastic Hamiltonian mechanics presented in Sect. 6.3 also provides an example of our stochastic Lagrangian mechanics. Consider the following Lagrangian defined on $\mathbb{R} \times TM$:

$$L_0(t, x, \dot{x}) = \frac{1}{2} |\dot{x} - b(t, x)|^2 - F(t, x), \tag{7.25}$$

where b is a given time-dependent vector field on M. It actually relates to the second-order Hamiltonian H in (6.26) via the second-order Legendre transform, which will be considered in Sect. 7.4. For such Lagrangian, we can directly figure out the relation between stochastic Euler–Lagrange equation (7.22) and Hamilton–Jacobi–Bellman

equation. We denote by $I_0^T(M)$ the set all M-valued diffusion processes over time interval $[0, T]$.

Theorem 7.18 (S-EL & HJB) *Let L_0 be as in* (7.25). *If $X \in I_0^T(M)$ satisfies*

$$D_\nabla X(t) = \nabla S(t, X(t)) + b(t, X(t)) \tag{7.26}$$

for a function $S : \mathbb{R} \times M \to \mathbb{R}$, then X is a solution of the stochastic Euler–Lagrange equation (7.22) *if and only if S solves the following Hamilton–Jacobi–Bellman equation*

$$\frac{\partial S}{\partial t} + \langle b, \nabla S \rangle + \frac{1}{2}|\nabla S|^2 + \frac{1}{2}\Delta S + F = f, \tag{7.27}$$

with f a function depending only on t.

Proof For a function g on $\mathbb{R} \times M$, we will denote by dg the exterior differential of g on M, i.e., with respect to coordinates (x^i). Condition (7.26) can be rewritten in local coordinates as

$$\dot{x} = \nabla S + b. \tag{7.28}$$

Then, it is clear that

$$d_{\dot{x}} L_0 = \frac{\partial L_0}{\partial \dot{x}^i} dx^i = g_{ij}(\dot{x}^j - b^j)dx^i = dS. \tag{7.29}$$

Since $\nabla g = 0$, we use Leibniz's rule to derive

$$
\begin{aligned}
d_x L_0(\partial_k) &= \frac{1}{2}d[g(\dot{x} - b, \dot{x} - b)](\partial_k) - dF(\partial_k) = -g\left(\nabla_{\partial_k} b, \dot{x} - b\right) - dF(\partial_k) \\
&= -dS\left(\nabla_{\partial_k} b\right) - dF(\partial_k).
\end{aligned} \tag{7.30}
$$

Now we take the differential with respect to x to the HJB equation (7.27). Obviously,

$$d\frac{\partial S}{\partial t} = \frac{\partial}{\partial t}dS = \frac{\partial}{\partial t}d_{\dot{x}} L_0.$$

For the second term,

$$
\begin{aligned}
d(\langle b, \nabla S \rangle)(\partial_k) &= d[dS(b)](\partial_k) = \left(\nabla_{\partial_k} dS\right)(b) + dS\left(\nabla_{\partial_k} b\right) \\
&= \nabla^2_{\partial_k, b} S + dS\left(\nabla_{\partial_k} b\right) = (\nabla_b dS)(\partial_k) + dS\left(\nabla_{\partial_k} b\right).
\end{aligned}
$$

For the third term, we use again $\nabla g = 0$. Then, we have

$$
\begin{aligned}
\frac{1}{2}d\left(|\nabla S|^2\right)(\partial_k) &= \frac{1}{2}d[dS \otimes dS(\check{g})](\partial_k) = \left((\nabla_{\partial_k} dS) \otimes dS\right)(\check{g}) \\
&= \left(\nabla_{\partial_k} dS\right)(\nabla S) = (\nabla_{\nabla S} dS)(\partial_k).
\end{aligned}
$$

For the fourth term, in the same way we have

$$
\begin{aligned}
&d(\Delta S)(\partial_k) \\
&= d\left(g^{ij}\nabla^2_{\partial_i,\partial_j}S\right)(\partial_k) = d\left(\nabla^2 S(\check{g})\right)(\partial_k) = \left(\nabla_{\partial_k}\nabla^2 S\right)(\check{g}) = g^{ij}\nabla^3_{\partial_k,\partial_i,\partial_j}S \\
&= g^{ij}\left[\left(\nabla^3_{\partial_k,\partial_i,\partial_j}S - \nabla^3_{\partial_i,\partial_k,\partial_j}S\right) + \left(\nabla^3_{\partial_i,\partial_k,\partial_j}S - \nabla^3_{\partial_i,\partial_j,\partial_k}S\right) + \nabla^3_{\partial_i,\partial_j,\partial_k}S\right] \\
&= g^{ij}\left[\left(\nabla^2_{\partial_k,\partial_i}dS - \nabla^2_{\partial_i,\partial_k}dS\right)(\partial_j) + 0 + \nabla^2_{\partial_i,\partial_j}dS(\partial_k)\right] \\
&= g^{ij}\left[R(\partial_k,\partial_i)dS(\partial_j) + \nabla^2_{\partial_i,\partial_j}dS(\partial_k)\right] \\
&= g^{ij}\left[-R(dS,\partial_j)\partial_i(\partial_k) + \nabla^2_{\partial_i,\partial_j}dS(\partial_k)\right] = [\Delta dS - \mathrm{Ric}(dS)](\partial_k) \\
&= \Delta_{\mathrm{LD}}(dS)(\partial_k).
\end{aligned}
$$

Combining these together and applying (7.26)–(7.30) as well as (7.7), we obtain

$$
\begin{aligned}
&d\left(\frac{\partial S}{\partial t} + \langle b, \nabla S\rangle + \frac{1}{2}|\nabla S|^2 + \frac{1}{2}\Delta S + F\right)(\partial_k) \\
&= \left(\frac{\partial}{\partial t} + \nabla_{b+\nabla S} + \frac{1}{2}\Delta_{\mathrm{LD}}\right)(dS)(\partial_k) + dS\left(\nabla_{\partial_k}b\right) + dF(\partial_k) \\
&= \frac{\overline{\mathbf{D}}}{dt}(dS)(\partial_k) + dS\left(\nabla_{\partial_k}b\right) + dF(\partial_k) = \left[\frac{\overline{\mathbf{D}}}{dt}(d_{\dot{x}}L_0) - d_x L_0\right](\partial_k).
\end{aligned}
$$

The result follows.   □

Remark 7.19 Equation (7.29) gives the relation between Lagrangians and second-order Hamilton's principal functions. It is valid for more general Lagrangians, see Remark 7.23.(i).

Theorem 7.18 strongly suggests some relations between stochastic Lagrangian (and also Hamiltonian) mechanics and Schrödinger's problem in the reinterpretation of optimal transport. In the setting of the latter (see, e.g., Cruzeiro et al. 2000; Léonard 2014; Léonard et al. 2014), there is a given reversible positive measure **R** on the path space $\mathcal{C}_0^T = C([0, T], M)$, called *reference measure*, as well as two probability distributions $\mu_0, \mu_T \in \mathcal{P}(M)$. Schrödinger's problem aims to minimize the following relative entropy:

$$
H(\mathbf{P}|\mathbf{R}) = \begin{cases} \int_{\mathcal{C}_0^T} \log\left(\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{R}}\right)\mathrm{d}\mathbf{P}, & \mathbf{P} \ll \mathbf{R}, \\ +\infty, & \text{otherwise}, \end{cases} \tag{7.31}
$$

over all probability measures **P** on $\mathcal{C}_0^T$ such that μ_0, μ_T are the initial and final time marginal distributions of **P**, i.e., $\mathbf{P}_0 = \mu_0$ and $\mathbf{P}_T = \mu_T$, where $\mathbf{P}_t := \mathbf{P} \circ (X(t))^{-1}$ is the time marginal distribution of **P** and $X(t) : \mathcal{C}_0^T \rightarrow M, X(t,\omega) = \omega(t)$ is the coordinate mapping. Denote, respectively, by $X_{\mathbf{R}}$ and $X_{\mathbf{P}}$, the coordinate process X

under the measure $\mathbf{R}$ and $\mathbf{P}$. Then, Girsanov theorem implies that (Léonard 2012a, Theorem 1) a necessary condition for the *finite entropy condition* $H(\mathbf{P}|\mathbf{R}) < \infty$ is $QX_{\mathbf{P}} = QX_{\mathbf{R}}$, $\mathbf{P}$-a.s.. Furthermore, if $\mathbf{R}$ is a diffusion measure, i.e., $X_{\mathbf{R}}$ is a diffusion process, then a similar application of Girsanov theorem yields that a necessary condition for $H(\mathbf{P}|\mathbf{R}) < \infty$ is that $\mathbf{P}$ is also a diffusion measure and there exists a time-dependent vector field w such that

$$(DX_{\mathbf{P}}(t), QX_{\mathbf{P}}(t)) = (DX_{\mathbf{R}}(t) + w(t, X(t)), QX_{\mathbf{R}}(t)), \quad \forall t \in [0, T], \text{ a.s..}$$

The solution $\mathbf{P}$ of Schrödinger's problem, i.e., minimizing (7.31), is related to the reference measure $\mathbf{R}$ by a time-symmetric version of Doob's h-transform (Léonard 2014, Section 3). Its coordinate process $X_{\mathbf{P}}$ is sometimes called a *Schrödinger bridge* or Schrödinger process. When the reference measure $\mathbf{R}$ is Markovian, i.e., the law of a Markov process, the solution process $X_{\mathbf{P}}$ is also called a reciprocal (Bernstein 1932; Jamison 1975) or Bernstein process (Cruzeiro et al. 2000; Cruzeiro and Vuillermot 2015).

If the manifold M is endowed with a Riemannian metric g, and the reference coordinate process $X_{\mathbf{R}}$ has generator

$$A^{X_{\mathbf{R}}} = \langle b, \nabla \rangle + \tfrac{1}{2}\Delta + F,$$

for some time-dependent vector field b on M, then the density $\mu(t, x) = \frac{\mathrm{d}\,\mathbf{P}_t^*}{\mathrm{d}\,\mathrm{Vol}}(x)$ of the minimizer $\mathbf{P}^*$ of (7.31) solves the following Kolmogorov forward equation

$$\begin{cases} \dfrac{\partial}{\partial t}\mu(t, x) + \mathrm{div}\,[\mu(\nabla S + b)] - \dfrac{1}{2}\Delta\mu(t, x) = 0, & (t, x) \in (0, 1] \times M, \\ \mu(0, x) = \mu_0(x), & x \in M. \end{cases} \tag{7.32}$$

where S solves the HJB equation (7.27) with $f \equiv 0$, or (6.28).

Moreover, an analog of Benamou–Brenier formula was derived (see Léonard 2014). Consider the problem of minimizing the average action

$$\int_0^T \int_M \left(\frac{1}{2}|v(t, x) - b(t, x)|^2 - F(t, x) \right) \rho(t, \mathrm{d}x)\mathrm{d}t \tag{7.33}$$

among all pairs (ρ, v), where is $\rho = (\rho(t))_{t \in [0, T]}$ is a measurable path in $\mathcal{P}(M)$, $v = (v(t))_{t \in [0, T]}$ is a measurable time-dependent vector field and the following constraints are satisfied (in the weak sense of PDEs):

$$\begin{cases} \dfrac{\partial}{\partial t}\rho + \mathrm{div}\,(\rho v) - \dfrac{1}{2}\Delta\rho = 0, \\ \rho(0) = \mu_0, \ \rho(T) = \mu_T, \end{cases} \tag{7.34}$$

The relation between ρ in (7.33) and $\mathbf{P}$ in (7.31) is just that ρ is the time marginal of $\mathbf{P}$, namely,

$$\rho(t) = \mathbf{P}_t = \mathbf{P} \circ (X(t))^{-1}. \tag{7.35}$$

The minimizer of (7.33) is the pair $(\mu, \nabla S + b)$ where μ solves (7.32) and S solves (6.28).

These results are summarized in the following equivalent relations:

$$
\begin{aligned}
\inf & \left\{ H(\mathbf{P}|\mathbf{R}) : \mathbf{P} \in \mathcal{P}(\mathcal{C}_0^T), \mathbf{P}_0 = \mu_0, \mathbf{P}_T = \mu_T \right\} - H(\mu_0|\mathbf{R}_0) \\
&= \inf \left\{ \int_0^T \int_M \left(\frac{1}{2}|v(t,x) - b(t,x)|^2 - F(t,x) \right) \rho(t, \mathrm{d}x)\mathrm{d}t : (\rho, v) \text{ satisfies (7.34)} \right\} \\
&= \int_0^T \int_M \left(\frac{1}{2}|\nabla S(t,x)|^2 - F(t,x) \right) \mu(t, \mathrm{d}x)\mathrm{d}t.
\end{aligned}
\tag{7.36}
$$

Now if the coordinate process $X_{\mathbf{R}}$ under the reference measure $\mathbf{R}$ is a nondegenerate M-valued diffusion in $I_0^T(M)$ which is diffusion-homogeneous, then assigning such a reference measure $\mathbf{R}$ amounts to assigning a pair $(b_{\mathbf{R}}, g_{\mathbf{R}}) \in \Gamma(TM \otimes \mathrm{Sym}^2(T^*M))$, where $g_{\mathbf{R}}$ is a positive-definite symmetric $(0,2)$-tensor, i.e., a Riemannian metric tensor. More precisely, we let $A^{X_{\mathbf{R}}} = (\mathfrak{b}, a) + F$ be the generator of $X_{\mathbf{R}}$. Since $X_{\mathbf{R}}$ is nondegenerate and diffusion-homogeneous, a is a time-independent nondegenerate symmetric $(2,0)$-tensor field. Let $g_{\mathbf{R}} = \hat{a}$ be the inverse of a, so that $g_{\mathbf{R}}$ is a Riemannian metric tensor. We then equip the Riemannian manifold $(M, g_{\mathbf{R}})$ with the associated Levi–Civita connection ∇. The isomorphism (2.19) implies that

$$A^{X_{\mathbf{R}}} = b_{\mathbf{R}}^i \partial_i + \tfrac{1}{2} g_{\mathbf{R}}^{ij} \nabla^2_{\partial_i, \partial_j} + F = \langle b_{\mathbf{R}}, \nabla \rangle + \tfrac{1}{2}\Delta + F,$$

where $b_{\mathbf{R}}$ is the time-dependent vector field given by $b_{\mathbf{R}}^i = (\mathfrak{b}^i + \tfrac{1}{2} g_{\mathbf{R}}^{jk} \Gamma^i_{jk})$, and ∇ and Δ are the gradient and Laplace–Beltrami operator with respect to $g_{\mathbf{R}}$, respectively.

We set that $\mathbf{P}$ is a diffusion measure and $QX_{\mathbf{P}} = QX_{\mathbf{R}} = \check{g}_{\mathbf{R}}$, $\mathbf{P}$-a.s., which is a necessary condition for $H(\mathbf{P}|\mathbf{R}) < \infty$. Then, by (3.4), the generator of $X_{\mathbf{P}}$ is given by

$$(DX_{\mathbf{P}}(t), QX_{\mathbf{P}}(t)) = (D_\nabla X_{\mathbf{P}})^i(t)\partial_i|_{X(t)} + \tfrac{1}{2}\Delta|_{X(t)}.$$

From (7.34) and (7.35), one can see that $v(t, X(t)) = D_\nabla X_{\mathbf{P}}(t)$ and the action (7.33) equals to

$$\mathbf{E}_{\mathbf{P}} \int_0^T \left(\frac{1}{2}|D_\nabla X(t) - b_{\mathbf{R}}(t, X(t))|^2 - F(t, X(t)) \right) \mathrm{d}t. \tag{7.37}$$

So the minimizing problem turns into minimizing the action (7.37) over all diffusion measures $\mathbf{P} \in \mathcal{P}(\mathcal{C}_0^T)$ with $\mathbf{P}_0 = \mu_0$, $\mathbf{P}_T = \mu_T$ and $QX_{\mathbf{P}} = \check{g}_{\mathbf{R}}$, $\mathbf{P}$-a.s.. If $\mu_0 = \delta_q$

and $\mu_T = \mu$, this brings us back to our stochastic variational problem, that is, to minimize the action functional $\mathcal{S}$ in (7.9) over $\mathcal{A}_{g_{\mathbf{R}}}([0, T]; q, \mu)$, with Lagrangian $L_0(t, x, \dot{x}) = \frac{1}{2}|\dot{x} - b_{\mathbf{R}}(t, x)|^2 - F(t, x)$. Note that in this case, since $\mathbf{P}_0 = \mu_0$ is Dirac, the relative entropy in (7.31) and $H(\mu_0|\mathbf{R}_0)$ are always infinite, while their difference $H(\mathbf{P}|\mathbf{R}) - H(\mu_0|\mathbf{R}_0)$ can be finite as in (7.36). Moreover, by Theorem 7.16 and 7.18, a necessary condition for $X_{\mathbf{P}}$ to be the minimizer of $\mathcal{S}$ is that $X_{\mathbf{P}}$ satisfies (7.26) and (7.27), which coincides with (7.32).

Remark 7.20 (i) Compared to the Lagrangian (7.25) used here for addressing Schrödinger's problem, there is another type of Lagrangians used in the Euclidean version of quantum mechanics in Cruzeiro and Zambrini (1991, Eq. (5.4)). The latter has an additional term of divergence of b, which helps to express part of the action functional as a Stratonovich integral. The stochastic Euler–Lagrange equation (7.22) applied to their Lagrangians recovers the equations of motion in Cruzeiro and Zambrini (1991, Theorem 5.3).

(ii) In the seminal paper (Otto 2001), F. Otto provided a geometric perspective for numerous PDEs by introducing a Riemannian structure in the Wasserstein space. It is known as Otto's calculus. A similar idea can ascend to V.I. Arnold, who established a geometric framework for hydrodynamics by studying the Riemannian nature of the infinite-dimensional group of diffeomorphisms (Arnold and Khesin 2021). The recent paper (Gentil et al. 2020) formulated Schrödinger's problem via Otto calculus, where the equation of motion is given by an infinite-dimensional Newton equation, cf. Khesin et al. (2021) and von Renesse (2012) on related matters. All these works can be called a "geometrization" of (stochastic) dynamics. In contrast, the present framework can be called a "stochastization" of geometric mechanics. The difference and relations between our framework and theirs are similar to those between two ways of producing HJ equations for quantum mechanics mentioned in the introduction. More precisely, while (second-order) HJB equations play a key role in our framework, various HJ equations with density-dependent potential terms were derived by them (see Gentil et al. 2020, Corollary 23; Khesin et al. 2021, Proposition 2.4).

7.4 Second-Order Legendre Transform

7.4.1 From $\mathcal{T}^{S*}M$ to $\mathcal{T}^{S}M$ and Back

Let us fix a linear connection ∇ on M. Here, for simplicity, we consider time-independent Hamiltonians and Lagrangians.

We first produce second-order Lagrangians from second-order Hamiltonians. To this end, we first reduce the second-order Hamiltonian to a classical one. Given a time-independent second-order Hamiltonian $H : \mathcal{T}^{S*}M \to \mathbb{R}$, its ∇-reduction is the classical Hamiltonian $H_0 = H \circ \hat{\iota}_{\nabla}^* : T^*M \to \mathbb{R}$, as in (6.42). If H_0 is hyperregular (see Abraham and Marsden 1978, Section 3.6), then its fiber derivative $\mathbf{F}H_0 : T^*M \to TM$, which is given in canonical coordinates by $\dot{x}^i = \frac{\partial H_0}{\partial p_i}$, is a diffeomorphism and defines the classical Legendre transform (Abraham and Marsden 1978, Section 3.6):

$$L_0(x, \dot{x}) = p_i \dot{x}^i - H_0(x, p) = p_i \dot{x}^i - H\left(x, p, \hat{o}\right), \tag{7.38}$$

where $(\hat{o}_{jk})$ is a family of auxiliary variables introduced in (6.43). Then, we lift L_0 to an admissible second-order Lagrangian $L : \mathcal{T}^S M \to \mathbb{R}$ as in Definition 7.9, that is, $L = L_0 \circ \varrho_\nabla$. Combining (7.38) with (7.8), the relation between L and H is

$$\begin{aligned}
L(x, Dx, Qx) &= p_i D_\nabla x^i - H\left(x, p, \hat{o}\right) \\
&= p_i D^i x + \tfrac{1}{2} \hat{o}_{jk} Q^{jk} x - H(x, p, \hat{o}).
\end{aligned} \tag{7.39}$$

We call (7.39) the *second-order Legendre transform*. In particular, if we restrict the admissible second-order Lagrangian L to the subbundle of $\mathcal{T}^S M$ with coordinate constraint $Q^{jk} x = g^{jk}(x)$ for some symmetric $(2, 0)$-tensor field g [which is just the condition in (7.10)], and let H be (g, ∇)-canonical, then by (6.45), we have

$$L(x, Dx, Qx) = p_i D^i x + \tfrac{1}{2} o_{jk} Q^{jk} x - H(x, p, o). \tag{7.40}$$

Consequently, we can find the relation between second-order Hamilton's principal functions and action functionals. By (6.41) and (7.40),

$$\mathbf{D}_t S = L(t, x, Dx, Qx) = L_0(t, x, D_\nabla x).$$

One concludes, from Dynkin's formula, that for an M-valued diffusion $X \in \mathcal{A}_g([0, T]; q, \mu)$,

$$\mathbf{E} S(T, X(T)) - S(0, q) = \mathbf{E} \int_0^T L_0\left(t, X(t), D_\nabla X(t)\right) \mathrm{d}t = \mathcal{S}[X; 0, T],$$

and

$$\begin{aligned}
S(t, x) - S(0, q) &= \mathbf{E}_{(t,x)}[S(t, X(t)) - S(0, X(0))] \\
&= \mathbf{E}_{(t,x)} \int_0^t L_0\left(s, X(s), D_\nabla X(s)\right) \mathrm{d}s,
\end{aligned}$$

where $\mathbf{E}_{(t,x)}$ is the conditional expectation $\mathbf{E}(\cdot | X(t) = x)$. These mean that the action functional is the expectation of second-order Hamilton's principal function (up to an undetermined constant), while the second-order Hamilton's principal function is the conditional expectation version of action functional.

