An Invitation to Second-Order Stochastic Differential Geometry

Michel Émery

IRMA
7 rue René Descartes
67084 Strasbourg Cedex
France
emery@math.u-strasbg.fr

> Iusques icy i'ay tasché de me rendre intelligible a tout le monde, mais pour ce traité ie crains, qu'il ne pourra estre leu que par ceux, qui sçauent desia ce qui est dans les liures de Geometrie.
>
> René Descartes. *La Geometrie.*

These notes are an expanded version of a set of lectures given at the 2005 Dimitsana Summer School; the added material consists of a few remarks or exercises, and of the proofs of some statements. Many proofs are not complete: the localization of manifold-valued semimartingales, and many other technicalities, have been omitted. Anyway, all proofs can be skipped; contrary to a general rule in mathematics, all important ideas in these notes are better expressed in the statements of the propositions and theorems than in their proofs. Another deviation from the oral lectures is a change in the order of the material: all sections having to do with Stratonovich integrals and Stratonovich stochastic differential equations have been grouped; intrinsic Itô integrals and Itô stochastic differential equations, which are far less useful and important, are put at the end.

I thank the organizers of the summer school for this opportunity to propagandize Schwartz' view of stochastic differential geometry; I hope these notes will contribute to popularize ideas which deserve to be more broadly known, not only for their intrinsic beauty, but also for the interesting light they shed on stochastic calculus in manifolds. I also thank the participants for their remarks and questions.

Introduction

Differential geometry deals with (local or global) properties of many differentiable structures, and among them with the behaviour of manifold-valued differentiable curves; its most basic tools are general theorems such as the chain-rule (how to differentiate a compound function between manifolds), uniqueness and local existence of the flow of a locally Lipschitz vector field, parallel transport and more general lifts of curves in a bundle via a connection, and so on.

Similarly, stochastic calculus also has its basic toolkit: chain rule with Itô's second-order terms, consistency of Stratonovich stochastic calculus, existence and uniqueness for stochastic differential equations with Lipschitz coefficients, and much more.

These geometric and stochastic tools are made powerful by their very generality; like a Swiss card-knife, one can resort to them in quite a few different situations.

Combining both theories, stochastic differential geometry mostly deals with continuous, but non-smooth random curves taking their values in manifolds. It is also grounded on very general, powerful tools; the most popular one is the Stratonovich transfer principle: "geometric constructions performed on differential curves extend intrinsically to random curves by replacing ordinary differentials with Stratonovich stochastic differentials". This universal recipe is a cornerstone of stochastic differential geometry, and one of the aims of these lectures is to present it; but I could not resist seizing this opportunity to lecture on a more general, albeit less popular, stochastic calculus in manifolds devised by Schwartz around 1980. Whereas Stratonovich calculus works by killing the second-order Itô correction terms of stochastic calculus, Schwartz' *second-order stochastic calculus in manifolds* takes the opposite view by incorporating these second-order terms into a more complete formalism which gives a broader picture. In my opinion, it also provides a better understanding on the nature (and limitations) of the Stratonovich transfer principle, which is included in this more general setting as its most important application.

It should be clear that second-order stochastic calculus in differentiable manifolds, as presented in these lectures, is a convenient language, a useful set of tools, but by no means a theory. One usually considers that true differential geometry starts only after curvature has been introduced (see Chapters ??? and ???); similarly, stochastic differential geometry really begins with the relationship between Brownian motion and Ricci curvature (see Chapter ???). But we shall stop short of reaching that level: only at the very end will Brownian motion in manifolds show up, and, although affine connections play a significant rôle, curvature will not even be defined.

Tangent and cotangent spaces and bundles

By a *manifold*, we mean a finite-dimensional, real differentiable manifold without boundary, having a countable atlas. The examples one can have in mind are any open subset of $\mathbb{R}^d$, or any submanifold of $\mathbb{R}^n$ (for instance, a curve or a surface in $\mathbb{R}^3$). The latter is in fact the most general manifold: Whitney's embedding theorem says that any manifold is always diffeomorphic to some closed submanifold of $\mathbb{R}^n$ for a suitable n (for instance, a Klein bottle can be embedded into $\mathbb{R}^4$).