Conversely, let us be given an admissible second-order Lagrangian $L : \mathcal{T}^S M \to \mathbb{R}$ which is the ∇-lift of a classical Lagrangian $L_0 : TM \to \mathbb{R}$. If L_0 is hyperregular, then its fiber derivative

$$\mathbf{F}L_0 : TM \to T^*M, \quad (x, \dot{x}) \mapsto (x, d_{\dot{x}} L_0), \tag{7.41}$$

which is written in coordinates as $p_i = \frac{\partial L_0}{\partial \dot{x}^i}$, is a diffeomorphism and defines the classical inverse Legendre transform:

$$H_0(x, p) = p_i \dot{x}^i - L_0(x, \dot{x}). \qquad (7.42)$$

We replace coordinates $(\dot{x}^i)$ by $(D^i_\nabla x)$, due to (3.2). Now, given a symmetric $(2, 0)$-tensor field g, we lift H_0 to the (g, ∇)-canonical $\overline{H}^g_0$ in (6.44). The relation between $\overline{H}^g_0$ and L is

$$
\begin{aligned}
\overline{H}^g_0(x, p, o) &= p_i D^i_\nabla x - L_0(x, D_\nabla x) + \tfrac{1}{2} g^{jk}(x) \left(o_{jk} - \Gamma^i_{jk}(x) p_i \right) \\
&= p_i D^i x + \tfrac{1}{2} o_{jk} Q^{jk} x - L(x, Dx, Qx) \\
&\quad + \tfrac{1}{2} \left(g^{jk}(x) - Q^{jk} x \right) o^\nabla_{jk},
\end{aligned}
\qquad (7.43)
$$

where (o^∇_{jk}) is the tensorial conjugate diffusivities defined in (5.6). We call (7.43) the (g, ∇)-*canonical inverse second-order Legendre transform*. When g is Riemannian and ∇ is the associated Levi–Civita connection, we call (7.43) the g-canonical inverse second-order Legendre transform. In particular, when restricting L onto the subbundle of $\mathcal{T}^S M$ with coordinate constraint $Q^{jk} x = g^{jk}(x)$, we have

$$\overline{H}^g_0(x, p, o) = p_i D^i x + \tfrac{1}{2} o_{jk} Q^{jk} x - L(x, Dx, Qx). \qquad (7.44)$$

Following the procedure in classical mechanics (Abraham and Marsden 1978, Definition 3.5.11), for a given classical Lagrangian $L_0 : TM \to \mathbb{R}$, we define a function $A_0 : TM \to \mathbb{R}$ by $A_0(v_x) = \mathbf{F}L_0(v_x) \cdot v_x$, and the *classical energy* $E_0 : TM \to \mathbb{R}$ by $E_0 = A_0 - L_0$. Notice that in local coordinates, $A_0 = \dot{x}^i \frac{\partial L_0}{\partial \dot{x}^i}$ and $E_0 = \dot{x}^i \frac{\partial L_0}{\partial \dot{x}^i} - L_0$.

Example 7.21 It is easy to check that the ∇-lift of the classical Lagrangian L_0 in (7.25) is the second-order Legendre transform of the second-order Hamiltonian H in (6.26). And conversely, the latter is the g-canonical inverse second-order Legendre transform of the former. The classical energy associated with this Lagrangian is given by

$$E_0(t, x, \dot{x}) = \frac{1}{2} |\dot{x} - b(t, x)|^2 + \langle \dot{x} - b(t, x), b(t, x) \rangle + F(t, x). \qquad (7.45)$$

Each term at RHS corresponds to a kinetic energy, a vector potential energy and a scalar potential energy, respectively.

7.4.2 Stochastic Hamiltonian Mechanics on Riemannian Manifolds

Given a reference metric tensor g, i.e., a geodesically complete Riemannian metric as in Sect. 7.2, let ∇ be the associated Levi–Civita connection. If a second-order Hamiltonian H is the g-canonical lift of a classical Hamiltonian H_0, namely, $H = \overline{H}^g_0$ as in (6.44), then the stochastic Hamilton's equations (6.17) can reduce to a simpler Hamilton-type system on $T^* M$, which is exactly equivalent to the stochastic Euler–Lagrange equation (7.22) via the classical Legendre transform (7.41) and (7.42).

Similarly to (7.19) and (7.20), we introduce, for a smooth function f on T^*M, the vertical gradient $\nabla_p f$ and horizontal differential $d_x f$ which are given in local coordinates (x, p) by

$$\nabla_p f = \frac{\partial f}{\partial p_i} \frac{\partial}{\partial x^i}, \quad d_x f = \left(\frac{\partial f}{\partial x^i} + \Gamma_{ij}^k p_k \frac{\partial f}{\partial p_j} \right) dx^i.$$

Both are invariant under change of coordinates. Still by the classical theory, the connection ∇ can uniquely determine a TT^*M-valued 1-form on T^*M horizontal over M, given by

$$\Gamma^* = dx^i \otimes \left(\frac{\partial}{\partial x^i} + \Gamma_{ij}^k p_k \frac{\partial}{\partial p_j} \right).$$

Hence, we have $d_x f = \Gamma^*(df)$. Given a 1-form η on M, $f \circ \eta : q \mapsto f(\eta_q)$ is a smooth function on M. Then, it is easy to verify that

$$d(f \circ \eta) = d_x f \circ \eta + \nabla_{(\nabla_p f \circ \eta)} \eta. \tag{7.46}$$

Theorem 7.22 *Given a smooth function $H_0 : T^*M \times \mathbb{R} \to \mathbb{R}$.*

(i) *Let $H = \overline{H}_0^g : \mathcal{T}^{S*}M \times \mathbb{R} \to \mathbb{R}$ be the g-canonical lift of H. Let $\mathbf{X}$ be the horizontal integral process of stochastic Hamilton's equations (6.17) corresponding to H and $X = \tau_M^{S*}(\mathbf{X})$. Define a T^*M-valued horizontal diffusion by $\mathbb{X} := \hat{\varrho}^*(\mathbf{X})$. Then, $\mathbb{X}(t) = p(t, X(t))$ solves the following system on T^*M,*

$$\begin{cases} D_\nabla X(t) = \nabla_p H_0(\mathbb{X}(t), t), \\ \dfrac{\overline{\mathbf{D}}}{dt} p(t, X(t)) = -d_x H_0(\mathbb{X}(t), t), \end{cases} \tag{7.47}$$

subject to $QX(t) = \check{g}(X(t))$, where $\frac{\overline{\mathbf{D}}}{dt}$ is the damped mean covariant derivative with respect to X. In this case, we refer to the system (7.47) as the g-canonical reduction of (6.17), or global stochastic Hamilton's equations.

(ii) *If H_0 is hyperregular, then the global stochastic Hamilton's equations (7.47) are equivalent to the stochastic Euler–Lagrange equation (7.22) via the classical Legendre transform $p = d_{\dot{x}} L_0$ and $H_0(x, p, t) = p \cdot \dot{x} - L_0(t, x, \dot{x})$.*

(iii) *Let $S \in C^\infty(M \times \mathbb{R})$. Then, the following statements are equivalent:*

(a) *for every M-valued diffusion X satisfying*

$$D_\nabla X(t) = \nabla_p H_0(dS(t, X(t)), t), \quad QX(t) = \check{g}(X(t)), \tag{7.48}$$

the T^*M-valued process $dS \circ X$ solves the global stochastic Hamilton's equations (7.47);

(b) *S satisfies the following Hamilton–Jacobi–Bellman equation*

$$\frac{\partial S}{\partial t} + H_0(dS, t) + \frac{1}{2}\Delta S = f(t), \tag{7.49}$$

for some function f depending only on t.

Proof (i) Since $H = \overline{H}_0^g = H_0 + \frac{1}{2}g^{jk}(o_{jk} - \Gamma^i_{jk}p_i)$, $(QX)^{jk} = 2\frac{\partial H}{\partial o_{jk}}$ if and only if $QX(t) = \check{g}(X(t))$. Since,

$$\frac{\partial H}{\partial p_i} = \frac{\partial H_0}{\partial p_i} - \frac{1}{2}g^{jk}\Gamma^i_{jk} = dx^i(\nabla_p H_0) - \frac{1}{2}(QX)^{jk}\Gamma^i_{jk},$$

we have $(DX)^i = \frac{\partial H}{\partial p_i}$ if and only if $D_\nabla X = \nabla_p H_0$, due to (2.20). This proves the first equation of (7.47). Furthermore,

$$\begin{aligned}
\frac{\partial H}{\partial x^i} &= \frac{\partial H_0}{\partial x^i} + \frac{1}{2}\partial_i g^{jk}\left(o_{jk} - \Gamma^l_{jk}p_l\right) - \frac{1}{2}g^{jk}\partial_i\Gamma^l_{jk}p_l \\
&= \frac{\partial H_0}{\partial x^i} - g^{jm}\Gamma^k_{im}\left(o_{jk} - \Gamma^l_{jk}p_l\right) - \frac{1}{2}g^{jk}\partial_i\Gamma^l_{jk}p_l.
\end{aligned}$$

On the other hand, by applying Lemma 7.8 (ii) and (iv), and the equation $D_\nabla X = \nabla_p H_0$, we have

$$\begin{aligned}
(D(p \circ \mathbf{X}))_i = \mathbf{D}_t p_i = \mathbf{D}_t[p(\partial_i)] &= \frac{\overline{\mathbf{D}}p}{dt}(\partial_i) + p\left(\frac{\overline{\mathbf{D}}\partial_i}{dt}\right) \\
&\quad + (QX)^{jk}(\nabla_{\partial_j}p)(\nabla_{\partial_k}\partial_i) \\
&= \frac{\overline{\mathbf{D}}p}{dt}(\partial_i) + p\left(\nabla_{D_\nabla X}\partial_i + \frac{1}{2}g^{jk}\nabla^2_{\partial_j,\partial_k}\partial_i + \frac{1}{2}g^{jk}R(\partial_i, \partial_j)\partial_k\right) \\
&\quad + g^{jk}(\nabla_{\partial_j}p)(\nabla_{\partial_k}\partial_i) \\
&= \frac{\overline{\mathbf{D}}p}{dt}(\partial_i) + p_l\left(\frac{\partial H_0}{\partial p_j}\Gamma^l_{ij} + \frac{1}{2}g^{jk}\partial_i\Gamma^l_{jk}\right) \\
&\quad + g^{jk}\Gamma^m_{ik}\left(\partial_j p_m - \Gamma^l_{jm}p_l\right).
\end{aligned} \tag{7.50}$$

Hence,

$$(D(p \circ \mathbf{X}))_i + \frac{\partial H}{\partial x^i} = \frac{\overline{\mathbf{D}}p}{dt}(\partial_i) + d_x H_0(\partial_i) + g^{jm}\Gamma^k_{im}\left(\partial_j p_k - o_{jk}\right).$$

The second equation of (7.47) follows from (6.15).

(ii) The equivalence between (7.47) and (7.22) follows from the following calculations:

$$\nabla_p H_0 = \nabla_p (p \cdot \dot{x} - L_0) = \dot{x},$$

$$d_x H_0 = \left(\frac{\partial H_0}{\partial x^i} + \Gamma_{ij}^k p_k \frac{\partial H_0}{\partial p_j} \right) dx^i$$

$$= \left(-\frac{\partial L_0}{\partial x^i} + \Gamma_{ij}^k \frac{\partial L_0}{\partial \dot{x}^k} \dot{x}^j \right) dx^i = -d_x L_0.$$

(iii) By (7.7), conditions (7.48) and (7.46),

$$\frac{\overline{\mathbf{D}}}{dt}(dS) = \left(\frac{\partial}{\partial t} + \nabla_{D_\nabla X} + \frac{1}{2} \Delta_{\mathrm{LD}} \right) (dS)$$

$$= d\frac{\partial S}{\partial t} + \nabla_{(\nabla_p H_0 \circ dS)} dS - \frac{1}{2}(dd^* + d^*d)dS$$

$$= d\frac{\partial S}{\partial t} + d(H_0 \circ dS) - d_x H_0 \circ dS - \frac{1}{2} dd^* dS$$

$$= d\left(\frac{\partial S}{\partial t} + H_0 \circ dS + \frac{1}{2} \Delta S \right) - d_x H_0 \circ dS.$$

The result follows.

$\square$

Remark 7.23 (i) Assertions (ii) and (iii) of Theorem 7.22 generalize Theorem 7.18, since from the Legendre transform $p = d_{\dot{x}} L_0$ we observe that the S-EL equation (7.22) is related to HJB equation (7.49) via Eq. (7.29). However, assertion (iii) is a special case of Theorem 6.19, since HJB equation (7.49) is just the one in (6.39) with $H = \overline{H}_0^g$ the g-canonical lift of H_0, due to the observation that $\overline{H}_0^g(d^2 S, t) = H_0(dS, t) + \frac{1}{2}\Delta S$.

(ii) The advantage of Theorem 7.22 is that it formulates stochastic Hamiltonian mechanics in a global way similar to stochastic Lagrangian mechanics, while its disadvantage is that it depends on the choice of Riemannian structures. However, unlike stochastic Hamiltonian mechanics of Sect. 6, neither global S-H equations (7.47) nor HJB equation (7.49) encodes any new symplectic or contact structures, as the Hamiltonian functions therein are still classical.

(iii) By a direct calculation similar to (7.50), one easily obtains the following local version of stochastic Euler–Lagrange equation (7.22):

$$\mathbf{D}_t \left(\frac{\partial L_0}{\partial \dot{x}^i} \right) = \frac{\partial L_0}{\partial x^i} + \frac{1}{2} g^{jk} \partial_i \Gamma_{jk}^l \frac{\partial L_0}{\partial \dot{x}^l} - \frac{1}{2} \partial_i g^{jk} \left(\frac{\partial^2 L_0}{\partial x^j \partial \dot{x}^k} - \Gamma_{jk}^l \frac{\partial L_0}{\partial \dot{x}^l} \right). \tag{7.51}$$

This local version is related to stochastic Hamilton's equations (6.10) via the canonical second-order Legendre transform (7.43).

(iv) Similarly to Remark 6.20, if we let $\tilde{H} = H - f$, then Theorem 7.22 holds with $\tilde{H}$ and zero function in place of H and f. We will refer to Eq. (7.49) with $f \equiv 0$ as the HJB equation associated with Hamiltonian H_0, or the HJB equation associated with the Lagrangian L_0 related to H_0 via the Legendre transform (when H_0 is hyperregular).

On Riemannian manifolds, canonical transformations of Sect. 6.5 can also be reduced to tangent bundles. We consider a bundle isomorphism $\mathbf{F}$ from $\mathcal{T}^{S*}M \times \mathbb{R}$ to $\mathcal{T}^{S*}N \times \mathbb{R}$, projecting to a time-change map $F^0 : \mathbb{R} \to \mathbb{R}$. The transformation $\mathbf{F}$ is a map from coordinates (x^i, p_i, o_{jk}, t) to (y^i, P_i, O_{jk}, s) satisfying $s = F^0(t)$. Both base manifolds M and N are equipped with some Riemannian metrics and the corresponding Levi–Civita connections.

By the inverse second-order Legendre transform (7.44) and the integrability condition (6.15), the action functional in (7.9) can be rewritten as

$$
\begin{aligned}
&\mathcal{S}[X; 0, T] \\
&= \mathbf{E} \int_0^T \left[p_i(t, X(t))(DX)^i(t) + \frac{1}{2} \frac{\partial p_j}{\partial x^k}(t, X(t))(QX)^{jk}(t) - \overline{H}_0^g(\mathbf{X}(t), t) \right] dt \\
&= \mathbf{E} \int_0^T \left[p_i(t, X(t)) \circ dX^i(t) - \overline{H}_0^g(\mathbf{X}(t), t) dt \right],
\end{aligned}
$$

where $\circ d$ denotes the Stratonovich stochastic differential and $\overline{H}_0^g = H_0 + \frac{1}{2} g^{jk}(o_{jk} - \Gamma_{jk}^i p_i)$. We denote simply $x^i = x^i \circ \mathbf{X}$, $p_i = p_i \circ \mathbf{X}$ and $H = \overline{H}_0^g$. Then, $\mathcal{S} = \mathbf{E} \int_0^T (p_i \circ dx^i(t) - H dt)$. Now we make a change of coordinates from (x^i, p_i, t) to (y^i, P_i, s) satisfying $s = F^0(t)$, and denote that $y^i = y^i \circ \mathbf{X}$ and $P_i = P_i \circ \mathbf{X}$. We have

$$
\mathcal{S} = \mathbf{E} \int_0^T \left(P_i \circ dy^i(s) - K ds \right) = \mathbf{E} \int_0^T \left(P_i \circ d(y^i \circ F^0)(t) - K \dot{F}^0 dt \right),
$$

where the function K plays the role of the second-order Hamiltonian in new coordinate system.

As in Sect. 6.5, the general condition for a transformation to be canonical is to preserve the form of stochastic Hamilton's system (7.47). This is equivalent to preserve the form of stochastic stationary-action principle (7.12), according to Theorem 7.22.(ii). It follows from $\delta\mathcal{S} = 0$ that

$$
\delta \mathbf{E} \int_0^T \left(p_i \circ dx^i(t) - H dt \right) = \delta \mathbf{E} \int_0^T \left(P_i \circ d(y^i \circ F^0)(t) - K \dot{F}^0 dt \right) = 0.
$$

Since the underlying process X has zero variation at the endpoints, both equalities will be satisfied if the integrands are related by the following SDE:

$$
p_i \circ dx^i - H dt = P_i \circ dy^i - K \dot{F}^0 dt + dG, \tag{7.52}
$$

where G is a function of phase space coordinates (x, p, t) or (y, P, s) or any mixture of them and called the generating function. Note that in contrast with the classical theory of canonical transformation and also (6.36), here Eq. (7.52) for canonical transformations is a stochastic differential equation, instead of equation for forms.

Consider the *type one* generating function G_1, that is, $G = G_1(x, y, t)$ is given as a function of the old and new generalized position coordinates (cf. Goldstein et al. 2002, Section 9.1). Then, using Itô's formula $dG_1 = \frac{\partial G_1}{\partial t} dt + \frac{\partial G_1}{\partial x^i} \circ dx^i + \frac{\partial G_1}{\partial y^i} \circ dy^i$, and vanishing the coefficients of every (stochastic) differentials $\circ dx$, $\circ dy$ and dt in (7.52), we get

$$p_i = \frac{\partial G_1}{\partial x^i}, \quad P_i = -\frac{\partial G_1}{\partial y^i}, \quad K \dot{F}^0 - H = \frac{\partial G_1}{\partial t},$$

which recovers (6.37). By taking $F^0 = \mathbf{Id}_{\mathbb{R}}$ (i.e., no time-change) and requiring the new Hamiltonian K_0 to be identically zero, and writing G_1 as S the last equation turns into the following HJB equation

$$\frac{\partial S}{\partial t}(x, y, t) + H_0\left(x^i, \frac{\partial S}{\partial x^i}(x, y, t), t\right) + \frac{1}{2}\Delta_x S(x, y, t) + \frac{1}{2}\Delta_y S(x, y, t) = 0,$$

where (x, y) are regarded as coordinates on the product manifold $M \times N$ equipped with the direct-sum Riemannian metric and its corresponding Levi–Civita connection, Δ_x and Δ_y are the Laplacian on M and N, respectively, so that $\Delta_x + \Delta_y$ is the Laplacian on $M \times N$ under the aforementioned connection.

In contrast to the mixed-order contact approach to canonical transformations of Sect. 6.5, since the changes of coordinates proceed on T^*M, one can easily formulate *four types* of generating functions that are related to each other through classical Legendre transforms in the same way as in classical mechanics (Goldstein et al. 2002, Section 9.1). For example, the *type two* generating function takes the form $G = G_2(x, P, t) - y^i P_i$, for which we have

$$p_i = \frac{\partial G_2}{\partial x^i}, \quad y^i = \frac{\partial G_2}{\partial P_i}, \quad K\dot{F}^0 - H = \frac{\partial G_2}{\partial t}. \tag{7.53}$$

In this case, since (x^i) and (y^i) are no longer independent variables, Riemannian structures on M and N should be related by the transformation. In view of this, we only consider point transformations, a subclass of canonical transformations. That is, we assume G_2 to be the form

$$G_2(x, P, t) = f^i(x, t)P_i + h(x, t)$$

for some diffeomorphisms $f : M \to N$'s and $h : M \to \mathbb{R}$. The second equation of (7.53) implies

$$y^i = f^i(x, t).$$

So we equip N with the (time-dependent) pushforward Riemannian metric of g on M by f, and with the Levi–Civita connection.

Example 7.24 (Canonical transformations for one-dimensional Bernstein's reciprocal processes) Consider the scalar case of Example 6.11, that is, the $\mathbb{R}$-valued Brownian reciprocal process with second-order Hamiltonian $H(x, p, o) = H_0(x, p) + \frac{1}{2}o = \frac{1}{2}|p|^2 + \frac{1}{2}o$. The equations of motion are $DDX = 0$, $QX = 1$ [cf. (6.33)]. In the following, we will consider two canonical transformations which transform Brownian reciprocal processes to reciprocal processes derived from diffusions with linear potentials and quadratic potentials, respectively.

(i) Consider the time-dependent change of coordinates from (x, p, t) to (y, P, t) (without time-change) induced by $G_2(x, P, t) = (x + \frac{1}{2}t^2)P - tx$. By (7.53),

$$y = x + \tfrac{1}{2}t^2, \quad p = P - t, \quad K = H + Pt - y + \tfrac{1}{2}t^2. \tag{7.54}$$

For the latent second-order coordinates, we have

$$O = \frac{\partial P}{\partial y} = \frac{\partial p}{\partial x} = o.$$

Hence, by the last equation of (7.53), the new second-order Hamiltonian is

$$K(y, P, O, t) = K_0(y, P, t) + \frac{1}{2}O = \frac{1}{2}|P|^2 - y + t^2 + \frac{1}{2}O,$$

which is still of the form (6.26), with $b \equiv 0$ and $F(t, y) = -y + t^2$. The equations of motion under new coordinates are $DDY = 1$ and $QY = 1$. By Remark 6.20, K share the same equations of motion with $\tilde{K}(y, P, O) = \frac{1}{2}|P|^2 - y + \frac{1}{2}O$. In other words, (7.54) transforms Brownian reciprocal processes to reciprocal processes derived from diffusions with linear potentials. This example is taken from Lescot and Zambrini (2007, Theorem 4.1.(1)), where the authors used (7.54) to transform free heat equations to heat equations with linear potentials. We refer readers to Lescot and Zambrini (2007) for more applications of canonical transformations of contact Hamiltonian systems to Euclidean quantum mechanics in Example 6.12.

(ii) Consider the following change of coordinates from (x, p, t) to (y, P, s) (with time-change)

$$x = y\sqrt{1 - t^2}, \quad P = p\sqrt{1 - t^2} + yt, \quad s = \operatorname{arctanh} t. \tag{7.55}$$

Clearly, the map $(x, p) \mapsto (y, P)$ is induced by the type three generating function $G_3(y, p, t) = -py\sqrt{1 - t^2} - \frac{y^2}{2}t$ via relations $x = -\frac{\partial G_3}{\partial p}$ and $P = -\frac{\partial G_3}{\partial y}$. The relation between the latent coordinates o and O is

$$O = \frac{\partial P}{\partial y} = \frac{\partial p}{\partial x}\frac{\partial x}{\partial y}\sqrt{1 - t^2} + t = (1 - t^2)o + t. \tag{7.56}$$

The new second-order Hamiltonian K satisfies $K\frac{ds}{dt} - H = \frac{\partial G_3}{\partial t}$. Hence, combining with (7.55) and (7.56), we obtain

$$K(y, P, O, s) = (1 - t^2)\left(\frac{1}{2}|p|^2 + \frac{pyt}{\sqrt{1 - t^2}} - \frac{1}{2}|y|^2 + \frac{1}{2}o\right)$$

$$= \frac{1}{2}|P|^2 - \frac{1}{2}|y|^2 + \frac{1}{2}O - \frac{1}{2}\tanh s.$$

This differs with the second-order Hamiltonian of Euclidean harmonic oscillators in Example 6.12.(ii), i.e., $\tilde{K}(y, P, O) = \frac{1}{2}|P|^2 - \frac{1}{2}|y|^2 + \frac{1}{2}O$, by a term depending only on time. So by virtue of Remark 6.20, K and $\tilde{K}$ share the same equations of motion $DDY = Y$, $QY = 1$. Therefore, (7.55) transforms free reciprocal processes to Euclidean harmonic oscillators.

Example 7.25 (Canonical transformations for vanishing potentials) Let (M, g) be Riemannian. Take $G_2(x, P, t) = x^i P_i - S(x, t)$ for some function S. Then,

$$y = x, \quad p_i = P_i - \frac{\partial S}{\partial x^i}, \quad K = H - \frac{\partial S}{\partial t}.$$

Since the transformation on base manifold M is identity, it does not change the Riemannian metric, and

$$o_{ij} = \frac{\partial p_i}{\partial x^j} = \frac{\partial P_i}{\partial y^j} - \frac{\partial^2 S}{\partial x^i \partial x^j}.$$

(i) We consider the Hamiltonian $H_0(x, p) \equiv b^i(x)p_i - F(x)$, whose corresponding second-order Hamiltonian $H = \overline{H_0^g}$ has a diffusion with generator $\frac{1}{2}\Delta + b \cdot \nabla - F$ for solution process (see Sect. 6.3.1). Then, the new Hamiltonian is

$$K_0(y, P, t) = b^i(y)P_i - \langle b(y), \nabla S(y, t)\rangle$$
$$- F(y) - \frac{1}{2}\Delta S(y, t) - \frac{\partial S}{\partial t}(y, t).$$

If we choose S solving the backward PDE (6.23), then $K_0(y, P) = b^i(y)P_i$ has a diffusion process with generator $\frac{1}{2}\Delta + \nabla_b$ for solution. In particular, such a canonical transformation can transform a diffusion process with a scalar potential into a free motion.

(ii) Consider the Hamiltonian $H_0(x, p, t) = \frac{1}{2}g^{ij}(x)p_i p_j + g^{ij}(x)p_i\frac{\partial S}{\partial x^j}(x, t) + b^i(x)p_i - F(x)$, whose corresponding second-order Hamiltonian $H = \overline{H_0^g}$ has a Schrödinger's bridge with vector potential $(b + \nabla S)$ and scalar potential $-F$ for solution process. Then, the new Hamiltonian is

$$K_0(y, P, t) = \frac{1}{2}g^{ij}(y)P_i P_j + b^i(y)P_i - \langle b(y), \nabla S(y, t)\rangle - \frac{1}{2}|\nabla S(y, t)|^2 - F(y)$$
$$- \frac{1}{2}\Delta S(y, t) - \frac{\partial S}{\partial t}(y, t).$$

To transform K_0 into the standard form $K_0(y, P, t) = \frac{1}{2} g^{ij}(y) P_i P_j + b^i(y) P_i$ whose solution is a Schrödinger's bridge with vector potential b, we only need to assume that S solves HJB equation (6.28). In particular, such a canonical transformation transforms a Schrödinger's bridge with a scalar potential into a free one.

Regarding the classical energy introduced in the end of Sect. 7.4.1, for a given classical Lagrangian $L_0 : \mathbb{R} \times TM \to \mathbb{R}$, we introduce its *generalized (or deformed) energy* $E : \mathbb{R} \times TM \to \mathbb{R}$ by

$$E(t, x, \dot{x}) = E_0(t, x, \dot{x}) + \tfrac{1}{2} \Delta S(t, x),$$

where S is the solution of the Hamilton–Jacobi–Bellman equation (7.49) associated with L_0 (with $f \equiv 0$). The term $\frac{1}{2} \Delta S$ stands for the stochastic deformation.

7.4.3 Small-Noise Limits

In this part, we will see, informally, how our stochastic framework degenerates into classical mechanics as the noise goes to zero. Let $\epsilon > 0$ be a small parameter which we refer to as *diffusivity*. The limit when $\epsilon \to 0$ is called the *small-noise limit*.

Let $\mathcal{A}_g^\epsilon([0, T]; q, \mu)$ be the small-noise version of the admissible class (7.10), that is, with the constraint $QX(t) = \epsilon \breve{g}(X(t))$. The ϵ-dependent stochastic variational problem is to minimize the action functional $\mathcal{S}[X; 0, T]$ in (7.9) among all $X \in \mathcal{A}_g^\epsilon([0, T]; q, \mu)$. Then, the same procedure as Sect. 7.2 yields the following ϵ-dependent stochastic Euler–Lagrange equation:

$$\frac{\overline{\mathbf{D}}^\epsilon}{dt} \big(d_{\dot{x}} L_0 \left(t, X_\epsilon(t), D_\nabla X_\epsilon(t) \right) \big) = d_x L_0 \left(t, X_\epsilon(t), D_\nabla X_\epsilon(t) \right), \qquad (7.57)$$

which is an equivalent condition for $X_\epsilon \in \mathcal{A}_g^\epsilon([0, T]; q, \mu)$ to be a stationary point of $\mathcal{S}$. Here $\frac{\overline{\mathbf{D}}^\epsilon}{dt}$ is the damped mean covariant derivative with respect to X_ϵ so that

$$\frac{\overline{\mathbf{D}}^\epsilon}{dt} = \frac{\partial}{\partial t} + \nabla_{D_\nabla X} + \frac{\epsilon}{2} \Delta_{\mathrm{LD}}.$$

Now as $\epsilon \to 0$, since $QX_\epsilon \to 0$, X_ϵ tends to some deterministic curve $\gamma = (\gamma(t))_{t \in [0,T]}$ (in a suitable probabilistic sense), and $D_\nabla X_\epsilon(t)$ tends to $\dot{\gamma}(t)$. Thus, we can write informally

$$\mathcal{A}_g^\epsilon([0, T]; q, \mu) \to \mathcal{A}_g^0([0, T]; q, \mu)$$
$$:= \Big\{ \gamma \text{ is adapted with paths in } C^2([0, T], M) : \gamma(0) = q, \mathbf{P} \circ (\gamma(T))^{-1} = \mu \Big\}.$$

The ϵ-dependent stochastic variational problem tends to the following deterministic variational problem

$$\min_{\gamma \in \mathcal{A}_g^0([0,T];q,\mu)} \int_0^T L_0\left(t, \gamma(t), \dot\gamma(t)\right) \mathrm{d}t. \tag{7.58}$$

And the ϵ-dependent stochastic Euler–Lagrange equation (7.57) tends to

$$\frac{D}{dt}\left(d_{\dot x}L_0\left(t, \gamma(t), \dot\gamma(t)\right)\right) = d_x L_0\left(t, \gamma(t), \dot\gamma(t)\right), \tag{7.59}$$

where, $\frac{D}{dt} = \frac{\partial}{\partial t} + \nabla_{\dot\gamma}$ is the material derivative along γ. This is the classical Euler–Lagrange equation in global form, cf. Villani (2009, p. 153).

We introduce the following ϵ-dependent version of the g-canonical lift (6.44):

$$H_\epsilon(x, p, o, t) := H_0(x, p, t) + \frac{\epsilon}{2}g^{jk}(x)\left(o_{jk} - \Gamma^i_{jk}(x)p_i\right).$$

Let $\mathbf{X}_\epsilon$ be a horizontal integral process of stochastic Hamilton's equations (6.10) corresponding to H_ϵ and $X_\epsilon = \tau_M^{S*}(\mathbf{X}_\epsilon)$. Since $(Q(x \circ \mathbf{X}_\epsilon))^{jk} = 2\frac{\partial H_\epsilon}{\partial o_{jk}} = \epsilon\check g \to 0$ as $\epsilon \to 0$, $\mathbf{X}_\epsilon$ converges to a T^*M-valued process. And since $\frac{\partial H_\epsilon}{\partial p_i} \to \frac{\partial H_0}{\partial p_i}$ and $\frac{\partial H_\epsilon}{\partial x^i} \to \frac{\partial H_0}{\partial x^i}$, the limit T^*M-valued process satisfies classical Hamilton's equations,

$$\begin{cases} \dot x^i(t) = \frac{\partial H_0}{\partial p_i}(x(t), p(t), t), \\ \dot p_i(t) = -\frac{\partial H_0}{\partial x^i}(x(t), p(t), t). \end{cases} \tag{7.60}$$

Let $\mathbb{X}_\epsilon := \hat\varrho^*(\mathbf{X}_\epsilon)$. Then, $\mathbb{X}_\epsilon(t) = p(t, X_\epsilon(t))$ solves the system of global stochastic Hamilton's equations (7.47), with $\mathbb{X}_\epsilon$, X_ϵ and $\frac{\overline{\mathbf{D}}^\epsilon}{dt}$ in place of $\mathbb{X}$, X and $\frac{\overline{\mathbf{D}}}{dt}$, respectively, subject to $QX_\epsilon(t) = \epsilon\check g(X_\epsilon(t))$. As ϵ goes to 0, this system tend to the following deterministic system,

$$\begin{cases} \dot x(t) = \nabla_p H_0(x(t), p(t), t), \\ \dfrac{D}{dt}p(t) = -d_x H_0(x(t), p(t), t), \end{cases} \tag{7.61}$$

This is indeed the global form of (7.60) which is equivalent to the global Euler–Lagrange equation (7.59) via the classical Legendre transform.

The corresponding ϵ-dependent Hamilton–Jacobi–Bellman equation is now

$$\frac{\partial S}{\partial t} + H_\epsilon(d^2S, t) = \frac{\partial S}{\partial t} + H_0(dS, t) + \frac{\epsilon}{2}\Delta S = f(t),$$

which, as $\epsilon \to 0$, goes to the classical Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + H_0(dS, t) = f(t).$$

The latter corresponds to (7.59)–(7.61) via classical Hamilton–Jacobi theory (e.g., Abraham and Marsden 1978, Chapter 5).

We list here some previous works that have independent interests in the above small-noise limits, in some special cases. The time-asymptotic large deviation for Brownian bridges of Example 6.11 was studied in Hsu (1990). The second author of the present paper and his collaborator proved in Privault et al. (2016) a large deviation result for one-dimensional Bernstein bridges which are solution processes of Euclidean quantum mechanics in Example 6.12. The paper (Léonard 2012b) proved that the Γ-limit of Schrödinger's problem in Sect. 7.3 with small variance is the Monge–Kantorovich problem. The latter is the optimal transport problem associated with the classical variational problem (7.58) (Villani 2009, Chapter 7). See Mikami (2021, Section 2.3) for more on small-noise limits of stochastic optimal transport.

Remark 7.26 There are various terminologies in other areas related to the small-noise limit. In thermodynamics (Huang and Zambrini 2023), ϵ stands for the Boltzmann constant which relates to the diffusion coefficient via Einstein relation, as consistent with Schrödinger's original statistical problem (Schrödinger 1932); when applied to quantum mechanics as in Example 6.12, the small-noise limit is called the *semiclassical limit* and the parameter ϵ stands for the reduced Planck constant $\hbar$; when/if applied to hydrodynamics (cf. Arnaudon et al. 2014; Chen et al. 2023), it is often called the *vanishing viscosity limit* and ϵ stands for the kinematic viscosity ν. The latter may be expected to solve Kolmogorov's conjecture that the "stochastization" of dynamical systems is related to hydrodynamic PDEs as viscosity vanishes (Arnold and Khesin 2021). In physics, diffusivity, Planck constant and viscosity are indeed related to each other (Trachenko and Brazhkin 2021).

7.4.4 Relations to Stochastic Optimal Control

Following the way of converting problems of classical calculus of variations into optimal control problems (see Fleming and Soner 2006), we can regard the stochastic variational problem of Sect. 7.2 as a stochastic optimal control problem.

Assume that (M, g) is compact (for simplicity). Consider a stochastic control model in which the state evolves according to an M-valued diffusion X governed by a system of MDEs on the time interval $[t, T]$, of the form

$$\begin{cases} D_\nabla X(s) = U(s), \\ Q X(s) = g(X(s)), \end{cases} \tag{7.62}$$

or equivalently, by an Itô SDE of the form

$$dX^i(s) = \left(U^i(s) - \frac{1}{2} g^{jk}(X(s)) \Gamma^i_{jk}(X(s)) \right) ds + \sigma^i_r(X(s)) dW^r(s), \quad s \in [t, T],$$

where σ is the positive-definite square root $(1, 1)$-tensor of g, i.e., $\sum_{r=1}^{d} \sigma^i_r \sigma^j_r = g^{ij}$, W is an $\mathbb{R}^d$-valued standard Brownian motion and, most importantly, U is a TM-valued process called the control process. There are no control constraints for U as it

is admissible in the sense of Fleming and Soner (2006, Definition 2.1). As endpoint condition, we require that $X(t) = x$.

The control problem on a finite time interval $s \in [t, T]$ is to choose U to minimize

$$J(t, x; X, U) := \mathbf{E}_{(t,x)} \left[\int_t^T L_0\left(s, X(s), U(s)\right) \mathrm{d}s - S_T(X(T)) \right], \quad (7.63)$$

among all pairs (X, U) satisfying the system (7.62) and the endpoint condition, where S_T is a given smooth function on M. The real-valued smooth function L_0 on $\mathbb{R} \times TM$ is called running cost function and J the payoff functional. The problem is called a *stochastic Bolza problem*. In the case $S_T \equiv 0$, this stochastic control problem is of the same form as our stochastic variational problem of Sect. 7.2. For this reason, we call the latter stochastic control problem to be in *Lagrange form*. By an argument similar to Theorem 7.16, one can derive the same equation as (7.22), but with boundary conditions $X(t) = x$ and $d_{\dot{x}} L_0(T, X(T), D_\nabla X(T)) = dS_T(X(T))$.

The starting point of *dynamic programming* is to regard the infimum of J being minimized as a function $S(t, x)$ of the initial data:

$$S(t, x) = - \inf_{(X,U)} J(t, x; X, U).$$

Then, Bellman's principle of dynamic programming (Fleming and Soner 2006, Section III.7) states that for $t \le t + \epsilon \le T$,

$$0 = \inf_{X \in I_0^T(M)} \mathbf{E}_{(t,x)} \left[\int_t^{t+\epsilon} L_0\left(s, X(s), D_\nabla X(s)\right) \mathrm{d}s - S(t + \epsilon, X(t + \epsilon)) + S(t, x) \right].$$

Divide the equation by ϵ, let $\epsilon \to 0^+$, and then use Dynkin's formula. We get the dynamic programming equation

$$0 = \inf \left[L_0(t, x, D_\nabla x) - (\mathbf{D}_t S)(t, x, Dx, Qx) \right], \quad (7.64)$$

subjected to terminal data $S(T, x) = S_T(x)$. By (4.5) and (7.62),

$$\mathbf{D}_t S = \partial_t S + D^i x \partial_i S + \tfrac{1}{2} Q^{ij} x \partial_i \partial_j S = \partial_t S + \left(D_\nabla^i x - \tfrac{1}{2} \Gamma_{jk}^i g^{jk} \right) \partial_i S + \tfrac{1}{2} g^{ij} \partial_i \partial_j S.$$

We let

$$H(x, p, o, t) = \sup \left[\left(D_\nabla^i x - \tfrac{1}{2} \Gamma_{jk}^i(x) g^{jk}(x) \right) p_i + \tfrac{1}{2} g^{ij}(x) o_{ij} - L_0(t, x, D_\nabla x) \right]$$

where the supremum can be ignored if L_0 is convex, so that $H = \overline{H_0^g}$ is exactly the canonical inverse second-order Legendre transform in (7.43). Then, the dynamic programming equation (7.64) can be written as the HJB equation (6.38), cf. Fleming and Soner (2006, Section IV.3).

There is also a stochastic version of Pontryagin's maximum principle (Yong and Zhou 1999, Theorem 3.3.2). The crucial objects in stochastic Pontryagin's principle are first- and second-order adjoint processes, p and o, respectively. Corresponding to the stochastic control problem (7.62)–(7.63), its adjoint processes p and o satisfy the following backward SDEs (Yong and Zhou 1999, Section 3.3.2) (where "backward" is again in a different sense from ours in Sect. 2),

$$
\begin{cases}
dp_i(t) = \left[\frac{1}{2} \left(\partial_i g^{kl} \Gamma_{kl}^j + g^{kl} \partial_i \Gamma_{kl}^j \right)(X(t)) p_j(t) - \sum_{r=1}^d \partial_i \sigma_r^j(X(t)) z_{jr}(t) \right. \\
\qquad\qquad \left. + \frac{\partial L_0}{\partial x^i}(t, X(t), U(t)) \right] dt \\
\qquad\qquad + z_{ir}(t) dW^r(t), \\
p_i(T) = \partial_i S_T(X(T)),
\end{cases}
\tag{7.65}
$$

and

$$
\begin{cases}
do_{ij}(t) = -\left[\frac{\partial^2 \overline{H}_0^g}{\partial x^i \partial x^j}(X(t), p(t), o(t), t) \right. \\
\qquad\quad + o_{ik}(t) \left(o_{jl}(t) \frac{\partial^2 H_0}{\partial p_k \partial p_l}(X(t), p(t), t) + 2 \frac{\partial^2 H_0}{\partial x^j \partial p_k}(X(t), p(t), t) \right) \\
\qquad\quad - o_{ik}(t) \left(\partial_j g^{lm} \Gamma_{lm}^k + g^{lm} \partial_j \Gamma_{lm}^k \right)(X(t)) \\
\qquad\quad \left. + \sum_{r=1}^d \left(\partial_j \sigma_r^k(X(t)) Z_{ikr}(t) + \partial_j \sigma_r^l(X(t)) Z_{ilr}(t) \right) \right] dt + Z_{ijr}(t) dW^r(t), \\
o_{ij}(T) = \partial_i \partial_j S_T(X(T)),
\end{cases}
\tag{7.66}
$$

which are called first- and second-order adjoint equation, respectively. The unknowns in (7.65) and (7.66) are the pairs (p, z) and (o, Z), respectively. Suppose that $p_i(t) = p_i(t, X(t))$ and $o_{ij}(t) = o_{ij}(t, X(t))$ for time-dependent second-order form (p, o) that satisfies second-order Maxwell relations (6.15). Then,

$$
z_{ir} = \frac{\partial p_i}{\partial x^j} \sigma_r^j, \quad Z_{ijr} = \frac{\partial o_{ij}}{\partial x^k} \sigma_r^k.
$$

Plugging them into (7.65) and (7.66), we get

$$
\begin{aligned}
D_i p &= \frac{1}{2} \left(\partial_i g^{kl} \Gamma_{kl}^j + g^{kl} \partial_i \Gamma_{kl}^j \right) p_j - \frac{1}{2} \partial_i g^{jk} \frac{\partial p_j}{\partial x^k} + \frac{\partial L_0}{\partial x^i} = -\frac{\partial \overline{H}_0^g}{\partial x^i}, \\
D_{ij} o &= -\left(\frac{\partial^2 \overline{H}_0^g}{\partial x^i \partial x^j} + \frac{\partial p_k}{\partial x^i} \frac{\partial p_l}{\partial x^j} \frac{\partial^2 \overline{H}_0^g}{\partial p_k \partial p_l} + 2 \frac{\partial p_k}{\partial x^i} \frac{\partial^2 \overline{H}_0^g}{\partial x^j \partial p_k} + 2 \frac{\partial o_{kl}}{\partial x^i} \frac{\partial^2 \overline{H}_0^g}{\partial x^j \partial o_{kl}} \right).
\end{aligned}
\tag{7.67}
$$

These coincide with the corresponding equations in the S-H system (6.10) for second-order Hamiltonian $\overline{H}_0^g$. The first equality of (7.67) also recovers (7.51).

7.5 Stochastic Variational Symmetries

Definition 7.27 Given an action functional $\mathcal{S}$ as in (7.9), a bundle automorphism F on $(\mathbb{R} \times M, \pi, \mathbb{R})$ projecting to F^0 is called a variational symmetry of $\mathcal{S}$ if, whenever $[t_1, t_2]$ is a subinterval of $[0, T]$, we have $\mathcal{S}[F \cdot X, F^0(t_1), F^0(t_2)] = \mathcal{S}[X, t_1, t_2]$. A π-projectable vector field V on $\mathbb{R} \times M$ is called an infinitesimal variational symmetry of $\mathcal{S}$, if its flow consists of variational symmetries of $\mathcal{S}$.

Lemma 7.28 *The π-projectable vector field V of the form (4.9) is an infinitesimal variational symmetry of $\mathcal{S}$ if and only if*

$$\left[(j^\nabla V)(L_0) + L_0 \dot{V}^0 \right] \left(j_t^\nabla X \right), \quad t \in [0, T]$$

is a martingale, for all $X \in I_0^T(M)$.

Proof As in the proof of Theorem 4.14, we let $\psi = \{(\psi_\epsilon^0, \bar{\psi}_\epsilon)\}_{\epsilon \in \mathbb{R}}$ be the flow generated by V, and denote $\tilde{X}_\epsilon = \psi_\epsilon \cdot X$. Then, by a change of variable $s = \psi_\epsilon^0(t)$,

$$\mathcal{S}[\tilde{X}_\epsilon, \psi_\epsilon^0(t_1), \psi_\epsilon^0(t_2)] = \mathbf{E} \int_{\psi_\epsilon^0(t_1)}^{\psi_\epsilon^0(t_2)} L_0 \left(s, \tilde{X}_\epsilon(s), D_\nabla \tilde{X}_\epsilon(s) \right) \mathrm{d}s$$

$$= \mathbf{E} \int_{t_1}^{t_2} L_0 \left(\psi_\epsilon^0(t), \bar{\psi}_\epsilon(t, X(t)), D_\nabla \tilde{X}_\epsilon \left(\psi_\epsilon^0(t) \right) \right) \frac{\mathrm{d}\psi_\epsilon^0}{\mathrm{d}t}(t) \mathrm{d}t.$$

Since for all $[t_1, t_2] \subset [0, T]$ and each ϵ, $\mathcal{S}[\tilde{X}_\epsilon, \psi_\epsilon^0(t_1), \psi_\epsilon^0(t_2)] = S[X, t_1, t_2]$, we have that the difference

$$L_0 \left(\psi_\epsilon^0(t), \bar{\psi}_\epsilon(t, X(t)), D_\nabla \tilde{X}_\epsilon(\psi_\epsilon^0(t)) \right) \frac{\mathrm{d}\psi_\epsilon^0}{\mathrm{d}t}(t) - L_0 \left(t, X(t), D_\nabla X(t) \right).$$

is a martingale (depending on ϵ). Taking derivatives with respect to ϵ and evaluating at $\epsilon = 0$ for the above equality, and recalling that $j^\nabla V = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\big|_{\epsilon=0} j^\nabla \psi_\epsilon$, we can obtain the desired result.   □

Definition 7.29 Given a smooth function $\Phi : \mathbb{R} \times M \to \mathbb{R}$. A π-projectable vector field V on $\mathbb{R} \times M$ is called an infinitesimal Φ-divergence symmetry of $\mathcal{S}$, if

$$\left[(j^\nabla V)(L_0) + L_0 \dot{V}^0 \right] \left(j_t^\nabla X \right) = \mathbf{D}_t \Phi \left(j_t^\nabla X \right),$$

for all $X \in I_0^T(M)$ and $t \in [0, T]$.

Recall that for the π-projectable vector field V of the form (4.9), we denote $\bar{V} = V^i \frac{\partial}{\partial x^i}$, as in Corollary 4.17.