Given a d-dimensional manifold M, a *local chart* is a system of d real functions $x^1, \dots, x^d$ (the *local coordinates*) defined in some open subset $D \subset M$ (called the domain of the chart), realizing a diffeomorphism between D and some open subset of $\mathbb{R}^d$. Every point of M has a neighbourhood which is the domain of some local chart; indeed, the definition of a manifold consists in demanding that M is coverable with countably many such domains, any two local charts being compatible (that is, they induce the same differentiable structure) on the intersection of their domains.

If $(x^1, \dots, x^d)$ is a local chart and f a function on M, we shall call $\mathrm{D}_k f$ the function defined on the domain of the chart and equal to the partial derivative of f with respect to the coordinate x^k when the other coordinates are kept constant. (Rigorously, $\mathrm{D}_k f(x) = \frac{\partial g}{\partial \xi^k}\big(x^1(x), \dots, x^d(x)\big)$, where g is defined on the range of the chart and $g \circ (x^1, \dots, x^d) = f$ on the domain of the chart.) Higher derivatives are similarly denoted: $\mathrm{D}_{ij} f = \mathrm{D}_i \mathrm{D}_j f$, etc.

If x is a point in a manifold M, a *tangent vector* to M at x can be defined in two equivalent ways: as an equivalence class of curves, or as a differential operator on functions. The former definition considers two C^1-curves γ^1 and $\gamma^2 : \mathbb{R} \to M$ such that $\gamma^1(t_0) = \gamma^2(t_0) = x$ to be equivalent if, for each C^1 function $f : M \to \mathbb{R}$, the derivative of $f \circ \gamma^1(t) - f \circ \gamma^2(t)$ vanishes at $t = t_0$. Both curves are then said to have the same *velocity* at $t = t_0$, and this velocity (which can abstractly be defined as the equivalence class) is denoted by $\dot\gamma(t_0)$ and is the most general tangent vector to M at point x. The other point of view, that of operators, defines the tangent vector $\dot\gamma(t_0)$ as the map $f \mapsto (f \circ \gamma)'(t_0)$. This map is a differential operator at point $\gamma(t_0) = x$, of order 1 and with no constant term, acting on C^1 functions $f : M \to \mathbb{R}$. Conversely, it is easily seen that such an operator is always of the form $\dot\gamma(t_0)$ for some curve γ such that $\gamma(t_0) = x$; so the (linear) space $\mathrm{T}_x M$ of all tangent vectors to M at x can conveniently be defined as the space of all first-order differential operators at x with no constant term.

Exercise. a) When M is a d-dimensional vector space, verify that both definitions of $\mathrm{T}_x M$ agree, and that $\mathrm{T}_x M$ is canonically isomorphic to M.

b) For a general manifold M with dimension d, using a local chart in a neighbourhood of x, show that both definitions of $\mathrm{T}_x M$ agree and that the d maps $f \mapsto \mathrm{D}_i f(x)$ form a basis of the vector space $\mathrm{T}_x M$.

The disjoint union $\mathrm{T}M = \bigcup_{x\in M} \mathrm{T}_x M$ is called the *tangent bundle*; each of its subsets $\mathrm{T}_x M$ is a *fibre* of the bundle. A point $y \in \mathrm{T}M$ is said to be *above* x if y belongs to the fibre $\mathrm{T}_x M$. The tangent bundle $\mathrm{T}M$ is itself a manifold (of class C^{p-1} if M is C^p); more precisely, it has the structure of a vector bundle over M, as explained in Chapter ???.

Like tangent vectors, which can be defined in two different ways, *vector fields*, too, admit equivalent definitions, which are recalled in the next proposition.

Proposition 1. *Let M be a manifold of class C^p with $1 \leqslant p \leqslant \infty$, and V be a real-valued function on $M \times \mathrm{C}^p(M)$; denote by $V(x)$ the map $V(x, .)$ and by Vf the function $V(\,.\,, f)$. The following three statements are equivalent:*

(i) for each $x \in M$, $V(x)$ is a vector in the fibre $\mathrm{T}_x M$ and $x \mapsto V(x)$ is of class C^{p-1};

(ii) for every local chart $(x^1, \dots, x^d)$ with domain D, there exist d functions $V^1, \dots, V^d$ in $\mathrm{C}^{p-1}(D)$, such that, for each $x \in D$ and each $f \in \mathrm{C}^p(M)$,

$$V(x, f) = \sum_{k=1}^{d} V^k(x)\, \mathrm{D}_k f(x)\ ;$$

(iii) $f \mapsto Vf$ is a linear map from $\mathrm{C}^p(M)$ to $\mathrm{C}^{p-1}(M)$ and, for all n, all $f^1, \dots, f^n \in \mathrm{C}^p(M)$ and all $\phi \in \mathrm{C}^p(\mathbb{R}^n, \mathbb{R})$,

$$V\big(\phi \circ (f^1, \dots, f^n)\big) = \sum_{q=1}^{n} \mathrm{D}_q \phi \circ (f^1, \dots, f^n)\, V f^q\ .$$