Proposition 7.30 *A vector field V of the form (4.9) is an infinitesimal Φ-divergence symmetry of $\mathcal{S}$ if and only if*

$$V^0 \partial_t L_0 + d_x L_0(\bar{V}) + d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}}\bar{V}}{dt} \right) - \dot{V}^0 E_0 = \mathbf{D}_t \Phi.$$

Proof It follows from Corollary 4.17 and (7.19), (7.20) that

$$
\begin{aligned}
\mathbf{D}_t \Phi &= V^0 \partial_t L_0 + V^i \partial_i L_0 \\
&\quad + \left[\left(\partial_t + \dot{x}^j \partial_j \right) V^i + \tfrac{1}{2} \left(\Delta \bar{V} + \mathrm{Ric}(\bar{V}) \right)^i - \dot{V}^0 \dot{x}^i \right] \partial_{\dot{x}^i} L_0 + \dot{V}^0 L_0 \\
&= V^0 \partial_t L_0 + d_x L_0(\bar{V}) + d_{\dot{x}} L_0 \left(\left(\partial_t + \nabla_{\dot{x}} + \tfrac{1}{2}\Delta_{\mathrm{LD}} \right) \bar{V} \right) - \dot{V}^0 \left(\dot{x}^i \partial_{\dot{x}^i} L_0 - L_0 \right) \\
&= V^0 \partial_t L_0 + d_x L_0(\bar{V}) + d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}}\bar{V}}{dt} \right) - \dot{V}^0 E_0.
\end{aligned}
$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Corollary 7.31 *Let $L_0 : \mathbb{R} \times TM \to \mathbb{R}$ be a hyperregular Lagrangian.*

Let V be a vector field of the form (4.9). Given a smooth function $\Phi : \mathbb{R} \times M \to \mathbb{R}$, define the Φ-extension of V by

$$V_\Phi = V + \Phi \frac{\partial}{\partial u}, \tag{7.68}$$

which is a vector field on $\mathbb{R} \times M \times \mathbb{R}$. Suppose that V satisfies

$$\frac{1}{2}\dot{V}^0 \Delta S = g^{ij} \nabla^2_{\partial_i, \nabla_{\partial_j}\bar{V}} S,$$

for S the solution of the Hamilton–Jacobi–Bellman equation (7.49) associated with L_0 (for $f \equiv 0$). Then, V is an infinitesimal Φ-divergence symmetry of $\mathcal{S}$ if and only if V_Φ is an infinitesimal symmetry of equation (7.49).

Proof By the classical jet bundle theory, we know that V is an infinitesimal symmetry of Hamilton–Jacobi–Bellman equation (7.49) if and only if (Olver 1998, Theorem 2.31)

$$j^{1,2} V \left(u_t + H_0(x, (u_i), t) + \tfrac{1}{2} g^{ij}(x) u_{ij} - \tfrac{1}{2} g^{ij}(x) \Gamma^k_{ij}(x) u_k \right) = 0, \tag{7.69}$$

where

$$j^{1,2} V = V^0 \frac{\partial}{\partial t} + V^i \frac{\partial}{\partial x^i} + \Phi \frac{\partial}{\partial u} + V_t \frac{\partial}{\partial u_t} + V_i \frac{\partial}{\partial u_i} + V_{ij} \frac{\partial}{\partial u_{ij}},$$

with coefficients given by Olver (1998, Theorem 2.36 or Example 2.38)

$$V_t = \frac{\partial \Phi}{\partial t} - \dot{V}^0 u_t - \frac{\partial V^i}{\partial t} u_i, \quad V_i = \frac{\partial \Phi}{\partial x^i} - \frac{\partial V^j}{\partial x^i} u_j,$$
$$V_{ij} = \frac{\partial^2 \Phi}{\partial x^i \partial x^j} - \frac{\partial^2 V^k}{\partial x^i \partial x^j} u_k - \frac{\partial V^k}{\partial x^i} u_{jk} - \frac{\partial V^k}{\partial x^j} u_{ik}.$$

Moreover, the jet coordinates (u_t, u_i, u_{ij}) satisfy

$$(u_t, u_i, u_{ij}) = (\partial_t S, \partial_i S, \partial_{ij} S) = (-E_0 - \tfrac{1}{2}\Delta S, \partial_{\dot{x}^i} L_0, \partial_{ij} S),$$

where we recall $dS = d_{\dot{x}} L_0$ from Eq. (7.29) and Remark 7.23, and also that $\partial_t S = -H_0(dS, t) - \tfrac{1}{2}\Delta S = -E_0 - \tfrac{1}{2}\Delta S$. Plugging these into (7.69) and using the fact that $\partial_t H_0 = -\partial_t L_0$ and $\partial_{x^i} H_0 = -\partial_{x^i} L_0$ due to classical Legendre transform, we have

$$
\begin{aligned}
0 =& V^0 \partial_t H_0 + V^i \left(\partial_{x^i} H_0 + \tfrac{1}{2}\partial_i g^{jk} u_{jk} - \tfrac{1}{2}\partial_i g^{jk}\Gamma^l_{jk} u_l - \tfrac{1}{2}g^{jk}\partial_i \Gamma^l_{jk} u_l \right) \\
&+ \left(\partial_t \Phi - \dot{V}^0 u_t - \partial_t V^i u_i \right) \\
&+ \left(\partial_i \Phi - \partial_i V^l u_l \right) \left(\partial_{p_i} H_0 - \tfrac{1}{2}g^{jk}\Gamma^i_{jk} \right) \\
&+ \tfrac{1}{2}g^{ij} \left(\partial_i \partial_j \Phi - \partial_i \partial_j V^k u_k - \partial_i V^k u_{jk} - \partial_j V^k u_{ik} \right) \\
=& V^0 \partial_t H_0 + V^i \partial_{x^i} H_0 - \left(\partial_t + \partial_{p_i} H_0 \partial_j \right) V^i u_i \\
&- \tfrac{1}{2}g^{ij} \left(\partial_i \partial_j V^k - \Gamma^l_{ij}\partial_l V^k + 2\Gamma^k_{il}\partial_j V^l + \partial_l \Gamma^k_{ij} V^l \right) u_k \\
&- \dot{V}^0 u_t - g^{ij} \left(\partial_j V^k + \Gamma^k_{jm} V^m \right) \left(u_{ik} - \Gamma^l_{ik} u_l \right) \\
&+ \left[\partial_t \Phi + \left(\partial_{p_i} H_0 - \tfrac{1}{2}g^{jk}\Gamma^i_{jk} \right)\partial_i \Phi + \tfrac{1}{2}g^{ij}\partial_i \partial_j \Phi \right] \\
=& - V^0 \partial_t L_0 - V^i \partial_{x^i} L_0 - \left(\partial_t + \dot{x}^j \partial_j \right) V^i \partial_{\dot{x}^i} L_0 - \tfrac{1}{2}\left[\Delta \bar{V} + \mathrm{Ric}(\bar{V}) \right]^k \partial_{\dot{x}^k} L_0 \\
&+ \dot{V}^0 \left(E_0 + \tfrac{1}{2}\Delta S \right) - g^{ij}\nabla^2_{\partial_i, \nabla_{\partial_j} \bar{V}} S + \left(\partial_t \Phi + \dot{x}^i \partial_i \Phi + \tfrac{1}{2}\Delta \Phi \right) \\
=& - \left[V^0 \partial_t L_0 + d_x L_0(\bar{V}) + d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}}\bar{V}}{dt} \right) - \dot{V}^0 E_0 \right] \\
&+ \left(\tfrac{1}{2}\dot{V}^0 \Delta S - g^{ij}\nabla^2_{\partial_i, \nabla_{\partial_j}\bar{V}} S \right) + \mathbf{D}_t \Phi,
\end{aligned}
\tag{7.70}
$$

where, in the last equality, we used the fact that $(QX)^{ij}(t) = g^{ij}(X(t))$ to derive $\mathbf{D}_t \Phi$. The result then follows from Proposition 7.30. $\qquad\square$

Theorem 7.32 (Stochastic Noether's theorem) *Let $L_0 : \mathbb{R} \times TM \to \mathbb{R}$ be a hyper-regular Lagrangian. Suppose that the vector field V_Φ in (7.68) is an infinitesimal symmetry of the Hamilton–Jacobi–Bellman equation (7.49) associated with L_0 (with $f \equiv 0$). Then, the following stochastic conservation law holds for the stochastic*

Euler–Lagrange equation (7.22),

$$\mathbf{D}_t \left[V^i \partial_{\dot{x}^i} L_0 - V^0 E - \Phi \right] = 0. \tag{7.71}$$

Proof Recall that $dS = d_{\dot{x}} L_0$ and $\partial_t S = -E_0 - \frac{1}{2}\Delta S = -E$. By applying Lemma 7.8.(iv) and (7.22), as well as the fact that $(QX)^{ij}(t) = g^{ij}(X(t))$, we have

$$\mathbf{D}_t \left[d_{\dot{x}} L_0(\bar{V}) \right] = d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}} \bar{V}}{dt} \right) + \frac{\overline{\mathbf{D}}(d_{\dot{x}} L_0)}{dt}(\bar{V}) + (QX)^{ij} (\nabla_{\partial_i} (d_{\dot{x}} L_0))(\nabla_{\partial_j} \bar{V})$$

$$= d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}} \bar{V}}{dt} \right) + d_x L_0(\bar{V}) + g^{ij} \nabla^2_{\partial_i, \nabla_{\partial_j} \bar{V}} S.$$

Then, we use HJB equation (7.49) (with $f \equiv 0$) and the classical Legendre transform $H_0 = d_{\dot{x}} L_0 \cdot \dot{x} - L_0$ to derive

$$\mathbf{D}_t E = -\mathbf{D}_t \partial_t S = -\partial_t \left(\partial_t + \nabla_{\dot{x}} + \frac{1}{2}\Delta \right) S = -\partial_t \left[dS \cdot \dot{x} + \left(\partial_t + \frac{1}{2}\Delta \right) S \right]$$

$$= -\partial_t \left(d_{\dot{x}} L_0 \cdot \dot{x} - H_0 \right) = -\partial_t L_0.$$

Combining these with the S-EL equation (7.22) and the criterion (7.70) for symmetries of the HJB equation (7.27), we have

$$\mathbf{D}_t \left[V^i \partial_{\dot{x}^i} L_0 - V^0 E - \Phi \right] = \mathbf{D}_t \left[d_{\dot{x}} L_0(\bar{V}) \right] - \dot{V}^0 E - V^0 \mathbf{D}_t E - \mathbf{D}_t \Phi$$

$$= d_{\dot{x}} L_0 \left(\frac{\overline{\mathbf{D}} \bar{V}}{dt} \right) + d_x L_0(\bar{V}) + g^{ij} \nabla^2_{\partial_i, \nabla_{\partial_j} \bar{V}} S$$

$$- \dot{V}^0 \left(E_0 + \frac{1}{2}\Delta S \right) + V^0 \partial_t L_0 - \mathbf{D}_t \Phi$$

$$= 0.$$

The result follows.   □

Remark 7.33 (i) In stochastic Hamiltonian formalism, (7.71) reads as $\mathbf{D}_t \left[V^i p_i - V^0 H - \Phi \right] = 0$.

(ii) The stochastic conservation law (6.19) of a time-independent g-canonical second-order Hamiltonian $H = \overline{H}_0^g$ can be regarded as a special case of the above stochastic Noether's theorem. Indeed, consider the infinitesimal unit time translation $V = \frac{\partial}{\partial t}$, i.e., $V^0 = 1$, $\bar{V} = 0$, $\Phi = 0$. Then, the criterion (7.70) reduces to $0 = \partial_t L_0 = -\partial_t H_0$, which means that $H = \overline{H}_0^g$ is time-independent. The resulting stochastic conservation law is $\mathbf{D}_t E = \mathbf{D}_t H = 0$.

Applying stochastic Noether's theorem to Schrödinger's problem of Sect. 7.3, we have the following corollary. Its Euclidean case with zero vector potential (i.e., $b \equiv 0$) has already been formulated in Thieullen and Zambrini (1997).

Corollary 7.34 (Stochastic Noether's theorem for Schrödinger's problem) *Let L_0 be the Lagrangian given in* (7.25). *Suppose that the vector field V_Φ in* (7.68) *is an infinitesimal symmetry of Hamilton–Jacobi–Bellman equation* (7.27) *with $f \equiv 0$. Then, the following stochastic conservation law holds for the coordinate process of the solution of Schrödinger's problem in* (7.33),

$$\mathbf{D}_t \left[g_{ij} \left(D^j_\nabla x - b^j \right) V^i - V^0 \left(E_0 + \tfrac{1}{2} \Delta S \right) - \Phi \right] = 0,$$

where E_0 is the classical energy given in (7.45) *and S is the solution of* (7.27).

Data Availability   Our manuscript has no associated data.

Appendix A Mixed-Order Tangent and Cotangent Bundles

A.1 Mixed-Order Tangent and Cotangent Maps

Clearly, the mixed-order tangent bundle $T\mathbb{R} \times \mathcal{T}^S M$ is a subbundle of the totally second-order tangent bundle $\mathcal{T}^S(\mathbb{R} \times M)$, and contains the tangent bundle $T(\mathbb{R} \times M) \cong T\mathbb{R} \times TM$ as a subbundle. Similar properties hold for the mixed-order cotangent bundle.

It is easy to verify that the mixed-order tangent bundle can be characterized as follows:

$$T\mathbb{R} \times \mathcal{T}^S M = \{ A \in \mathcal{T}^S(\mathbb{R} \times M) : \pi^S_*(A) \in T\mathbb{R} \}.$$

We also define the stochastic analog of the vertical bundle as

$$V^S \pi = \{ A \in T\mathbb{R} \times \mathcal{T}^S M : \pi^S_*(A) = 0 \}.$$

Then, it is easy to see that $V^S \pi \cong \mathbb{R} \times \mathcal{T}^S M$.

Given a smooth map $F : \mathbb{R} \times M \to \mathbb{R} \times N$, we can define its second-order pushforward F^S_* as in Definitions 5.5 and 5.7, so that F^S_* is a bundle homomorphism

from $\tau^S_{\mathbb{R} \times M}$ to $\tau^S_{\mathbb{R} \times N}$. In general, F^S_* neither maps the mixed-order tangent bundle to the mixed-order tangent bundle, nor maps the vertical bundle to the vertical bundle. But if F is projectable, then it does.

Lemma A.1 *Let M and N be two smooth manifolds and M be connected. Let F : $\mathbb{R} \times M \to \mathbb{R} \times N$ be a smooth map. Then, the following statements are equivalent:*

(i) *F is a bundle homomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$;*
(ii) *$F^S_*(T\mathbb{R} \times T^S M) \subset T\mathbb{R} \times T^S N$;*
(iii) *$F^S_*(V^S \pi) \subset V^S \rho$.*

Proof We first prove that (i) implies both (ii) and (iii). Suppose that F is a bundle homomorphism projecting to F^0. Then, $\rho \circ F = F^0 \circ \pi$ and hence, for any $A \in T^S(\mathbb{R} \times M)$,

$$\rho^S_*(F^S_*(A)) = (F^0)^S_* \pi^S_*(A).$$

If $A \in T\mathbb{R} \times T^S M$, then $\pi^S_*(A) \in T\mathbb{R}$ and thus $\rho^S_*(F^S_*(A)) \in (F^0)^S_*(T\mathbb{R}) = (F^0)_*(T\mathbb{R}) \subset T\mathbb{R}$. This implies $F^S_*(A) \in T\mathbb{R} \times T^S N$. If $A \in V^S\pi$, then $\pi^S_*(A) = 0$, it follows $\rho^S_*(F^S_*(A)) = 0$ and therefore $F^S_*(A) \in V^S \rho$.

Next we prove either (ii) or (iii) implies (i). Choose local coordinates (t, x^i) around $(t_0, q) \in \mathbb{R} \times M$ and (s, y^j) around $F(t_0, q)$. Suppose F has a local expression $F = (F^0, \bar{F}^j)$. Let $A \in T\mathbb{R} \times T^S M|_{(t_0, q)}$ having the following local expression:

$$A = A^0 \frac{\partial}{\partial t}\Big|_{t_0} + A^i \frac{\partial}{\partial x^i}\Big|_q + A^{jk} \frac{\partial^2}{\partial x^j \partial x^k}\Big|_q. \qquad (A.1)$$

Then, Lemma 5.6 yields

$$F^S_* A = (AF^0) \frac{\partial}{\partial s}\Big|_{F^0(t_0, q)} + (A\bar{F}^i) \frac{\partial}{\partial y^i}\Big|_{\bar{F}(t_0, q)} + \Gamma_A(F^0, F^0) \frac{\partial^2}{\partial s^2}\Big|_{F^0(t_0, q)}$$
$$+ \Gamma_A(\bar{F}^i, \bar{F}^j) \frac{\partial^2}{\partial y^i \partial y^j}\Big|_{\bar{F}(t_0, q)} + 2\Gamma_A(F^0, \bar{F}^i) \frac{\partial^2}{\partial s \partial y^i}\Big|_{F(t_0, q)}.$$

If (ii) holds, then $F^S_*(A) \in T\mathbb{R} \times T^S N|_{F(t_0, q)}$. It then follows

$$\Gamma_A(F^0, F^0) = A^{jk} \frac{\partial F^0}{\partial x^j} \frac{\partial F^0}{\partial x^k} = 0 \quad \text{and} \quad \Gamma_A(F^0, \bar{F}^i) = A^{jk} \frac{\partial F^0}{\partial x^j} \frac{\partial \bar{F}^i}{\partial x^k} = 0. (A.2)$$

Since A is arbitrary, we know that $\frac{\partial F^0}{\partial x^i} = 0$ for all i. Then, by the connectness of M, F^0 is independent of $q \in M$. This implies that F is a bundle homomorphism. Now assume that $A \in V^S M|_{(t_0, q)}$ has a local expression in (A.1) with $A^0 = 0$. If (iii) holds, then $F^S_*(A) \in V^S N|_{F(t_0, q)}$. This amounts to (A.2) together with

$$AF^0 = A^i \frac{\partial F^0}{\partial x^i} + A^{jk} \frac{\partial^2 F^0}{\partial x^j \partial x^k} = 0.$$

Again, the arbitrariness of A yields that $\frac{\partial F^0}{\partial x^i} = 0$ for all i. Thus, F is a bundle homomorphism. $\square$

It is easy to deduce from the proof that if $F = (F^0, \bar{F})$ is a bundle homomorphism from π to ρ, then $F_*^S|_{T\mathbb{R} \times \mathcal{T}^S M}$ is a bundle homomorphism from $\tau_R \times \tau_M^S$ to $\tau_\mathbb{R} \times \tau_N^S$.

When $F : \mathbb{R} \times M \to \mathbb{R} \times N$ is a diffeomorphism, we can also consider the second-order pullback map F^{S*} which is a bundle homomorphism from $\tau_{\mathbb{R}\times M}^{S*}$ to $\tau_{\mathbb{R}\times N}^{S*}$. But when we restrict F^{S*} to the mixed-order cotangent bundle $T^*\mathbb{R} \times \mathcal{T}^S M$, there are difficulties. We can check that even if F is a bundle homomorphism, F^{S*} does not necessarily map $T^*\mathbb{R} \times \mathcal{T}^S M$ into $T^*\mathbb{R} \times \mathcal{T}^{S*} M$. The reason is basically that the restrictions of second-order pullbacks to the cotangent bundle do not coincide with usual pullbacks. To overcome this, we consider the dual map of $F_*^S|_{T\mathbb{R} \times \mathcal{T}^S M}$. This motivates the following definition, which contrasts with Definitions 5.5 and 5.7.

Definition A.2 (*Mixed-order pushforward and pullback*) Let F be a bundle homomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$. The mixed-order tangent map of F at $(t, q) \in \mathbb{R} \times M$ is the linear map $d^\circ F_{(t,q)} : T\mathbb{R} \times \mathcal{T}^S M|_{(t,q)} \to T\mathbb{R} \times \mathcal{T}^S N|_{F(t,q)}$ defined by

$$d^\circ F_{(t,q)} = d^2 F_{(t,q)}|_{T_t\mathbb{R} \times \mathcal{T}_q^S M}.$$

The mixed-order cotangent map of F at $(t, q) \in \mathbb{R} \times M$ is the linear map $d^\circ F_{(t,q)}^* : T^*\mathbb{R} \times \mathcal{T}^{S*} N|_{F(t,q)} \to T^*\mathbb{R} \times \mathcal{T}^{S*} M|_{(t,q)}$ dual to $d^\circ F_{(t,q)}$, that is,

$$d^\circ F_{(t,q)}^*(\alpha)(A) = \alpha(d^\circ F_{(t,q)}(A)), \quad \text{for } A \in T_t\mathbb{R} \times \mathcal{T}_q^S M, \alpha \in T^*\mathbb{R} \times \mathcal{T}^{S*} N|_{F(t,q)}.$$

The mixed-order pushforward by F is the bundle homomorphism $F_*^R : (T\mathbb{R} \times \mathcal{T}^S M, \tau_\mathbb{R} \times \tau_M^S, \mathbb{R} \times M) \to (T\mathbb{R} \times \mathcal{T}^S N, \tau_\mathbb{R} \times \tau_N^S, \mathbb{R} \times N)$ defined by

$$F_*^R|_{T_t\mathbb{R} \times \mathcal{T}_q^S M} = d^\circ F_{(t,q)}.$$

Given a mixed-order form α on $\mathbb{R} \times N$, the mixed-order pullback of α by F is the mixed-order form $F^{R*}\alpha$ on $\mathbb{R} \times M$ defined by

$$(F^{R*}\alpha)_{(t,q)} = d^\circ F_{(t,q)}^*\left(\alpha_{F(t,q)}\right), \quad (t, q) \in \mathbb{R} \times M.$$

If, moreover, F is a bundle isomorphism, then the mixed-order pullback by F is the bundle isomorphism $F^{R*} : (T\mathbb{R} \times \mathcal{T}^{R*} N, \tau_\mathbb{R} \times \tau_N^{S*}, \mathbb{R} \times N) \to (T\mathbb{R} \times \mathcal{T}^{S*} M, \tau_\mathbb{R} \times \tau_M^{S*}, \mathbb{R} \times M)$ defined by

$$F^{R*}|_{\mathcal{T}_s\mathbb{R} \times \mathcal{T}_{q'}^{S*} N} = d^\circ F_{F^{-1}(s,q')}^*.$$

Given a mixed-order vector field A on $\mathbb{R} \times M$, the mixed-order pushforward of A by F is the mixed-order vector field $F_*^R A$ on $\mathbb{R} \times N$ defined by

$$(F_*^R A)_{(s,q')} = d^\circ F_{F^{-1}(s,q')}\left(A_{F^{-1}(s,q')}\right), \quad (s, q') \in \mathbb{R} \times N.$$

Clearly, the mixed-order pushforward F_*^R is nothing but $F_*^S|_{T\mathbb{R} \times \mathcal{T}^S M}$. Write $F = (F^0, \bar{F})$. Then, in local coordinates, F_*^R acts on A of (A.1) as follows:

$$
\begin{aligned}
F_*^R A &= A^0 \frac{dF^0}{dt}(t_0) \frac{\partial}{\partial s}\Big|_{F^0(t_0)} + (A\bar{F}^i)(t_0, q) \frac{\partial}{\partial y^i}\Big|_{\bar{F}(t_0,q)} \\
&\quad + A^{kl} \frac{\partial \bar{F}^i}{\partial x^k} \frac{\partial \bar{F}^j}{\partial x^l}(t_0, q) \frac{\partial^2}{\partial y^i \partial y^j}\Big|_{\bar{F}(t_0,q)} \\
&= A^0 \frac{dF^0}{dt}(t_0) \frac{\partial}{\partial s}\Big|_{F^0(t_0)} + \Big[A^0 \frac{\partial \bar{F}^i}{\partial t}(t_0, q) + A^j \frac{\partial \bar{F}^i}{\partial x^j}(t_0, q) \qquad \text{(A.3)} \\
&\quad + A^{jk} \frac{\partial^2 \bar{F}^i}{\partial x^j \partial x^k}(t_0, q)\Big] \frac{\partial}{\partial y^i}\Big|_{\bar{F}(t_0,q)} \\
&\quad + A^{kl} \frac{\partial \bar{F}^i}{\partial x^k} \frac{\partial \bar{F}^j}{\partial x^l}(t_0, q) \frac{\partial^2}{\partial y^i \partial y^j}\Big|_{\bar{F}(t_0,q)}.
\end{aligned}
$$

And F^{R*} acts on the mixed-order cotangent vector $\alpha = \alpha_0 ds|_{F^0(t_0)} + \alpha_i d^2 y^i|_{\bar{F}(t_0,q)} + \alpha_{ij} dy^i \cdot dy^j|_{\bar{F}(t_0,q)} \in T\mathbb{R} \times \mathcal{T}^{S*}N|_{F(t_0,q)}$ by

$$
\begin{aligned}
F^{R*}\alpha &= \alpha_0 \frac{dF^0}{dt}(t_0) dt|_{t_0} + \alpha_i d^\circ F^i|_{(t_0,q)} + \alpha_{ij} \frac{\partial \bar{F}^i}{\partial x^k} \frac{d\bar{F}^i}{dx^l}(t_0, q) dx^k \cdot dx^l|_q \\
&= \Big[\alpha_0 \frac{dF^0}{dt}(t_0) + \alpha_i \frac{\partial \bar{F}^i}{\partial t}(t_0, q)\Big] dt|_{t_0} + \alpha_i \frac{\partial \bar{F}^i}{\partial x^j}(t_0, q) d^2 x^j|_q \qquad \text{(A.4)} \\
&\quad + \Big[\frac{\alpha_i}{2} \frac{\partial^2 \bar{F}^i}{\partial x^k \partial x^l}(t_0, q) + \alpha_{ij} \frac{\partial \bar{F}^i}{\partial x^k} \frac{d\bar{F}^j}{dx^l}(t_0, q)\Big] dx^k \cdot dx^l|_q.
\end{aligned}
$$

By virtue of these local expressions, one easily deduce that

$$
\begin{aligned}
F_*^R|_{T_t\mathbb{R} \times \mathcal{T}_q^S M} &= F(q)_*|_{T_t\mathbb{R}} \times \bar{F}(t)_*^S|_{\mathcal{T}_q^S M}, \\
F^{R*}|_{T_s^*\mathbb{R} \times \mathcal{T}_{q'}^{S*}N} &= F(q)^*|_{T_s^*\mathbb{R}} \times \bar{F}(t)^{S*}|_{\mathcal{T}_{q'}^{S*}N}.
\end{aligned}
$$

And in turn, these verify the linearity of F_*^R and F^{R*}. The following property is easy to check.

Lemma A.3 *Let F be a bundle isomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ and A be a mixed-order vector field. Let f be a smooth functions on $\mathbb{R} \times N$. Then, $((F_*^R A)f) \circ F = A(f \circ F)$.*

A.2 Pushforwards of Generators

A smooth map $F : M \to N$ can be associated naturally with a bundle homomorphism $\mathbf{Id}_{\mathbb{R}} \times F : (\mathbb{R} \times M, \pi, \mathbb{R}) \to (\mathbb{R} \times N, \rho, \mathbb{R})$ that projects to the identity on $\mathbb{R}$. In this

case, the pushforward of a diffusion X by $\mathbf{Id}_{\mathbb{R}} \times F$ is just $(\mathbf{Id}_{\mathbb{R}} \times F) \cdot X = F(X)$. The stochastic prolongations of the bundle homomorphism $\mathbf{Id}_{\mathbb{R}} \times F$ is then

$$j(\mathbf{Id}_{\mathbb{R}} \times F)(j_{(t,q)}X) = j_{(t,F(q))}(F(X)).$$

Corollary A.4 *Let $F : M \to N$ be a diffeomorphism. If a diffusion X on M has a generator $A = (A_t)$, then the process $F(X)$ is a diffusion on N, with generator $F_*^S A = (F_*^S A_t)$.*

Proof Assume $X \in I_{t_0}(M)$. For every $f \in C^\infty(N)$, $f \circ F \in C^\infty(M)$, by the assumption, we have

$$f \circ F(X(t)) - f \circ F(X(t_0)) - \int_{t_0}^t A_s(f \circ F)(X(s))\mathrm{d}s$$

$$= f(F(X(t))) - f(F(X(t_0))) - \int_{t_0}^t \left((F_*^S A_s)f\right)(F(X(s)))\mathrm{d}s$$

is a real-valued continuous $\{\mathcal{P}_t\}$-martingale. This proves that $F(X) \in I_{t_0}(N)$ has generator $F_*^S A$. □

This corollary together with the identification between $\mathbb{R} \times \mathcal{T}^S M$ and $\mathbb{R} \times \mathcal{T}^E M$ in (3.6) and (3.7), give rise to the relation between prolongations and pushforwards as follows:

$$j(\mathbf{Id}_{\mathbb{R}} \times F)(t, A_q) = j(\mathbf{Id}_{\mathbb{R}} \times F)\left(j_{(t,q)}X^A\right) = j_{(t,F(q))}(F \circ X^A) = \left(t, \left(F_*^S A_t\right)_{F(q)}\right)$$

$$= \left(t, d^2 F_q(A_{(t,q)})\right) = \left(t, d^2 F_q(A_q)\right) = (\mathbf{Id}_{\mathbb{R}} \times F_*^S)(t, A_q),$$

so that $j(\mathbf{Id}_{\mathbb{R}} \times F) = \mathbf{Id}_{\mathbb{R}} \times F_*^S$.

The following corollary is an extension of Corollary A.4 and a straightforward consequence of Lemma 4.8. Here, we will present another proof, using notions of "Appendix A.1."

Corollary A.5 *Let F be a bundle isomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to F^0. If X is a diffusion on M with respect to $\{\mathcal{P}_t\}$ and has a extended generator $\frac{\partial}{\partial t} + A$ where A is a time-dependent second-order vector field, then the pushforward $F \cdot X$ is a diffusion on N with respect to $\{\mathcal{F}_{(F^0)^{-1}(s)}\}$, with extended generator*

$$\frac{\mathrm{d}(F^0)^{-1}}{\mathrm{d}s} F_*^R \left(\frac{\partial}{\partial t} + A\right).$$

Proof Assume that $X \in I_{t_0}(M)$ and $F = (F^0, \bar{F})$. For every $f \in C^\infty(\mathbb{R} \times N)$, Lemma A.3 yields that the process

$$f \circ F(t, X(t)) - f \circ F(t_0, X(t_0)) - \int_{t_0}^t \left(\frac{\partial}{\partial t} + A \right) (f \circ F)(u, X(u)) du$$

$$= f\left(F^0(t), \bar{F}(t, X(t)) \right) - f\left(F^0(t_0), \bar{F}(t_0, X(t_0)) \right)$$

$$- \int_{t_0}^t F_*^R \left(\frac{\partial}{\partial t} + A \right) f\left(F^0(u), \bar{F}(u, X(u)) \right) du$$

is a continuous $\{\mathcal{P}_t\}$-martingale. Denote $s_0 = F^0(t_0)$. By substituting $t = (F^0)^{-1}(s)$ which can be done because F^0 is an isomorphism, and using the change of variable $u = (F^0)^{-1}(v)$, and recalling that $F \cdot X(s) = \bar{F}\left((F^0)^{-1}(s), X((F^0)^{-1}(s)) \right)$, the process

$$f(s, F \cdot X(s)) - f(s_0, F \cdot X(s_0))$$

$$- \int_{(F^0)^{-1}(s_0)}^{(F^0)^{-1}(s)} F_*^R \left(\frac{\partial}{\partial t} + A \right) f\left(F^0(u), \bar{F}(u, X(u)) \right) du$$

$$= f(s, F \cdot X(s)) - f(s_0, F \cdot X(s_0))$$

$$- \int_{s_0}^s \frac{d(F^0)^{-1}}{ds}(v) F_*^R \left(\frac{\partial}{\partial t} + A \right) f(v, F \cdot X(v)) dv$$

is a continuous $\{\mathcal{F}_{(F^0)^{-1}(s)}\}$-martingale. The result follows. □

Remark A.6 (i) As a consequence, the generator of the pushforward $F \cdot X$ is given in local coordinates by

$$\frac{d(F^0)^{-1}}{ds} \left[\left(\frac{\partial}{\partial t} + A \right) \bar{F}^i \circ F^{-1} \right] \frac{\partial}{\partial y^i} + \frac{d(F^0)^{-1}}{ds} \left[\left(A^{kl} \frac{\partial \bar{F}^i}{\partial x^k} \frac{\partial \bar{F}^j}{\partial x^l} \right) \circ F^{-1} \right] \frac{\partial^2}{\partial y^i \partial y^j}.$$

This coincides with Lemma 4.8.

(ii) This corollary together with Lemma A.1 indicates that the bundle homomorphisms from $\mathbb{R} \times M$ to $\mathbb{R} \times N$ are the only (deterministic) smooth maps between them that map diffusions to diffusions. Indeed, if a smooth map F from $\mathbb{R} \times M$ to $\mathbb{R} \times N$ pushes forward a diffusion to another diffusion, then a similar argument as in Corollary A.5 implies that F_*^S would map the extended generator of the former diffusion to that of the latter, whereas Lemma A.1 says such F_*^S must be the second-order pushforward of some bundle homomorphism.

(iii) In particular, if F is a smooth map from M to N and X is a diffusion on M with generator A, then $F(X)$ is a diffusion on N with respect to the same filtration, with generator $F_*^S(A)$.

A.3 Pushforwards and Pullbacks by Diffusions

Definition A.7 (*Pushforwards and pullbacks by diffusions*) Let X be an M-valued diffusion process. Let $(\mathbb{R} \times U, (t, x^i))$ be a coordinate chart on $\mathbb{R} \times M$. The pushforward map X_* from $T_t \mathbb{R}$ to $T_t \mathbb{R} \times \mathcal{T}^S_{X(t)} M$ is defined in the local coordinate by

$$
X_* \left(\tau \frac{\mathrm{d}}{\mathrm{d}t} \Big|_{t_0} \right) = \tau \left(\frac{\partial}{\partial t} \Big|_{t_0} + (DX)^i(t_0) \frac{\partial}{\partial x^i} \Big|_{X(t_0)} + \frac{1}{2}(QX)^{jk}(t_0) \frac{\partial^2}{\partial x^j \partial x^k} \Big|_{X(t_0)} \right).
\tag{A.5}
$$

The pullback map X^* from $\mathcal{T}^*_t \mathbb{R} \times \mathcal{T}^{S*}_{X(t)} M$ to $\mathcal{T}^*_t \mathbb{R}$ is defined by

$$
X^* \left(\alpha_0 dt|_{t_0} + \alpha_i d^2 x^i|_{X(t_0)} + \tfrac{1}{2}\alpha_{jk} dx^j \cdot dx^k|_{X(t_0)} \right)
$$
$$
= \left(\alpha_0 + \alpha_i (DX)^i(t_0) + \tfrac{1}{2}\alpha_{jk}(QX)^{jk}(t_0) \right) dt|_{t_0}.
\tag{A.6}
$$

Remark A.8 Recall that in classical differential geometry, the pushforward by a smooth curve $\gamma = (\gamma(t))_{t \in [-1,1]}$ on M is a map $\gamma_* : T\mathbb{R} \to TM$ given by $\gamma_*(\frac{\mathrm{d}}{\mathrm{d}t}|_{t_0}) = \dot{\gamma}^i(t_0) \frac{\partial}{\partial x^i}|_{\gamma(t_0)}$. While if we look at the graph of γ as a section of the trivial bundle $(\mathbb{R} \times M, \pi, \mathbb{R})$, denoted by $\bar{\gamma}$, then the pushforward map by $\bar{\gamma}$ is $\bar{\gamma}_*(\frac{\mathrm{d}}{\mathrm{d}t}|_{t_0}) = \frac{\mathrm{d}}{\mathrm{d}t}|_{t_0} + \dot{\gamma}^i(t_0) \frac{\partial}{\partial x^i}|_{\gamma(t_0)}$. For this reason, it would be more appropriate to call X_* and X^* in Definition A.7 the pushforward and pullback by graph of X, or by random section corresponding to X, instead of by X itself. But we avoid that for simplicity.

One can see from the definition that the pushforward X_* maps the time vector $\frac{\mathrm{d}}{\mathrm{d}t}|_{t_0}$ to the value of the extended generator of X at $(t_0, X(t_0))$. There is an informal way to look at the pullback map X^*: one first replace all x's by X's in the brackets at LHS of (A.6) and obtain

$$
\alpha_0 dt + \alpha_i dX^i + \tfrac{1}{2}\alpha_{jk} dX^j \cdot dX^k;
$$

then substituting dX^i and $dX^j \cdot dX^k$, and following Itô's calculus,

$$
dX^i = (DX)^i dt + \text{martingale part}, \qquad dX^j \cdot dX^k = (QX)^{jk} dt,
$$

and getting rid of the martingale part, we get the RHS of (A.6).

The following corollary is straightforward. We will see that pushforward and pull-back maps by diffusions are also closely related to the concept of "total derivatives."

Corollary A.9 (i) *Let X be an M-valued diffusion process. For all $\tau \frac{\mathrm{d}}{\mathrm{d}t}|_{t_0} \in \mathcal{T}_{t_0} \mathbb{R}$ and $\alpha \in \mathcal{T}^*_{t_0} \mathbb{R} \times \mathcal{T}^{S*}_{X(t_0)} M$,*

$$
\left\langle X^*(\alpha), \tau \tfrac{\mathrm{d}}{\mathrm{d}t}|_{t_0} \right\rangle = \left\langle \alpha, X_*(\tau \tfrac{\mathrm{d}}{\mathrm{d}t}|_{t_0}) \right\rangle.
\tag{A.7}
$$

(ii) *If $X \in I_{(t_0,q)}(M)$, f is a smooth function on $\mathbb{R} \times M$ and g a smooth function on M, then*

$$
\begin{aligned}
\langle X^*(d^\circ f), \tfrac{d}{dt} \rangle \big|_{t_0} &= X_*(\tfrac{d}{dt})(f) \big|_{(t_0,q)} \\
&= (\mathbf{D}_t f)(j_{(t_0,q)} X) = \langle \tfrac{\partial}{\partial t} + A^X, d^\circ f \rangle (t_0, q), \\
\langle X^*(dg \cdot dg), \tfrac{d}{dt} \rangle \big|_{t_0} &= \langle dg \cdot dg, X_*(\tfrac{d}{dt}) \rangle \big|_{(t_0,q)} = (\mathbf{Q}_t g)(j_{(t_0,q)} X).
\end{aligned}
$$

(iii) *Let X, Y be M-valued diffusion processes satisfying $X(t) = Y(t)$ a.s.. Then, $j_t X = j_t Y$ a.s. if and only if $X_*(\tfrac{d}{dt}|_t) = Y_*(\tfrac{d}{dt}|_t)$ a.s.. In particular, if $X, Y \in I_{(t,q)}(M)$, then $j_{(t,q)} X = j_{(t,q)} Y$ if and only if $X_*(\tfrac{d}{dt}|_t) = Y_*(\tfrac{d}{dt}|_t)$.*

(iv) *Let F be a bundle homomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ projecting to F^0, and X be an M-valued diffusion process. Then, $F_*^R \circ X_* = (F \cdot X)_* \circ (F^0)_*$.*

(v) *Let F be a smooth function from M to M, and X be an M-valued diffusion process. Then, $(\mathbf{Id}_{T\mathbb{R}} \times F_*^S) \circ X_* = (F \circ X)_*$.*

Proof Assertions (i), (ii) and (iii) are easy to deduce from the definitions. We prove (iv) using local expressions. Assume that $F = (F^0, \bar{F})$ and denote $\tilde{X} = F \cdot X$. Recall that $\tilde{X}(F^0(t)) = \bar{F}(t, X(t))$. Then,

$$
\begin{aligned}
F_*^R \circ X_* \left(\frac{d}{dt} \Big|_t \right) =& \frac{dF^0}{dt}(t) \frac{\partial}{\partial s} \Big|_{F^0(t)} + \left[\frac{\partial \bar{F}^i}{\partial t}(t, X(t)) + (DX)^j(t) \frac{\partial \bar{F}^i}{\partial x^j}(t, X(t)) \right. \\
&+ \frac{1}{2}(QX)^{jk}(t) \frac{\partial^2 \bar{F}^i}{\partial x^j \partial x^k}(t, X(t)) \Big] \frac{\partial}{\partial y^i} \Big|_{\bar{F}(t,X(t))} \\
&+ \frac{1}{2}(QX)^{kl}(t) \frac{\partial \bar{F}^i}{\partial x^k} \frac{\partial \bar{F}^j}{\partial x^l}(t, X(t)) \frac{\partial^2}{\partial y^i \partial y^j} \Big|_{\bar{F}(t,X(t))} \\
=& \frac{dF^0}{dt}(t) \left[\frac{\partial}{\partial s} \Big|_{F^0(t)} + (D\tilde{X})^i(F^0(t)) \frac{\partial}{\partial y^i} \Big|_{\tilde{X}(F^0(t))} \right. \\
&+ \frac{1}{2}(Q\tilde{X})^{ij}(F^0(t)) \frac{\partial^2}{\partial y^i \partial y^j} \Big|_{\tilde{X}(F^0(t))} \Big] \\
=& \frac{dF^0}{dt}(t)(F \cdot X)_* \left(\frac{\partial}{\partial s} \Big|_{F^0(t)} \right) \\
=& (F \cdot X)_* \circ (F^0)_* \left(\frac{d}{dt} \Big|_t \right).
\end{aligned}
$$

The result follows. $\qquad \square$

Lie Derivatives

Definition A.10 (*Lie derivatives*) Let V be a vector field on M and $\psi = \{\psi_\epsilon\}_{\epsilon \in \mathbb{R}}$ be its flow. Let A be a second-order vector field and α be a second-order form on M. The Lie derivative of A with respect to V is a second-order vector field on M, denoted by

$\mathcal{L}_V A$, and defined by

$$(\mathcal{L}_V A)_q = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} (\psi_{-\epsilon})^S_* (A_{\psi_\epsilon(q)}) = \lim_{\epsilon\to 0} \frac{(\psi_{-\epsilon})^S_* (A_{\psi_\epsilon(q)}) - A_q}{\epsilon}.$$

The Lie derivative of α with respect to V is a second-order form on M, denoted by $\mathcal{L}_V \alpha$, and defined by

$$(\mathcal{L}_V \alpha)_q = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} (\psi_\epsilon)^{S*} (\alpha_{\psi_\epsilon(q)}) = \lim_{\epsilon\to 0} \frac{(\psi_\epsilon)^{S*} (\alpha_{\psi_\epsilon(q)}) - \alpha_q}{\epsilon}.$$

For sufficient small $\epsilon \neq 0$, ψ_ϵ is defined in a neighborhood of $q \in M$ and $\psi_{-\epsilon}$ is the inverse of ψ_ϵ. So the difference quotients in the above definitions of Lie derivatives make sense. It is easy to verify that the derivatives exist for each $q \in M$, and $\mathcal{L}_V A$ is a smooth second-order vector field, $\mathcal{L}_V \alpha$ is a smooth second-order covector field. Likewise, the restrictions of $\mathcal{L}_V$ to $\mathcal{T}_q M$ and $\mathcal{T}^*_{F(q)} N$ coincide with the classical Lie derivatives. In the following, we seek properties of $\mathcal{L}$. Some of them can be found in Meyer (1981b, Section 6.(d)).

Lemma A.11 *Let V be a vector field and f be a smooth function. Let A and α be a second-order vector field and second-order form, respectively. Then,*

(i) *$\mathcal{L}_V A = [V, A]$, where the RHS denotes the commutator of V and A as linear operators;*

(ii) *$\mathcal{L}_V(fA) = (Vf)A + f\mathcal{L}_V A$;*

(iii) *$\langle \mathcal{L}_V \alpha, A \rangle = V(\langle \alpha, A \rangle) - \langle \alpha, \mathcal{L}_V A \rangle$;*

(iv) *$\mathcal{L}_V(f\alpha) = (Vf)\alpha + f\mathcal{L}_V \alpha$;*

(v) *$\mathcal{L}_V(d^2 f) = d^2(Vf)$.*

Remark A.12 Note that the commutator $[V, A]$ is a second-order vector field. Indeed, if V and A have coordinate expressions $V = V^i \frac{\partial}{\partial x^i}$ and $A = A^i \frac{\partial}{\partial x^i} + A^{ij} \frac{\partial^2}{\partial x^i \partial x^j}$, then the following local expression for $[V, A]$ is easy to verify:

$$[V, A] = \left(V^j \frac{\partial A^i}{\partial x^j} - A^j \frac{\partial V^i}{\partial x^j} - A^{jk} \frac{\partial^2 V^i}{\partial x^j \partial x^k} \right) \frac{\partial}{\partial x^i} + V^i \frac{\partial A^{jk}}{\partial x^i} \frac{\partial^2}{\partial x^j \partial x^k}$$
$$- A^{jk} \left(\frac{\partial V^i}{\partial x^j} \frac{\partial^2}{\partial x^i \partial x^k} + \frac{\partial V^i}{\partial x^k} \frac{\partial^2}{\partial x^i \partial x^j} \right).$$

Proof (i) For a function $f \in C^\infty(M)$,

$$(\mathcal{L}_V A)_q f = \lim_{\epsilon\to 0} \frac{(\psi_{-\epsilon})^S_* (A_{\psi_\epsilon(q)}) f - A_q f}{\epsilon} = \lim_{\epsilon\to 0} \frac{(A_{\psi_\epsilon(q)})(f \circ \psi_{-\epsilon}) - A_q f}{\epsilon}$$
$$= \lim_{\epsilon\to 0} \frac{(A_{\psi_\epsilon(q)})(f \circ \psi_{-\epsilon} - f)}{\epsilon} + \lim_{\epsilon\to 0} \frac{(A_{\psi_\epsilon(q)}) f - A_q f}{\epsilon}.$$

Then, a similar argument to the derivation of classical Lie derivatives yields

$$(\mathcal{L}_V A)_q f = -A_q(Vf) + V_q(Af) = [V, A]_q f.$$

(ii) $\mathcal{L}_V(fA)g = [V, fA]g = V(fAg) - fAVg = VfAg + fVAg - fAVg = VfAg + f(\mathcal{L}_V A)g$.

(iii) For a second-order vector field A,

$$
\begin{aligned}
\langle \mathcal{L}_V \alpha, A \rangle &= \lim_{\epsilon \to 0} \frac{\langle (\psi_\epsilon)^{S*}(\alpha_{\psi_\epsilon(q)}), A \rangle - \langle \alpha_q, A \rangle}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{\langle \alpha_{\psi_\epsilon(q)}, (\psi_\epsilon)^S_* A \rangle - \langle \alpha_q, A \rangle}{\epsilon} \\
&= \lim_{t \to 0} \frac{\langle \alpha_{\psi_\epsilon(q)} - \alpha_q, (\psi_\epsilon)^S_* A \rangle}{\epsilon} + \lim_{\epsilon \to 0} \frac{\langle \alpha_q, (\psi_\epsilon)^S_* A - A \rangle}{t} \\
&= \lim_{\epsilon \to 0} \frac{\langle \alpha_{\psi_\epsilon(q)} - \alpha_q, A \rangle}{\epsilon} - \lim_{\epsilon \to 0} \frac{\langle \alpha_q, (\psi_{-\epsilon})^S_* A - A \rangle}{\epsilon} \\
&= V(\langle \alpha, A \rangle) - \langle \alpha, \mathcal{L}_V A \rangle.
\end{aligned}
$$

(iv) Use (iii) to derive

$$
\begin{aligned}
\langle \mathcal{L}_V(f\alpha), A \rangle &= V(f\langle \alpha, A \rangle) - f\langle \alpha, \mathcal{L}_V A \rangle \\
&= (Vf)\langle \alpha, A \rangle + fV(\langle \alpha, A \rangle) - f\langle \alpha, \mathcal{L}_V A \rangle \\
&= (Vf)\langle \alpha, A \rangle + f\langle \mathcal{L}_V \alpha, A \rangle.
\end{aligned}
$$

(v) Again using (iii) we have $\langle \mathcal{L}_V(d^2 f), A \rangle = V(\langle d^2 f, A \rangle) - \langle d^2 f, \mathcal{L}_V A \rangle = VAf - [V, A]f = AVf = \langle d^2(Vf), A \rangle$.