If furthermore $p = \infty$, the next two statements are equivalent to each other and to the preceding three ones:

(iv) $V : \mathrm{C}^\infty(M) \to \mathrm{C}^\infty(M)$ is linear, and satisfies $V(f^2) = 2f\,Vf$ for all $f \in \mathrm{C}^\infty(M)$;

(v) $V : \mathrm{C}^\infty(M) \to \mathrm{C}^\infty(M)$ is linear, $V1 = 0$, and $V(f^2)(x) = 0$ for all $f \in \mathrm{C}^\infty(M)$ and all $x \in M$ such that $f(x) = 0$.

The proof is left an an exercise. (Hint: If $f \in \mathrm{C}^p(M)$ vanishes on a neighbourhood of some point x, there exists $g \in \mathrm{C}^p(M)$ such that $g = 1$ on a neighbourhood of x and $fg = 0$. Use this property to show that any V satisfying (iii), (iv) or (v) is local: if $f^1 = f^2$ in a neighbourhood of x, then $Vf^1(x) = Vf^2(x)$.) Property (i) says that V is a section of the tangent bundle, (iii) says that V is a differential operator of order 1 with no constant term, and (iv) that V is a derivation.

Exactly like when $M = \mathbb{R}^d$, the *flow* of a (locally Lipschitz) vector field can be defined, locally in time and space: some integral curves of the vector field may explode (i.e., exit from all compact sets) in finite time.

Another important vector bundle is the *cotangent bundle*, called T^*M. It is defined as $\mathrm{T}^*M = \bigcup_{x\in M} \mathrm{T}^*_x M$, where $\mathrm{T}^*_x M$ is the *cotangent space* at x, that is, the dual to the tangent space $\mathrm{T}_x M$. The elements of T^*M (resp. $\mathrm{T}^*_x M$) are called *cotangent vectors* (resp. *cotangent vectors at* x). If $f \in \mathrm{C}^1(M)$, the cotangent vector $\mathrm{d}f(x) \in \mathrm{T}^*_x M$ is defined by its action on tangent vectors:

$$<\mathrm{d}f(x), V> = Vf(x) \ .$$

(Here and in the sequel, we use the sharp brackets $<\ ,\ >$ to denote the duality bilinear form pairing a vector space and its dual.)

If $(x^1, \ldots, x^d)$ is a local chart in a domain containing x, the d cotangent vectors $\mathrm{d}x^i(x)$ form a basis of the cotangent space $\mathrm{T}^*_x M$; every $\sigma \in \mathrm{T}^*_x M$ can be written $\sigma = \sum_{k=1}^d \sigma_k \, \mathrm{d}x^k(x)$, where the d real numbers σ_k are the coefficients of σ in the local chart. The duality with a tangent vector $V \in \mathrm{T}_x M$ given by $Vf = \sum_k V^k \mathrm{D}_k f(x)$ is expressed by $<\sigma, V> = \sum_k \sigma_k V^k$.

When x varies, the cotangent vectors $\mathrm{d}f(x)$ form the cotangent vector field $\mathrm{d}f$ such that $<V, \mathrm{d}f> = Vf$ for all vector fields V. Cotangent vector fields are usually called 1-forms; this stands for 'forms of degree 1'. (In view of the second order geometry about to be introduced, they could be more precisely called forms of degree 1 and order 1.)

Exercise. Given $x \in M$ and $\sigma \in \mathrm{T}^*_x M$, there always exists a function f on M such that $\sigma = \mathrm{d}f(x)$. This holds for individual covectors, but not for covector fields (= 1-forms): there may exist 1-forms which can not be written $\mathrm{d}f$ for any f. (In fact, the only exception is when all connected components of M are diffeomorphic to $\mathbb{R}$.).