$\square$

Corollary A.13 (i) $\mathcal{L}_V(df \cdot dg) = d(Vf) \cdot dg + df \cdot d(Vg)$.
(ii) $\mathcal{L}_V(\omega \cdot \eta) = \mathcal{L}_V \omega \cdot \eta + \omega \cdot \mathcal{L}_V \eta$.
(iii) $\mathcal{L}_V$ *commutes with the symmetric product operator* $\bullet$.

Proof For the first assertion,

$$
\begin{aligned}
&\langle \mathcal{L}_V(df \cdot dg), A \rangle \\
&= V(\langle df \cdot dg, A \rangle) - \langle df \cdot dg, \mathcal{L}_V A \rangle = V(\Gamma_A(f, g)) - \Gamma_{[V,A]}(f, g) \\
&= V(A(fg) - fAg - gAf) - ([V, A](fg) - f[V, A]g - g[V, A]f) \\
&= VA(fg) - VfAg - fVAg - VgAf - gVAf \\
&\quad - (VA(fg) - AV(fg) - fVAg + fAVg - gVAf + gAVf) \\
&= AV(fg) - VfAg - VgAf - fAVg - gAVf \\
&= [A(Vfg) - VfAg - gAVf] - [A(fVg) - VgAf - fAVg] \\
&= \langle d(Vf) \cdot dg, A \rangle + \langle df \cdot d(Vg), A \rangle.
\end{aligned}
$$

We use the local expressions to prove the second assertion. Assume, locally, that $\omega = \omega_i dx^i$ and $\eta = \eta_i dx^i$. Then, by (5.4), Lemma A.11.(ii) and Corollary A.11.(iv),

$$
\begin{aligned}
\mathcal{L}_V(\omega \cdot \eta) &= \mathcal{L}_V(\omega_i \eta_j dx^i \cdot dx^j) = V(\omega_i \eta_j) dx^i \cdot dx^j + \omega_i \eta_j \mathcal{L}_V(dx^i \cdot dx^j) \\
&= (\eta_j V \omega_i + \omega_i V \eta_j) dx^i \cdot dx^j + \omega_i \eta_j (dV^i \cdot dx^j + dx^i \cdot dV^j) \\
&= (V \omega_i dx^i + \omega_i dV^i) \cdot (\eta_j dx^j) + (\omega_i dx^i) \cdot (V \eta_j dx^j + \eta_j dV^j) \\
&= \mathcal{L}_V \omega \cdot \eta + \omega \cdot \mathcal{L}_V \eta.
\end{aligned}
$$

The last assertion is a consequence of the second one. Indeed,

$$
\begin{aligned}
\mathcal{L}_V(\bullet(\omega \otimes \eta)) &= \mathcal{L}_V(\omega \cdot \eta) = \mathcal{L}_V \omega \cdot \eta + \omega \cdot \mathcal{L}_V \eta \\
&= \bullet(\mathcal{L}_V \omega \otimes \eta + \omega \otimes \mathcal{L}_V \eta) = \bullet(\mathcal{L}_V(\omega \otimes \eta)). \qquad \square
\end{aligned}
$$

Given a vector field V on $\mathbb{R} \times M$, the Lie derivative $\mathcal{L}_V$ can also be defined for second-order vector fields and second-order forms on $\mathbb{R} \times M$, as in Definition A.10, without any changes. But when restricting to the mixed-order vector fields and mixed-order forms, it is necessary that the flow in Definition A.10 consists of bundle homomorphisms on $(\mathbb{R} \times M, \pi, \mathbb{R})$, so that its mixed-order pushforwards and pullbacks are well defined. This feeding back to the vector field V amounts to V is π-*projectable*. In this case, we just replace the second-order pushforwards and pullbacks in Definition A.10 by mixed-order pushforwards and pullbacks, to define the Lie derivative $\mathcal{L}_V$ for mixed-order vector fields and mixed-order forms on $\mathbb{R} \times M$.

Now let V be a π-projectable vector field on $\mathbb{R} \times M$. Then, Lemma A.11 (i)–(iv) still holds for smooth functions f on $\mathbb{R} \times M$, mixed-order vector fields A and mixed-order forms α on $\mathbb{R} \times M$. The assertion (v) will hold with the mixed differential instead of the second-order differential, that is, $\mathcal{L}_V(d^\circ f) = d^\circ(Vf)$. Moreover, if V and A have coordinate expressions $V = V^0 \frac{\partial}{\partial t} + V^i \frac{\partial}{\partial x^i}$ and $A = A^0 \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + A^{ij} \frac{\partial^2}{\partial x^i \partial x^j}$ where V^0 only depends on time, then the Lie derivative $\mathcal{L}_V A$ has the following expression:

$$
\begin{aligned}
\mathcal{L}_V A = [V, A] = {} & \left(V^0 \frac{\partial A^0}{\partial t} + V^j \frac{\partial A^0}{\partial x^j} - A^0 \frac{\partial V^0}{\partial t} \right) \frac{\partial}{\partial t} \\
& + \left(V^0 \frac{\partial A^i}{\partial t} + V^j \frac{\partial A^i}{\partial x^j} - A^0 \frac{\partial V^i}{\partial t} - A^j \frac{\partial V^i}{\partial x^j} - A^{jk} \frac{\partial^2 V^i}{\partial x^j \partial x^k} \right) \frac{\partial}{\partial x^i} \\
& + \left(V^0 \frac{\partial A^{jk}}{\partial t} + V^i \frac{\partial A^{jk}}{\partial x^i} \right) \frac{\partial^2}{\partial x^j \partial x^k} \\
& - A^{jk} \left(\frac{\partial V^i}{\partial x^j} \frac{\partial^2}{\partial x^i \partial x^k} + \frac{\partial V^i}{\partial x^k} \frac{\partial^2}{\partial x^i \partial x^j} \right).
\end{aligned}
$$

Appendix B The Mixed-Order Contact Structure on $\mathbb{R} \times \mathcal{T}^S M$

B.1 Mixed-Order Total Derivatives and Mixed-Order Contact Forms

We denote by $\pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M)$ the pullback bundle (see Saunders 1989, Definition 1.4.5) of $\tau_{\mathbb{R}} \times \tau_M^S$ by $\pi_{1,0}$. It is a fiber bundle over $\mathbb{R} \times \mathcal{T}^S M$.

Definition B.1 (*Mixed-order holonomic lift*) Let $t \in \mathbb{R}$, $q \in M$, $X \in I_{(t,q)}(M)$ and $\tau \frac{d}{dt}|_t \in T_t \mathbb{R}$. The mixed-order holonomic lift of $\tau \frac{\partial}{\partial t}|_t$ by X is defined to be

$$\left(X_*(\tau \tfrac{d}{dt}|_t), \, j_{(t,q)} X \right) \in \pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M).$$

The set of all mixed-order holonomic lifts is denoted by $H^R \pi_{1,0}$, that is,

$$H^R \pi_{1,0} := \left\{ \left(X_*(\tau \tfrac{d}{dt}|_t), \, j_{(t,q)} X \right) \in \pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M) : j_{(t,q)} X \in \mathbb{R} \times \mathcal{T}^S M, \, \tau \tfrac{d}{dt}|_t \in T_t \mathbb{R} \right\}.$$

Since X_* depends only upon the mean derivatives of X at t, the holonomic lift of a tangent vector is completely determined by $j_{(t,q)} X$ and does not depend on the choice of the representative diffusion X. In particular, the set $H^R \pi_{1,0}$ is well defined and is clearly a subbundle of $\pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M)$.

Lemma B.2 *The fiber bundle* $(\pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M), \pi_{1,0}^*(\tau_{\mathbb{R}} \times \tau_M^S), \mathbb{R} \times \mathcal{T}^S M)$ *can be written as the Whitney sum of two subbundles*

$$\pi_{1,0}^*(V^S \pi) \times_{\mathbb{R} \times \mathcal{T}^S M} H^R \pi_{1,0}.$$

Proof Suppose that $(A, j_{(t,q)} X) \in \pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M)$. Then, $A \in T\mathbb{R} \times \mathcal{T}^S M$, and

$$\left(X_*(\pi_*^R(A)), \, j_{(t,q)} X \right) \in H^R \pi_{1,0}.$$

It follows easily from the definition of pushforward (A.5) that $\pi_*^R(A - X_*(\pi_*^R(A))) = 0$. Hence, $A - X_*(\pi_*^R(A)) \in V^S \pi$ and

$$\left(A - X_*(\pi_*^R(A)), \, j_{(t,q)} X \right) \in \pi_{1,0}^*(V^S \pi).$$

The result follows.   □

The decomposition of $(A, j_{(t,q)} X) \in \pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M)$ may then be found by letting

$$A = A^0 \frac{\partial}{\partial t}\bigg|_t + A^i \frac{\partial}{\partial x^i}\bigg|_q + A^{jk} \frac{\partial^2}{\partial x^j \partial x^k}\bigg|_q$$

$$= \left(A^i - A^0 D^i x(j_{(t,q)} X) \right) \frac{\partial}{\partial x^i}\bigg|_q + \left(A^{jk} - A^0 Q^{jk} x(j_{(t,q)} X) \right) \frac{\partial^2}{\partial x^j \partial x^k}\bigg|_q$$

$$+ A^0 \left(\frac{\partial}{\partial t}\bigg|_t + D^i x(j_{(t,q)} X) \frac{\partial}{\partial x^i}\bigg|_q + \frac{1}{2} Q^{jk} x(j_{(t,q)} X) \frac{\partial^2}{\partial x^j \partial x^k}\bigg|_q \right).$$

Definition B.3 A section of the bundle $(H^R \pi_{1,0}, \pi_{1,0}^*(\tau_{\mathbb{R}} \times \tau_M^S)|_{H^R \pi_{1,0}}, \mathbb{R} \times \mathcal{T}^S M)$ is called a mixed-order total derivative. The specific section

$$\frac{\partial}{\partial t} + D^i x \frac{\partial}{\partial x^i} + \frac{1}{2} Q^{jk} x \frac{\partial^2}{\partial x^j \partial x^k}$$

is called the coordinate mixed-order total derivative, and is denoted by $\mathbf{D}_t$.

The coordinate mixed-order total derivative is just the total mean derivative in Definition 4.7. The dual construction is the mixed-order contact cotangent vector, which may be described as being in the kernel of X^*.

Definition B.4 An element $(\alpha, j_{(t,q)}X) \in \pi_{1,0}^*(T^*\mathbb{R} \times \mathcal{T}^{S*}M)$ is called a mixed-order contact cotangent vector if $X^*(\alpha) = 0$. The set of all mixed-order contact cotangent vectors is denoted by $C^{R*}\pi_{1,0}$, that is,

$$C^{R*}\pi_{1,0} := \left\{ (\alpha, j_{(t,q)}X) \in \pi_{1,0}^*(T^*\mathbb{R} \times \mathcal{T}^{S*}M) : j_{(t,q)}X \in \mathbb{R} \times \mathcal{T}^S M, X^*(\alpha) = 0 \right\}.$$

It is straightforward to check that the vanishing of X^* does not depend on the particular choice of the representative diffusion X. The dual relation between X^* and X_* in (A.7) implies that the mixed-order contact and holonomic elements annihilate each other.

To express a mixed-order contact cotangent vector $(\alpha, j_{(t,q)}X)$ in coordinates, let us consider

$$\alpha = \alpha_0 dt|_t + \alpha_i d^2 x^i|_q + \alpha_{jk} dx^j \cdot dx^k|_q. \tag{B.1}$$

Using the definition (A.6) we get

$$0 = X^*(\alpha) = \left(\alpha_0 + \alpha_i (DX)^i + \alpha_{jk} (QX)^{jk} \right) dt|_t.$$

There are two basic nontrivial solutions of the above equation, say,

$$\begin{cases} \alpha_0 = -\alpha_i (DX)^i, \\ \alpha_{jk} = 0, \end{cases} \quad \text{and} \quad \begin{cases} \alpha_0 = -\alpha_{jk} (QX)^{jk}, \\ \alpha_i = 0. \end{cases}$$

Plugging these solutions in (B.1), we get two basic types of mixed-order contact cotangent vectors

$$(d^2 x^i - D^i x dt)|_{j_{(t,q)}X} \quad \text{and} \quad (dx^j \cdot dx^k - Q^{jk} x dt)|_{j_{(t,q)}X}.$$

Thus, every mixed-order contact cotangent vector in $(C^{R*}\pi_{1,0})_{j_{(t,q)}X}$ is a linear combination of these basic mixed-order contact cotangent vectors.

Lemma B.5 *The fiber bundle* $(\pi_{1,0}^*(T^*\mathbb{R} \times \mathcal{T}^{S*}M), \pi_{1,0}^*(\tau_{\mathbb{R}}^* \times \tau_M^{S*}), \mathbb{R} \times \mathcal{T}^S M)$ *can be written as the Whitney sum of two subbundles*

$$\pi_1^*(T^*\mathbb{R}) \times_{\mathbb{R} \times \mathcal{T}^S M} C^{R*}\pi_{1,0}.$$

Proof Suppose that $(\alpha, j_{(t,q)}X) \in \pi_{1,0}^*(T^*\mathbb{R} \times \mathcal{T}^{S*}M)$. Then, $\alpha \in T^*\mathbb{R} \times \mathcal{T}^{S*}M$, and the definition of pullback yields

$$\left(X^*(\alpha), j_{(t,q)}X\right) \in \pi_1^*(T^*\mathbb{R}).$$

Since $X^*(\alpha - X^*(\alpha)) = 0$, it follows that

$$\left(\alpha - X^*(\alpha), j_{(t,q)}X\right) \in C^{R*}\pi_{1,0}.$$

This ends the proof. □

The decomposition of $(\alpha, j_{(t,q)}X) \in \pi_{1,0}^*(T\mathbb{R} \times \mathcal{T}^S M)$ may then be found by letting

$$
\begin{aligned}
\alpha &= \alpha_0 dt|_t + \alpha_i d^2 x^i|_q + \alpha_{jk} dx^j \cdot dx^k|_q \\
&= \left(\alpha_0 + \alpha_i D^i x(j_{(t,q)}X) + \alpha_{jk} Q^{jk} x(j_{(t,q)}X)\right) dt|_t \\
&\quad + \alpha_i \left(d^2 x^i - D^i x(j_{(t,q)}X)dt\right)\Big|_{(t,q)} + \alpha_{jk}\left(dx^j \cdot dx^k - Q^{jk} x(j_{(t,q)}X)dt\right)\Big|_{(t,q)}.
\end{aligned}
$$

Definition B.6 A section of the bundle $(C^{R*}\pi_{1,0}, \pi_{1,0}^*(\tau_{\mathbb{R}}^* \times \tau_M^{S*})|_{C^{R*}\pi_{1,0}}, \mathbb{R} \times \mathcal{T}^S M)$ is called a mixed-order contact form. The following specific sections

$$d^2 x^i - D^i x \, dt, \quad dx^j \cdot dx^k - Q^{jk} x \, dt, \quad 1 \le i, j, k \le d,$$

are called basic mixed-order contact forms.

It follows from the construction that the set of basic mixed-order contact forms defines a local frame of the bundle $\pi_{1,0}^*(\tau_{\mathbb{R}}^* \times \tau_M^{S*})|_{C^{R*}\pi_{1,0}}$.

Remark B.7 In contrast, we recall the classical contact forms on the first-order jet bundle $J^1\pi = \mathbb{R} \times TM$. Using the coordinates $(t, x^i, \dot{x}^i)$, the classical basic contact forms are $dx^i - \dot{x}^i dt$, $1 \le i \le d$. See Saunders (1989, Section 4.3) and Olver (1995, Theorem 4.23), also cf. Geiges (2008, p. 9), for a one-dimensional example.

Corollary B.8 *Let $(\mathbb{R} \times U, (t, x^i))$ be a coordinate chart on $\mathbb{R} \times M$. Let $\mathbf{X}$ be a $\mathcal{T}^S M$-valued diffusion process. In local coordinates, the pushforward map $\mathbf{X}_*$ from $T\mathbb{R}$ to $T\mathbb{R} \times \mathcal{T}^S \mathcal{T}^S M$ is given by*

$$
\begin{aligned}
\mathbf{X}_*\left(\tau \frac{d}{dt}\Big|_t\right) = \tau\Bigg(&\frac{\partial}{\partial t} + D^i(x \circ \mathbf{X})\frac{\partial}{\partial x^i} + D^i(Dx \circ \mathbf{X})\frac{\partial}{\partial D^i x} + D^{jk}(Qx \circ \mathbf{X})\frac{\partial}{\partial Q^{jk} x} \\
&+ \frac{1}{2}Q^{jk}(x \circ \mathbf{X})\frac{\partial^2}{\partial x^j \partial x^k} + \frac{1}{2}Q^{jk}(Dx \circ \mathbf{X})\frac{\partial^2}{\partial D^j x \partial D^k x} \\
&+ \frac{1}{2}Q^{jklm}(Qx \circ \mathbf{X})\frac{\partial^2}{\partial Q^{jk} x \partial Q^{lm} x} \\
&+ \frac{1}{2}Q^{jk}(x \circ \mathbf{X}, Dx \circ \mathbf{X})\frac{\partial^2}{\partial x^j \partial D^k x} + \frac{1}{2}Q^{jkl}(x \circ \mathbf{X}, Qx \circ \mathbf{X})\frac{\partial^2}{\partial x^j \partial Q^{kl} x} \\
&+ \frac{1}{2}Q^{jkl}(Dx \circ \mathbf{X}, Qx \circ \mathbf{X})\frac{\partial^2}{\partial D^j x \partial Q^{kl} x}\Bigg)\Bigg|_{(t, \mathbf{X}(t))}.
\end{aligned}
$$

The pullback map $\mathbf{X}^*$ from $T^*\mathbb{R} \times \mathcal{T}^{S*}\mathcal{T}^S M$ to $T^*\mathbb{R}$ is given by

$$
\begin{aligned}
\mathbf{X}^* \Big(& \alpha_0 dt + \alpha_i d^2 x^i + \alpha_i^1 d^2 D^i x + \alpha_{jk}^2 d^2 Q^{jk} x + \alpha_{jk} dx^j \cdot dx^k \\
& + \alpha_{jk}^1 d D^j x \cdot d D^k x + \alpha_{jklm}^2 d Q^{jk} x \cdot d Q^{lm} x \\
& + \alpha_{jk}^{01} dx^j \cdot d D^k x + \alpha_{jkl}^{02} dx^j \cdot d Q^{kl} x + \alpha_{jkl}^{12} d D^j x \cdot d Q^{kl} x \Big) \Big|_{(t,\mathbf{X}(t))} \\
= \Big(& \alpha_0 + \alpha_i D^i (x \circ \mathbf{X}) + \alpha_i^1 D^i (Dx \circ \mathbf{X}) + \alpha_{jk}^2 D^{jk} (Qx \circ \mathbf{X}) \\
& + \alpha_{jk} Q^{jk} (x \circ \mathbf{X}) + \alpha_{jk}^1 Q^{jk} (Dx \circ \mathbf{X}) + \alpha_{jklm}^2 Q^{jklm} (Qx \circ \mathbf{X}) \\
& + \alpha_{jk}^{01} Q^{jk} (x \circ \mathbf{X}, Dx \circ \mathbf{X}) + \alpha_{jkl}^{02} Q^{jkl} (x \circ \mathbf{X}, Qx \circ \mathbf{X}) \\
& + \alpha_{jkl}^{12} Q^{jkl} (Dx \circ \mathbf{X}, Qx \circ \mathbf{X}) \Big) dt |_t.
\end{aligned}
$$

Corollary B.9 *Let α be a section of $(T^*\mathbb{R} \times \mathcal{T}^{S*}\mathcal{T}^S M, \tau_{\mathbb{R}}^* \times \tau_{\mathcal{T}^S M}^{S*}, \mathbb{R} \times \mathcal{T}^S M)$. Then, α is a mixed-order contact form if and only if for every $t \in \mathbb{R}$ and every $X \in \cup_{q \in M} I_{(t,q)}(M)$,*

$$
(jX)^* (\alpha |_{j_{(t,q)} X}) = 0.
$$

Proof We first let $\alpha = \alpha_0 dt + \alpha_i d^2 x^i + \alpha_{jk} dx^j \cdot dx^k$ be a mixed-order contact form and let $X \in I_{(t,q)}(M)$. Then,

$$
(jX)^* (\alpha |_{j_{(t,q)} X}) = \Big(\alpha_0 + \alpha_i D^i x + \alpha_{jk} Q^{jk} x \Big) (j_{(t,q)} X) dt |_t = X^* (\alpha |_{j_{(t,q)} X}) = 0.
$$

(B.2)

To prove the converse, we suppose

$$
\begin{aligned}
\alpha = & \alpha_0 dt + \alpha_i d^2 x^i + \alpha_i^1 d^2 D^i x + \alpha_{jk}^2 d^2 Q^{jk} x + \alpha_{jk} dx^j \cdot dx^k \\
& + \alpha_{jk}^1 d D^j x \cdot d D^k x + \alpha_{jklm}^2 d Q^{jk} x \cdot d Q^{lm} x \\
& + \alpha_{jk}^{01} dx^j \cdot d D^k x + \alpha_{jkl}^{02} dx^j \cdot d Q^{kl} x + \alpha_{jkl}^{12} d D^j x \cdot d Q^{kl} x
\end{aligned}
$$

Fix a particular index i_0 with $1 \le i_0 \le d$. Let $Y \in I_{(t,q)}(M)$ such that $j_{(t,q)} X = j_{(t,q)} Y$, $D^i DY = D^i DX + \delta_{i_0}^i$ and

$$
\begin{aligned}
& \Big(D^{jk} QY, Q^{jk} DY, Q^{jklm} QY, Q^{jk}(Y, DY), Q^{jkl}(Y, QY), Q^{jkl}(DY, QY) \Big) \\
& = \Big(D^{jk} QX, Q^{jk} DX, Q^{jklm} QX, Q^{jk}(X, DX), Q^{jkl}(X, QX), Q^{jkl}(DX, QX) \Big).
\end{aligned}
$$

Then,

$$
0 = (jY)^* (\alpha |_{j_{(t,q)} Y}) = (jX)^* (\alpha |_{j_{(t,q)} X}) + \alpha_i^1 \delta_{i_0}^i = \alpha_{i_0}^1.
$$

It follows from the arbitrariness of i_0 that $\alpha_i^1 = 0$ for all $1 \le i \le d$. Similarly, all α_{jk}^1, α_{jk}^2 and α_{jklm}^2 vanish. Consequently, $\alpha = \alpha_0 dt + \alpha_i d^2 x^i + \alpha_{jk} dx^j \cdot dx^k$. As in (B.2), we have $(jX)^*(\alpha|_{j_{(t,q)}X}) = X^*(\alpha|_{j_{(t,q)}X}) = 0$. Hence, α is a mixed-order contact form. $\qquad\square$

Corollary B.10 *Let* $\mathbf{X}$ *be a* $\mathcal{T}^S M$-*valued diffusion process. Then,* $\mathbf{X} = jX$, *with* X *an M-valued diffusion process, if and only if* $\mathbf{X}^*(\alpha) = 0$ *for every mixed-order contact form* α *on* $\mathbb{R} \times \mathcal{T}^S M$.

Proof We first suppose $\mathbf{X} = jX$ with X an M-valued diffusion process. Then, for a mixed-order contact form α,

$$\mathbf{X}^*(\alpha) = (jX)^*(\alpha) = X^*(\alpha) = 0.$$

To prove the converse, it suffices to show, in local coordinates, that

$$D^i x(\mathbf{X}) = D^i (x \circ \mathbf{X}), \qquad Q^{jk} x(\mathbf{X}) = Q^{jk} (x \circ \mathbf{X}).$$

This can be done as soon as we let α be a basic mixed-order contact form. For example, let $\alpha = d^2 x^i - D^i x dt$, then

$$0 = \mathbf{X}^*(\alpha) = \left(D^i (x \circ \mathbf{X}) - D^i x \circ \mathbf{X} \right) dt,$$

which leads to $D^i x(\mathbf{X}) = D^i (x \circ \mathbf{X})$. $\qquad\square$

B.2 The Mixed-Order Cartan Distribution and Its Symmetries

The model bundle $\mathbb{R} \times \mathcal{T}^S M$ is a trivial bundle over $\mathbb{R}$ in its own right, and so we may consider its mixed-order tangent bundle $(T\mathbb{R} \times \mathcal{T}^S \mathcal{T}^S M, \tau_{\mathbb{R}} \times \tau_{\mathcal{T}^S M}^S, \mathbb{R} \times \mathcal{T}^S M)$.

Definition B.11 The bundle endomorphism $(v, \mathbf{Id}_E)$ of $\pi_{1,0}^*(\tau_{\mathbb{R}} \times \tau_M^S)$ is defined by

$$v(A^h + A^v) = A^v,$$

where $A^h \in H^R \pi_{1,0}$ and $A^v \in \pi_{1,0}^*(V^S \pi)$.

Definition B.12 (*Mixed-order Cartan distribution*) The mixed-order Cartan distribution is the kernel of the vector bundle homomorphism over $\mathbf{Id}_{\mathbb{R} \times \mathcal{T}^S M}$

$$v \circ \left(\pi_{1,0*}, \tau_{\mathbb{R}} \times \tau_{\mathcal{T}^S M}^S \right) : T\mathbb{R} \times \mathcal{T}^S \mathcal{T}^S M \to \pi_{1,0}^* \left(\tau_{\mathbb{R}} \times \tau_M^S \right)$$

and is denoted by $C^R \pi_{1,0}$.

Note that $C^R \pi_{1,0}$ is a subbundle of $\tau_{\mathbb{R}} \times \tau_{\mathcal{T}^S M}^S$. It follows from the above two definitions that

$$C^R \pi_{1,0} = \left(\pi_{1,0*}, \tau_{\mathbb{R}} \times \tau_{\mathcal{T}^S M}^S \right)^{-1} H^R \pi_{1,0}.$$

Hence, for each $X \in I_{(t,q)}(M)$,

$$C^R \pi_{1,0}|_{j_{(t,q)}X} = (jX)_*(T_t \mathbb{R}) \oplus V^S \pi_{1,0}|_{j_{(t,q)}X}.$$

Similarly to the proof of Lemma B.2, we can decompose an element $\mathbf{A} \in C^R \pi_{1,0}|_{j_{(t,q)}X}$ as

$$\mathbf{A} = (jX)_*((\pi_1)^R_*(\mathbf{A})) + \left[\mathbf{A} - (jX)_*((\pi_1)^R_*(\mathbf{A})) \right], \tag{B.3}$$

where $(jX)_*((\pi_1)^R_*(\mathbf{A})) \in (jX)_*(T_t \mathbb{R})|_{j_{(t,q)}X}$ and $\mathbf{A} - (jX)_*((\pi_1)^R_*(\mathbf{A})) \in V^S_{j_{(t,q)}X}\pi_{1,0}$.

From the duality relations it also follows that $(\tau^*_{\mathbb{R}} \times \tau^{S*}_{T^S M})|_{C^{R*}\pi_{1,0}}$ is the annihilator of $(\tau_{\mathbb{R}} \times \tau^S_{T^S M})|_{C^R \pi_{1,0}}$, or in other words, the basic mixed-order contact forms are local defining forms for the mixed-order contact distribution $C^R \pi_{1,0}$. A typical element $\mathbf{A} \in C^R \pi_{1,0}|_{j_{(t,q)}X}$ may be written in coordinates as

$$
\begin{aligned}
\mathbf{A} = \mathbf{A}^0 & \left(\frac{\partial}{\partial t}\bigg|_{j_{(t,q)}X} + D^i x(j_{(t,q)}X) \frac{\partial}{\partial x^i}\bigg|_{j_{(t,q)}X} + \frac{1}{2} Q^{jk} x(j_{(t,q)}X) \frac{\partial^2}{\partial x^j \partial x^k}\bigg|_{j_{(t,q)}X} \right) \\
& + \mathbf{A}^i_1 \frac{\partial}{\partial D^i x}\bigg|_{j_{(t,q)}X} + \mathbf{A}^{jk}_2 \frac{\partial}{\partial Q^{jk}x}\bigg|_{j_{(t,q)}X} + \mathbf{A}^{jk}_{11} \frac{\partial^2}{\partial D^j x \partial D^k x}\bigg|_{j_{(t,q)}X} \\
& + \mathbf{A}^{jklm}_{22} \frac{\partial^2}{\partial Q^{jk} x \partial Q^{lm} x}\bigg|_{j_{(t,q)}X} \\
& + \mathbf{A}^{jk}_{01} \frac{\partial^2}{\partial x^j \partial D^k x}\bigg|_{j_{(t,q)}X} + \mathbf{A}^{jkl}_{02} \frac{\partial^2}{\partial x^j \partial Q^{kl} x}\bigg|_{j_{(t,q)}X} + \mathbf{A}^{jkl}_{12} \frac{\partial^2}{\partial D^j x \partial Q^{kl} x}\bigg|_{j_{(t,q)}X}.
\end{aligned}
\tag{B.4}
$$

From this it is easy to deduce $(\pi_{1,0})^R_* \mathbf{A} \in H^R \pi_{1,0}$.

Definition B.13 A symmetry of the mixed-order Cartan distribution on $\mathbb{R} \times T^S M$ is a bundle automorphism $\mathbf{F}$ of $\mathbb{R} \times T^S M$ which satisfies $\mathbf{F}^R_*(C^R \pi_{1,0}) = C^R \pi_{1,0}$.

It follows by duality that symmetries of the mixed-order Cartan distribution are those bundle automorphisms which satisfy $\mathbf{F}^{R*}(C^{R*}\pi_{1,0}) = C^{R*}\pi_{1,0}$. For this reason, $\mathbf{F}$ is also called a mixed-order contact transformation. Similarly, $\mathbf{F}$ may be characterized by the fact that whenever α is a mixed-order contact form then so is $\mathbf{F}^{R*}(\alpha)$.

Proposition B.14 Let $\mathbf{F}$ be a bundle homomorphism from $(\mathbb{R} \times T^S M, \pi_1, \mathbb{R})$ to $(\mathbb{R} \times T^S N, \rho_1, \mathbb{R})$ that projects to a diffeomorphism $F^0 : \mathbb{R} \to \mathbb{R}$. Then, $\mathbf{F}^R_*(C^R \pi_{1,0}) \subset C^R \rho_{1,0}$ if and only if $\mathbf{F} = jF$ where F is a bundle homomorphism from $(\mathbb{R} \times M, \pi, \mathbb{R})$ to $(\mathbb{R} \times N, \rho, \mathbb{R})$ that projects to F^0.

Proof First, we prove the sufficiency. Let $\mathbf{A} \in C^R \pi_{1,0}|_{j_{(t,q)}X}$. According to (B.3), we decompose $\mathbf{A}$ by $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ with $\mathbf{A}_1 = (jX)_*((\pi_1)^R_*(\mathbf{A})) \in (jX)_*(T_t \mathbb{R})$ and $\mathbf{A}_2 \in V^S_{j_{(t,q)}X}\pi_{1,0}$. Then, since by Corollaries 4.6 and A.9.(iv), $(jF)^R_* \circ (jX)_* =$

$(jF \cdot jX)_* \circ (F^0)_* = (j\tilde{X})_* \circ (F^0)_*$ where $\tilde{X} = F \cdot X$ is the pushforward of X by F, we have

$$\mathbf{F}_*^R(\mathbf{A}_1) = (jF)_*^R(\mathbf{A}_1) = (jF)_*^R(jX)_*(\pi_1)_*^R\mathbf{A} = (j\tilde{X})_*(F^0)_*(\pi_1)_*^R\mathbf{A} \in (j\tilde{X})_*(\mathcal{T}_{F^0(t)}\mathbb{R}).$$

Besides, since $jF : \pi_{1,0} \to \rho_{1,0}$ is a bundle homomorphism projecting to F by Corollary 4.5.(ii), we have $\rho_{1,0} \circ jF = F \circ \pi_{1,0}$. Then,

$$(\rho_{1,0})_*^S(\mathbf{F}_*^R(\mathbf{A}_2)) = (\rho_{1,0})_*^S((jF)_*^R(\mathbf{A}_2)) = F_*^S(\pi_{1,0})_*^S(\mathbf{A}_2) = 0,$$

which yields $\mathbf{F}_*^R(\mathbf{A}_2) \in V^S\rho_{1,0}$. This proves $\mathbf{F}_*^R(C^R\pi_{1,0}) \subset C^R\rho_{1,0}$.

For the necessity, we first prove that $\mathbf{F}$ is bundle homomorphism from $\pi_{1,0}$ to $\rho_{1,0}$ by showing $\mathbf{F}_*^S(V^S\pi_{1,0}) \subset V^S\rho_{1,0}$, by virtue of Lemma A.1. Let $\mathbf{A} \in V^S\pi_{1,0}$. Set $\mathbf{F}_*^R\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$, where $\mathbf{A}_1 \in (jY)_*(\mathcal{T}_{F^0(t)}\mathbb{R})$ and $\mathbf{A}_2 \in V^S\rho_{1,0}$ for some diffusion Y. Since $\mathbf{F}$ projects to F^0,

$$(\rho_1)_*^S(\mathbf{F}_*^S\mathbf{A}) = (F^0)_*^S(\pi_1)_*^S\mathbf{A} = (F^0)_*^S\pi_*^S(\pi_{1,0})_*^S\mathbf{A} = 0,$$

while $(\rho_1)_*^S\mathbf{A}_2 = \rho_*^S(\rho_{1,0})_*^S\mathbf{A}_2 = 0$. Thus, $(\rho_1)_*^S\mathbf{A}_1 = 0$. Since $\mathbf{A}_1 \in (jY)_*(\mathcal{T}_{F^0(t)}\mathbb{R})$, we set $\mathbf{A}_1 = (jY)_*(\tau \frac{\partial}{\partial s}|_{F^0(t)})$. Then, $(\rho_1)_*^S\mathbf{A}_1 = \tau \frac{\partial}{\partial s}|_{F^0(t)} = 0$. Hence, $\tau = 0$ and so $\mathbf{A}_1 = 0$. This leads to $\mathbf{F}_*^R(V^S\pi_{1,0}) \subset V^S\rho_{1,0}$ and so that $\mathbf{F}$ is bundle homomorphism from $\pi_{1,0}$ to $\rho_{1,0}$. Denote the projection of $\mathbf{F}$ onto a map from $\mathbb{R} \times M$ to $\mathbb{R} \times N$ by F. It follows that

$$\rho \circ F \circ \pi_{1,0} = \rho \circ \rho_{1,0} \circ \mathbf{F} = \rho_1 \circ \mathbf{F} = F^0 \circ \pi_1 = F^0 \circ \pi \circ \pi_{1,0}.$$

Since $\pi_{1,0}$ is surjective, we obtain $\rho \circ F = F^0 \circ \pi$, so that F is a bundle homomorphism from π to ρ projecting to F^0. We shall write $F = (F^0, \bar{F})$ and $\mathbf{F} = (F^0, \bar{\mathbf{F}})$.

Next, we will show $\mathbf{F} = jF$. Fix a $j_{(t,q)}X \in \mathbb{R} \times \mathcal{T}^SM$. Let $\mathbf{F}(j_{(t,q)}X) = j_{(s,q')}Y$. Then, $s = F^0(t)$ and $(s, q') = F(t, q)$. For an element $\mathbf{A} \in C^R\pi_{1,0}|_{j_{(t,q)}X}$ with local expression in (B.4), we have from (A.3) that

$$\mathbf{F}_*^R\mathbf{A} = \mathbf{A}^0 \frac{\mathrm{d}F^0}{\mathrm{d}t}(t) \frac{\partial}{\partial s}\bigg|_{j_{(s,q')}Y} + (\mathbf{A}\bar{F}^i)(j_{(t,q)}X) \frac{\partial}{\partial y^i}\bigg|_{j_{(s,q')}Y}$$

$$+ \frac{\mathbf{A}^0}{2}Q^{jk}x(j_{(t,q)}X) \frac{\partial \bar{F}^i}{\partial x^k} \frac{\partial \bar{F}^j}{\partial x^l}(t,q) \frac{\partial^2}{\partial y^i \partial y^j}\bigg|_{j_{(s,q')}Y}$$

$$+ \text{terms} \left(\frac{\partial}{\partial D^i y}\bigg|_{j_{(s,q')}Y}, \frac{\partial}{\partial Q^{ij}y}\bigg|_{j_{(s,q')}Y}, \cdots \right).$$

Since $\bar{F}$ only depends on the variables on $\mathbb{R} \times M$, we have

$$
\begin{aligned}
(\mathbf{A}\bar{F}^i)(j_{(t,q)}X) &= \left((\pi_{1,0})_*^R \mathbf{A} \right) \bar{F}^i(j_{(t,q)}X) \\
&= \mathbf{A}^0 \left[\frac{\partial \bar{F}^i}{\partial t}(t,q) + D^j x(j_{(t,q)}X) \frac{\partial \bar{F}^i}{\partial x^j}(t,q) \right. \\
&\quad \left. + \frac{1}{2} Q^{jk} x(j_{(t,q)}X) \frac{\partial^2 \bar{F}^i}{\partial x^j \partial x^k}(t,q) \right].
\end{aligned}
$$

Then, the local expressions for jF in (4.7) and (4.8) yield

$$
\begin{aligned}
\mathbf{F}_*^R \mathbf{A} &= \mathbf{A}^0 \frac{\mathrm{d}F^0}{\mathrm{d}t}(t) \left[\frac{\partial}{\partial s} \bigg|_{j_{(s,q')}Y} + D^i y \circ jF(j_{(t,q)}X) \frac{\partial}{\partial y^i} \bigg|_{j_{(s,q')}Y} \right. \\
&\quad \left. + \frac{1}{2} Q^{ij} y \circ jF(j_{(t,q)}X) \frac{\partial^2}{\partial y^i \partial y^j} \bigg|_{j_{(s,q')}Y} \right].
\end{aligned}
$$

Since $\mathbf{F}_*^R \mathbf{A} \in C^R \pi_{1,0}|_{j_{(s,q')}Y}$ by the assumption, it follows that $jF(j_{(t,q)}X) = j_{(s,q')}Y = \mathbf{F}(j_{(t,q)}X)$. This proves that $\mathbf{F} = jF$. □

Corollary B.15 *Let $\mathbf{F}$ be a bundle automorphism on $(\mathbb{R} \times \mathcal{T}^S M, \pi_1, \mathbb{R})$ projecting to a diffeomorphism $F^0 : \mathbb{R} \to \mathbb{R}$. Then, $\mathbf{F}$ is a symmetry of $C^R \pi_{1,0}$ if and only if $\mathbf{F} = jF$ where F is a bundle automorphism on $(\mathbb{R} \times M, \pi, \mathbb{R})$ that projects to F^0.*

Proof If $\mathbf{F}$ is a symmetry, then $\mathbf{F}_*^R(C^R \pi_{1,0}) \subset C^R \pi_{1,0}$ and $(\mathbf{F}^{-1})_*^R(C^R \pi_{1,0}) \subset C^R \pi_{1,0}$. By Proposition B.14, $\mathbf{F} = jF$ and $\mathbf{F}^{-1} = jG$ for some bundle endomorphisms F and G on $(\mathbb{R} \times M, \pi_1, \mathbb{R})$ that projects to F^0 and $(F^0)^{-1}$, respectively. Then, Corollary 4.5.(iii) implies that $j(F \circ G) = jF \circ jG = F \circ F^{-1} = \mathbf{Id}_{\mathbb{R} \times \mathcal{T}^S M}$ and hence $F \circ G = \mathbf{Id}_{\mathbb{R} \times M}$. For the same reason, $G \circ F = \mathbf{Id}_{\mathbb{R} \times M}$. Thus, F is a bundle automorphism on π. Conversely, if $\mathbf{F} = jF$ and F is a bundle automorphism, then $\mathbf{F} \circ jF^{-1} = jF^{-1} \circ F = \mathbf{Id}_{\mathbb{R} \times \mathcal{T}^S M}$, which yields $\mathbf{F}^{-1} = jF^{-1}$ and hence $\mathbf{F}$ is a bundle automorphism on π_1. □

B.3 Infinitesimal Symmetries

Definition B.16 An infinitesimal symmetry of the mixed-order Cartan distribution is a π_1-projectable vector field $\mathbf{V}$ on $\mathbb{R} \times \mathcal{T}^S M$ with the property that, whenever the mixed-order vector field $\mathbf{A}$ belongs to $C^R \pi_{1,0}$, then so does the mixed-order vector field $\mathcal{L}_\mathbf{V}\mathbf{A}$.

Like in the classical case, an infinitesimal symmetry of the mixed-order Cartan distribution may also be called an infinitesimal mixed-order contact transformation. By duality, $\mathbf{V}$ is such an infinitesimal symmetry precisely when $\mathcal{L}_\mathbf{V}\alpha$ is a contact form for every mixed-order contact form α.

The following lemma is a consequence of the definition of Lie derivatives.

Lemma B.17 *Let* $\mathbf{V}$ *be a* π_1*-projectable vector field on* $\mathbb{R} \times \mathcal{T}^S M$ *with flow* $\Psi = \{\Psi_\epsilon\}_{\epsilon \in \mathbb{R}}$*. Then,* $\mathbf{V}$ *is an infinitesimal symmetry of the mixed-order Cartan distribution if and only if for each* ϵ*, the diffeomorphism* Ψ_ϵ *is a symmetry of the mixed-order Cartan distribution.*

The following result is the infinitesimal version of Corollary B.15. It can be deduced directly from Lemma B.17 and Corollary B.15. But here we give a computational proof based on the Lie derivative of mixed-order contact forms.

Theorem B.18 *Let* $\mathbf{V}$ *be a* π_1*-projectable vector field on* $\mathbb{R} \times \mathcal{T}^S M$*. Then,* $\mathbf{V}$ *is an infinitesimal symmetry of the mixed-order Cartan distribution if and only if* $\mathbf{V}$ *is the prolongation of a* π*-projectable vector field* V *on* $\mathbb{R} \times M$*.*

Proof Let the vector field $\mathbf{V}$ having the following local expression:

$$
\mathbf{V} = \mathbf{V}^0 \frac{\partial}{\partial t} + \mathbf{V}^i \frac{\partial}{\partial x^i} + \mathbf{V}^i_1 \frac{\partial}{\partial D^i x} + \mathbf{V}^i_2 \frac{\partial}{\partial Q^{jk} x},
$$

where $\mathbf{V}^0$ only depends on time due to the projectability of $\mathbf{V}$. We then derive the Lie derivative $\mathcal{L}_{\mathbf{V}}$ of the basic mixed-order contact forms $d^2 x^i - D^i x dt$ and $dx^j \cdot dx^k - Q^{jk} x dt$ as follows:

$$
\begin{aligned}
&\mathcal{L}_{\mathbf{V}}(d^2 x^i - D^i x dt) \\
&= d^\circ \mathbf{V}^i - \mathbf{V}^i_1 dt - D^i x d\mathbf{V}^0 \\
&= \frac{\partial \mathbf{V}^i}{\partial t} dt + \frac{\partial \mathbf{V}^i}{\partial x^j} d^2 x^j + \frac{1}{2} \frac{\partial^2 \mathbf{V}^i}{\partial x^j \partial x^k} dx^j \cdot dx^k + \text{terms} \left(\frac{\partial \mathbf{V}^i}{\partial D^j x}, \frac{\partial \mathbf{V}^i}{\partial Q^{jk} x}, \cdots \right) \\
&\quad - \mathbf{V}^i_1 dt - D^i x \frac{d\mathbf{V}^0}{dt} dt \\
&= \frac{\partial \mathbf{V}^i}{\partial x^j} (d^2 x^j - D^j x dt) + \frac{1}{2} \frac{\partial^2 \mathbf{V}^i}{\partial x^j \partial x^k} (dx^j \cdot dx^k - Q^{jk} x dt) \\
&\quad + \left(\frac{\partial \mathbf{V}^i}{\partial t} + \frac{\partial \mathbf{V}^i}{\partial x^j} D^j x + \frac{1}{2} \frac{\partial^2 \mathbf{V}^i}{\partial x^j \partial x^k} Q^{jk} x - \mathbf{V}^i_1 - D^i x \frac{d\mathbf{V}^0}{dt} \right) dt \\
&\quad + \text{terms} \left(\frac{\partial \mathbf{V}^i}{\partial D^j x}, \frac{\partial \mathbf{V}^i}{\partial Q^{jk} x}, \cdots \right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathcal{L}_{\mathbf{V}}(dx^j \cdot dx^k - Q^{jk} x dt) \\
&= d\mathbf{V}^j \cdot dx^k + dx^j \cdot d\mathbf{V}^k - \mathbf{V}^{jk}_2 dt - Q^{jk} x d\mathbf{V}^0 \\
&= \frac{\partial \mathbf{V}^j}{\partial x^i} dx^i \cdot dx^k + \frac{\partial \mathbf{V}^k}{\partial x^i} dx^j \cdot dx^i - \mathbf{V}^{jk}_2 dt - Q^{jk} x d\mathbf{V}^0 \\
&= \frac{\partial \mathbf{V}^j}{\partial x^i} (dx^i \cdot dx^k - Q^{ik} x dt) + \frac{\partial \mathbf{V}^k}{\partial x^i} (dx^i \cdot dx^j - Q^{ij} x dt) \\
&\quad + \left(\frac{\partial \mathbf{V}^j}{\partial x^i} Q^{ik} x + \frac{\partial \mathbf{V}^k}{\partial x^i} Q^{ij} x - \mathbf{V}^{jk}_2 - Q^{jk} x \frac{d\mathbf{V}^0}{dt} \right) dt.
\end{aligned}
$$

Thus, the mixed-order forms $\mathcal{L}_\mathbf{V}(d^2 x^i - D^i x dt)$ and $\mathcal{L}_\mathbf{V}(dx^j \cdot dx^k - Q^{jk} x dt)$ are mixed-order contact forms if and only if

$$\text{terms } \frac{\partial \mathbf{V}^i}{\partial D^j x}, \frac{\partial \mathbf{V}^i}{\partial Q^{jk} x}, \text{ etc, vanish and} \tag{B.5}$$

$$\frac{\partial \mathbf{V}^i}{\partial t} + \frac{\partial \mathbf{V}^i}{\partial x^j} D^j x + \frac{1}{2} \frac{\partial^2 \mathbf{V}^i}{\partial x^j \partial x^k} Q^{jk} x - \mathbf{V}_1^i - D^i x \frac{d\mathbf{V}^0}{dt} = 0, \tag{B.6}$$

$$\frac{\partial \mathbf{V}^j}{\partial x^i} Q^{ik} x + \frac{\partial \mathbf{V}^k}{\partial x^i} Q^{ij} x - \mathbf{V}_2^{jk} - Q^{jk} x \frac{d\mathbf{V}^0}{dt} = 0. \tag{B.7}$$

Now (B.5) means that $\mathbf{V}^i$'s only depend on the variables on $\mathbb{R} \times M$, so that the vector field $\mathbf{V}$ is also $\pi_{1,0}$-projectable. The two equations (B.6) and (B.7) are just restatements of the prolongation formulae in Theorem 4.14. $\qquad\square$

Appendix C Stochastic Maupertuis's Principle

Based on Definition 7.11, if we further consider the variation caused by time-change, as in classical mechanics (cf. Abraham and Marsden 1978, Definition 3.8.4 or the so called Δ-variation in Goldstein et al. 2002, Section 8.6), then we need to impose the constraint of constant energy. So the path space $\mathcal{A}_g([0, T]; q, \mu)$ in (7.10) is modified to

$$\begin{aligned}
\mathcal{A}_g([0, T]; q, \mu; e) := \Big\{ &(X, \tau) : \tau \in C^2([0, T], \mathbb{R}), \tau' > 0, X \in I_{(\tau(0),q)}^{(\tau(T),\mu)}(M), \\
&QX(t) = \check{g}(X(t)), \forall t \in [\tau(0), \tau(T)], \text{ a.s.}, \\
&\mathbf{E} E_0(t, X(t), D_\nabla X(t)) = e, \forall t \in [\tau(0), \tau(T)] \Big\},
\end{aligned}$$

where $e \in \mathbb{R}$ is a regular value of E_0.

Definition C.1 Given $v \in \mathcal{H}([0, T]; q)$ and $\varsigma \in \mathcal{C}^1([0, T], \mathbb{R})$, by a variation of the pair $(X, \tau) \in \mathcal{A}_g([0, T]; q, \mu; e)$ along (v, ς), we mean a family of pairs $\{(X_\epsilon^{v,\varsigma}, \tau_\epsilon^\varsigma)\}_{\epsilon \in (-\varepsilon, \varepsilon)}$ where $\tau_0^\varsigma = \tau$, $\frac{\partial}{\partial t} \tau_\epsilon^\varsigma > 0$, such that for each ϵ, $\frac{\partial}{\partial \epsilon} \tau_\epsilon^\varsigma|_{\epsilon=0} = \varsigma$, $X_\epsilon^{v,\varsigma} \in I_{(\tau_\epsilon^\varsigma(0),q)}^{(\tau_\epsilon^\varsigma(T),\mu)}(M)$, and for each $t \in [\tau_\epsilon^\varsigma(0), \tau_\epsilon^\varsigma(T)]$, $\mathbf{E} E_0(t, X_\epsilon^{v,\varsigma}(t), D_\nabla X_\epsilon^{v,\varsigma}(t)) = e$, $X_\epsilon^{v,\varsigma}(t)$ satisfies the ODE

$$\frac{\partial}{\partial \epsilon} X_\epsilon^{v,\varsigma}(t) = \Gamma(X_\epsilon^{v,\varsigma})_{\tau_\epsilon^\varsigma(0)}^t v(t), \quad X_0^{v,\varsigma}(t) = X(t). \tag{C.1}$$

Define a functional $\mathcal{I} : \mathcal{A}_g([0, T]; q, \mu; e) \to \mathbb{R}$ by

$$\mathcal{I}[X, \tau] := \mathbf{E} \int_{\tau(0)}^{\tau(T)} A_0(t, X(t), D_\nabla X(t)) \, dt.$$

The pair $(X, \tau) \in \mathcal{A}_g([0, T]; q, \mu; e)$ is called a stationary point of $\mathcal{I}$, if

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} \mathcal{I}[X_\epsilon^{v,\varsigma}, \tau_\epsilon^\varsigma] = 0, \quad \text{for all } v \in \mathcal{H}([0, T]; q) \text{ and } \varsigma \in \mathcal{C}^1([0, T], \mathbb{R}).$$

As in Lemma 7.13, it is easy to deduce from (C.1) that $QX_\epsilon^{v,\varsigma}(t) = \check{g}(X_\epsilon^{v,\varsigma}(t))$ for each $t \in [\tau_\epsilon^\varsigma(0), \tau_\epsilon^\varsigma(T)]$ so that $X_\epsilon^{v,\varsigma} \in \mathcal{A}_g([0, T]; q, \mu; e)$. Moreover, formula (7.13) still holds for all $t \in [\tau(0), \tau(T)]$, with $X_\epsilon^{v,\varsigma}$ in place of X_ϵ^v.

Lemma C.2 *Keep the notations in Definition* C.1*. Then, in normal coordinates (x^i) we have*

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \mathbf{E}\left[\left(X_\epsilon^{v,\varsigma}\right)^i\left(\tau_\epsilon^\varsigma(s)\right)\big|\mathcal{P}_{\tau(s)}\right] = \left(\Gamma(X)_{\tau(0)}^{\tau(s)} v(\tau(s))\right)^i + \varsigma(s)(D_\nabla X)^i(\tau(s)).$$

Proof Without loss of generality, we assume $\tau_\epsilon^\varsigma(s) \geq \tau(s)$. It follows from (C.1) and Definition 2.5 that

$$
\begin{aligned}
\mathrm{LHS} &= \lim_{\epsilon \to 0} \mathbf{E}\left[\frac{\left(X_\epsilon^{v,\varsigma}\right)^i\left(\tau_\epsilon^\varsigma(s)\right) - X^i(\tau(s))}{\epsilon}\bigg|\mathcal{P}_{\tau(s)}\right] \\
&= \lim_{\epsilon \to 0} \mathbf{E}\left[\frac{\left(X_\epsilon^{v,\varsigma}\right)^i\left(\tau_\epsilon^\varsigma(s)\right) - X^i\left(\tau_\epsilon^\varsigma(s)\right)}{\epsilon}\bigg|\mathcal{P}_{\tau(s)}\right] \\
&\quad + \lim_{\epsilon \to 0} \mathbf{E}\left[\frac{X^i\left(\tau_\epsilon^\varsigma(s)\right) - X^i(\tau(s))}{\tau_\epsilon^\varsigma(s) - \tau(s)}\bigg|\mathcal{P}_{\tau(s)}\right]\varsigma(s) \\
&= \mathrm{RHS}.
\end{aligned}
$$

Done.   □

Theorem C.3 (Stochastic Maupertuis's principle) *Let L_0 be a regular Lagrangian on $\mathbb{R} \times TM$. Let $X \in I_{(0,q)}^{(T,\mu)}(M)$ such that $(X, \mathbf{Id}_{[0,T]}) \in \mathcal{A}_g([0, T]; q, \mu; e)$. Then, the pair $(X, \mathbf{Id}_{[0,T]})$ is a stationary point of $\mathcal{I}$ if and only if X satisfy the stochastic Euler–Lagrange equation* (7.22).

Proof Since all diffusions in $\mathcal{A}_g([0, T]; q, \mu; e)$ have the same average energy e, we have

$$\mathcal{I}[X, \tau] := \mathbf{E}\int_{\tau(0)}^{\tau(T)} [L_0(t, X(t), D_\nabla X(t)) + e]\mathrm{d}t.$$

Denote $V(t) = \Gamma(X)_0^t v(t)$. As in (7.23),

$$
\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} I[X_\epsilon^{v,\varsigma}, \tau_\epsilon^\varsigma]
$$

$$
= \mathbf{E} \int_0^T \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} L_0\left(t, X_\epsilon^{v,\varsigma}(t), D_\nabla X_\epsilon^{v,\varsigma}(t)\right) \mathrm{d}t + \varsigma(t)\mathbf{E}[L_0\left(t, X(t), D_\nabla X(t)\right) + e]\big|_0^T
$$

$$
= \mathbf{E} \int_0^T \left[d_x L_0\left(V(t)\right) + d_{\dot{x}} L_0\left(\Gamma(X)_0^t \dot{v}(t)\right) + \frac{1}{2}(QX)^{ij}(t) d_{\dot{x}} L_0\left(R(V(t), \partial_i)\partial_j\right) \right] \mathrm{d}t
$$

$$
+ \varsigma(t)\mathbf{E}[L_0\left(t, X(t), D_\nabla X(t)\right) + e]\big|_0^T.
$$

We apply (7.24) and notice that in the present situation we do not have $v(0) = v(T) = 0$ in general. Hence,

$$
\mathbf{E} \int_0^T d_{\dot{x}} L_0\left(\Gamma(X)_0^t \dot{v}(t)\right) \mathrm{d}t = \mathbf{E} \int_0^T \Gamma(X)_t^0 (d_{\dot{x}} L_0)\left(\dot{v}(t)\right) \mathrm{d}t
$$

$$
= \mathbf{E}[d_{\dot{x}} L_0\left(V(t)\right)]\big|_0^T - \mathbf{E} \int_0^T \frac{\mathbf{D}}{dt}(d_{\dot{x}} L_0)\left(V(t)\right) \mathrm{d}t.
$$

One the other hand, since for all ϵ, $X_\epsilon^{v,\varsigma}(\tau_\epsilon^\varsigma(0)) = q$ and $\mathbf{P} \circ (X_\epsilon^{v,\varsigma}(\tau_\epsilon^\varsigma(T)))^{-1} = \mu$. It follows from Lemma C.2 that

$$
V(s) + \varsigma(s) D_\nabla X(s) = 0, \quad \text{for } s = 0 \text{ or } s = T.
$$

Therefore,

$$
\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} I\left[X_\epsilon^{v,\varsigma}, \tau_\epsilon^\varsigma\right] = \mathbf{E} \int_0^T \left(d_x L_0 - \frac{\overline{\mathbf{D}}}{dt}(d_{\dot{x}} L_0) \right)\left(V(t)\right) \mathrm{d}t
$$

$$
+ \varsigma(t)\mathbf{E}\left[L_0\left(t, X(t), D_\nabla X(t)\right) - (d_{\dot{x}} L_0)\left(D_\nabla X(t)\right) + e \right]\big|_0^T.
$$

By the definition of the energy E_0, we know that

$$
\mathbf{E}\left[L_0\left(t, X(t), D_\nabla X(t)\right) - (d_{\dot{x}} L_0)\left(D_\nabla X(t)\right)\right] = -\mathbf{E}E_0\left(t, X(t), D_\nabla X(t)\right) = -e.
$$

The result follows. $\square$

References

Abraham, R., Marsden, J.E.: Foundations of Mechanics, 2nd edn. Addison-Wesley Publishing Company, Boston (1978)

Albeverio, S., Yasue, K., Zambrini, J.-C.: Euclidean quantum mechanics: analytical approach. Ann. l'IHP Physique théorique **50**, 259–308 (1989)

Albeverio, S., Rezende, J., Zambrini, J.-C.: Probability and quantum symmetries. II. The theorem of Noether in quantum mechanics. J. Math. Phys. **47**(6), 062107 (2006)

Angst, J., Bailleul, I., Tardif, C.: Kinetic Brownian motion on Riemannian manifolds. Electron. J. Probab. **20**, 1–40 (2015)

Arnaudon, M., Thalmaier, A.: Complete lifts of connections and stochastic Jacobi fields. J. Math. Pures Appl. **77**(3), 283–315 (1998)

Arnaudon, M., Chen, X., Cruzeiro, A.B.: Stochastic Euler–Poincaré reduction. J. Math. Phys. **55**(8), 081507 (2014)

Arnold, V.I.: Mathematical Methods of Classical Mechanics, vol. 60, 2nd edn. Springer, New York (1989)

Arnold, V.I., Khesin, B.A.: Topological Methods in Hydrodynamics, vol. 125, 2nd edn. Springer, Cham (2021)

Asorey, M., Carinena, J.F., Ibort, L.A.: Generalized canonical transformations for time-dependent systems. J. Math. Phys. **24**(12), 2745–2750 (1983)

Belopolskaya, Y.I., Dalecky, Y.L.: Stochastic Equations and Differential Geometry. Kluwer Academic Publishers, Amsterdam (1990)

Bernstein, S.: Sur les liaisons entre les grandeurs aléatoires. Verh. Int. Math. Kongr. Zurich, Band I (1932)

Bismut, J.-M.: Mécanique Aléatoire, vol. 866. Springer, Berlin Heidelberg (1981)

Çetin, U., Danilova, A.: Markov bridges: SDE representation. Stoch. Process. Appl. **126**(3), 651–679 (2016)

Chen, X., Cruzeiro, A.B., Ratiu, T.S.: Stochastic variational principles for dissipative equations with advected quantities. J. Nonlinear Sci. **33**(1), 5 (2023)

Chung, K.L., Zambrini, J.-C.: Introduction to Random Time and Quantum Randomness, vol. 1. World Scientific, Singapore (2003)

Cruzeiro, A.B., Vuillermot, P.-A.: Forward-backward stochastic differential equations generated by Bernstein diffusions. Stoch. Anal. Appl. **33**(1), 91–109 (2015)

Cruzeiro, A.B., Zambrini, J.-C.: Malliavin calculus and Euclidean quantum mechanics. I. Functional calculus. J. Funct. Anal. **96**(1), 62–95 (1991)

Cruzeiro, A.B., Wu, L., Zambrini, J.-C.: Bernstein processes associated with a Markov process. In: Stochastic Analysis and Mathematical Physics, pp. 41–72. Springer (2000)

Dahlqvist, A., Diehl, J., Driver, B.K.: The parabolic Anderson model on Riemann surfaces. Probab. Theory Relat. Fields **174**(1), 369–444 (2019)

Dirac, P.A.M.: The Lagrangian in quantum mechanics. Phys. Z. Sowjetunion Band **3**(Heft 1), 64–72 (1933)

Dohrn, D., Guerra, F.: Geodesic correction to stochastic parallel displacement of tensors. In: Stochastic Behavior in Classical and Quantum Hamiltonian Systems, pp. 241–249. Springer (1979)

Driver, B.K.: A Cameron–Martin type quasi-invariance theorem for Brownian motion on a compact Riemannian manifold. J. Funct. Anal. **110**(2), 272–376 (1992)

Dynkin, E.B.: Diffusion of tensors. In: Doklady Akademii Nauk SSSR, vol. 179, pp. 1264–1267. Russian Academy of Sciences (1968)

Elworthy, K.D.: Stochastic Differential Equations on Manifolds, vol. 70. Cambridge University Press, Cambridge (1982)

Emery, M.: Stochastic Calculus in Manifolds. Springer, Berlin, Heidelberg (1989)

Emery, M.: An invitation to second-order stochastic differential geometry. hal-00145073 (2007)

Fang, S., Malliavin, P.: Stochastic analysis on the path space of a Riemannian manifold: I. Markovian stochastic calculus. J. Funct. Anal. **118**(1), 249–274 (1993)

Feynman, R.P.: Space–time approach to non-relativistic quantum mechanics. Rev. Mod. Phys. **118**, 367–387 (1948)

Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions, vol. 25, 2nd edn. Springer, Berlin (2006)

Fock, V.A.: Fundamentals of Quantum Mechanics, 2nd edn. Mir Publishers, Moscow (1978)

Gaeta, G., Quintero, N.R.: Lie-point symmetries and stochastic differential equations. J. Phys. A Math. Gen. **32**(48), 8485–8505 (1999)

Geiges, H.: An Introduction to Contact Topology, vol. 109. Cambridge University Press, Cambridge (2008)

Gentil, I., Léonard, C., Ripani, L.: Dynamical aspects of the generalized Schrödinger problem via Otto calculus-a heuristic point of view. Rev. Mat. Iberoam. **36**(4), 1071–1112 (2020)

Gliklikh, Y.E.: Global and Stochastic Analysis with Applications to Mathematical Physics. Springer, London (2011)

Goldstein, H., Poole, C., Safko, J.: Classical Mechanics, 3rd edn. Pearson Education, Pearson (2002)

Hairer, M.: A theory of regularity structures. Invent. Math. **198**(2), 269–504 (2014)

Haussmann, U.G.: A Stochastic Maximum Principle for Optimal Control of Diffusions, vol. 151. Longman Scientific and Technical (1986)

Holm, D.D., Schmah, T., Stoica, C.: Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions, vol. 12. Oxford University Press, Oxford (2009)

Hsu, P.: Brownian bridges on Riemannian manifolds. Probab. Theory Relat. Fields **84**(1), 103–118 (1990)

Hsu, E.P.: Quasi-invariance of the Wiener measure on the path space over a compact Riemannian manifold. J. Funct. Anal. **134**, 417–450 (1995)

Hsu, E.P.: Stochastic Analysis on Manifolds, vol. 38. American Mathematical Society, Providence (2002)

Huang, Q., Zambrini, J.-C.: Stochastic geometric mechanics in nonequilibrium thermodynamics: Schrödinger meets Onsager. J. Phys. A Math. Theor. **56**(13), 134003 (2023)

Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, vol. 24, 2nd edn. North-Holland Publishing Company, Amsterdam (1989)

Itô, K.: The Brownian motion and tensor fields on Riemannian manifold. In: Proceedings of the International Congress of Mathematicians 1962, pp. 536–539. Almqvist & Wiksells (1962)

Itô, K.: Stochastic parallel displacement. In: Probabilistic Methods in Differential Equations, pp. 1–7. Springer (1975)

Jamison, B.: The Markov processes of Schrödinger. Z. Wahrscheinlichkeitstheorie Verw. Gebiete **32**(4), 323–331 (1975)

Jost, J.: Riemannian Geometry and Geometric Analysis, 7th edn. Springer, Berlin (2017)

Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, vol. 113. Springer, New York (1991)

Khesin, B., Misiołek, G., Modin, K.: Geometric hydrodynamics and infinite-dimensional Newton's equations. Bull. Am. Math. Soc. **58**(3), 377–442 (2021)

Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, vol. 1. Interscience Publishers, Geneva (1963)

Lang, S.: Fundamentals of Differential Geometry, vol. 191. Springer, Berlin (1999)

Lassalle, R., Zambrini, J.-C.: A weak approach to the stochastic deformation of classical mechanics. J. Geom. Mech. **8**(2), 221 (2016)

Lázaro-Camí, J.-A., Ortega, J.-P.: Stochastic Hamiltonian dynamical systems. Rep. Math. Phys. **61**(1), 65–122 (2008)

Lázaro-Camí, J.-A., Ortega, J.-P.: The stochastic Hamilton–Jacobi equation. J. Geom. Mech. **1**(3), 295 (2009)

Lee, J.M.: Introduction to Smooth Manifolds, vol. 218, 2nd edn. Springer, New York (2013)

Léonard, C.: Girsanov theory under a finite entropy condition. In: Séminaire de Probabilités XLIV, pp. 429–465. Springer (2012a)

Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. J. Funct. Anal. **262**(4), 1879–1920 (2012b)

Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. Discrete Contin. Dyn. Syst. **34**(4), 1533–1574 (2014)

Léonard, C., Rœlly, S., Zambrini, J.-C.: Reciprocal processes: a measure-theoretical point of view. Probab. Surv. **11**, 237–269 (2014)

Lescot, P., Zambrini, J.-C.: Probabilistic deformation of contact geometry, diffusion processes and their quadratures. In: Seminar on Stochastic Analysis, Random Fields and Applications V, vol. 59, pp. 203–226. Springer (2007)

Li, X.-M.: Limits of random differential equations on manifolds. Probab. Theory Relat. Fields **166**(3), 659–712 (2016)

Malliavin, P.: Stochastic Analysis, vol. 313. Springer, Berlin, Heidelberg (1997)

Marsden, J.E., Ratiu, T.S.: Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems, vol. 17, 2nd edn. Springer, Berlin (1999)

Meyer, P.-A.: Formes differentielles d'ordre $n > 1$. Publication IRMA, Université Louis Pasteur, Strasbourg, 80 (1979)

Meyer, P.-A.: A differential geometric formalism for the Itô calculus. In: Stochastic Integrals, vol. 851 of LNM, pp. 256–270. Springer (1981a)

Meyer, P.-A: Géométrie stochastique sans larmes. In: Séminaire de Probabilités XV 1979/80, pp. 44–102. Springer (1981b)

Mikami, T.: Stochastic Optimal Transportation: Stochastic Control with Fixed Marginals. Springer, Berlin (2021)

Munkres, J.R.: Topology, 2nd edn. Prentice Hall Inc, Hoboken (1975)

Nelson, E.: Dynamical Theories of Brownian Motion, vol. 106, 2nd edn. Princeton University Press, Princeton (2001)

Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer, Berlin Heidelberg (2010)

Olver, P.J.: Equivalence, Invariants and Symmetry. Cambridge University Press, Cambridge (1995)

Olver, P.J.: Applications of Lie Groups to Differential Equations, vol. 107, 2nd edn. Springer, New York (1998)

Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**(1–2), 107–174 (2001)

Petersen, P.: Riemannian Geometry, vol. 171, 3rd edn. Springer, Berlin (2016)

Peyré, G., Chizat, L., Vialard, F.-X., Solomon, J.: Quantum entropic regularization of matrix-valued optimal transport. Eur. J. Appl. Math. **30**(6), 1079–1102 (2019)

Privault, N., Yang, X., Zambrini, J.-C.: Large deviations for Bernstein bridges. Stoch. Process. Appl. **126**(5), 1285–1305 (2016)

Saunders, D.J.: The Geometry of Jet Bundles, vol. 142. Cambridge University Press, Cambridge (1989)

Schrödinger, E.: Quantization as a problem of proper values (part I). Ann. Phys. **101**, 25–32 (1926)

Schrödinger, E.: Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. Ann. l'inst. Henri Poincaré **2**, 269–310 (1932)

Schwartz, L.: Semi-martingales sur des Variétés, et Martingales Conformes sur des Variétés Analytiques Complexes, vol. 780. Springer-Verlag, Berlin, Heidelberg (1980)

Schwartz, L.: Géométrie différentielle du 2 ème ordre, semi-martingales et équations différentielles stochastiques sur une variété différentielle. In: Séminaire de Probabilités XVI, 1980/81 Supplément: Géométrie Différentielle Stochastique, pp. 1–148. Springer (1982)

Schwartz, L.: Semimartingales and their stochastic calculus on manifolds. Gaetan Morin Editeur Ltee (1984)

Thieullen, M., Zambrini, J.-C.: Probability and quantum symmetries I: The theorem of Noether in Schrödinger's Euclidean quantum mechanics. Ann. Inst. Henri Poincaré **67**(3), 297–338 (1997)

Trachenko, K., Brazhkin, V.V.: The quantum mechanics of viscosity. Phys. Today **74**(12), 66–67 (2021)

Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Berlin, Heidelberg (2009)

von Renesse, M.-K.: An optimal transport view of Schrödinger's equation. Can. Math. Bull. **55**(4), 858–869 (2012)

Yong, J., Zhou, X.Y.: Stochastic Controls: Hamiltonian Systems and HJB Equations, vol. 43. Springer, New York (1999)

Zambrini, J.-C.: Variational processes and stochastic versions of mechanics. J. Math. Phys. **27**(9), 2307–2330 (1986)

Zambrini, J.-C.: The research program of stochastic deformation (with a view toward geometric mechanics). In: Stochastic Analysis: A Series of Lectures, vol. 68 of Progress in Probability, pp. 359–393. Springer, Basel (2015)