Exercise. Let M and N be two manifolds, $\phi : M \to N$ a map of class C^1 at least, and x a point in M. Define the *tangent map* $\phi_{*x} : \mathrm{T}_x M \to \mathrm{T}_{\phi(x)} N$, also denoted by $\mathrm{T}_x \phi$, and the *cotangent map* $\phi^*_x : \mathrm{T}^*_{\phi(x)} N \to \mathrm{T}^*_x M$, also written $\mathrm{T}^*_x \phi$. They are adjoint to each other; ϕ_{*x} is also called the *push-forward* of vectors, and ϕ^*_x the *pull-back* of forms. If σ is a 1-form on N, define its pull-back $\phi^*\sigma$, a 1-form on M. Explain why, in general, the push-forward $\phi_* V$ of a vector field V on M cannot be defined as a vector field on N.

Second order tangent and cotangent spaces and bundles

From now on, all manifolds are of class C^p with $p \geqslant 2$.

Definition. Let M be a manifold and x a point in M. The *second-order tangent space to* M *at* x is the vector space $\mathbb{T}_x M$ consisting of all second-order differential operators at x with no constant term. The elements of $\mathbb{T}_x M$ will be called *diffusors* (the names *second-order tangent vectors* and *tangent vectors of order 2* are more frequent in the literature).

Given a local chart $(x^1,\dots,x^d)$ in a neighbourhood of x, every diffusor $L\in\mathbb{T}_xM$ is given by an expression of the form

$$Lf=\sum_{i=1}^d\sum_{j=1}^d L^{ij}\mathrm{D}_{ij}f(x)+\sum_{k=1}^d L^k\mathrm{D}_kf(x)\ ,$$

where $f\in\mathrm{C}^2(M)$ and L^{ij} and L^k are real coefficients. Observe that, because $\mathrm{D}_{ij}f(x)=\mathrm{D}_{ji}f(x)$, the vector space $\mathbb{T}_xM$ has dimension $\frac{1}{2}d(d{+}1)+d$, and not d^2+d: there is some arbitrariness in the choice of the coefficients L^{ij}. One can for instance demand that $L^{ij}=L^{ji}$; in that case, all coefficients are uniquely defined by L.

Exercise. The expression of L in another local chart $(y^\alpha,\ \alpha\in\{1,\dots,d\})$ whose domain also contains x is $Lf=\sum_{\alpha\beta}L^{\alpha\beta}\mathrm{D}_{\alpha\beta}f(x)+\sum_\gamma L^\gamma\mathrm{D}_\gamma f(x)$, with

$$L^{\alpha\beta}=\sum_{ij}L^{ij}\,\mathrm{D}_iy^\alpha(x)\,\mathrm{D}_jy^\beta(x)$$

$$L^\gamma=Ly^\gamma=\sum_{ij}L^{ij}\mathrm{D}_{ij}y^\gamma(x)+\sum_k L^k\mathrm{D}_ky^\gamma(x)\ .$$

Deduce therefrom that the decomposition of L into its “first-order part” $\sum L^k\mathrm{D}_k$ and “purely second-order” part $\sum L^{ij}\mathrm{D}_{ij}$ depends upon the choice of the chart: it has no intrinsic meaning, even though the vector space $\mathbb{T}_xM$ has a non-trivial intrinsic structure (this will be seen in Proposition 5).

The first example of diffusor is a vector. Clearly, $\mathrm{T}_xM\subset\mathbb{T}_xM$ (but the first-order part of an $L\in\mathbb{T}_xM$ does not exist, as seen in the preceding exercise; in other words, there exists in $\mathbb{T}_xM$ no canonical subspace supplementary to T_xM).

Another important example of a diffusor is the acceleration of a curve. If $\gamma:\mathbb{R}\to M$ is a C^2 curve, its acceleration at $t=t_0$ is the diffusor $\ddot\gamma(t_0)\in\mathbb{T}_{\gamma(t_0)}M$ equal to the second-order differential operator $f\mapsto(f\circ\gamma)''(t_0)$.

Exercise. a) In a local chart around $\gamma(t_0)$, denoting by $\gamma^i(t)$ the coordinates of the point $\gamma(t)$, the acceleration $L=\ddot\gamma(t_0)$ is given by $L^{ij}=(\gamma^i)'(t_0)\,(\gamma^j)'(t_0)$ and $L^k=(\gamma^k)''(t_0)$. (It may come as a surprise that the second derivatives appear only in the coefficients L^k, which one could be tempted to consider as “first order”.)

b) Fix $x\in M$. When γ ranges over all C^2 curves such that $\gamma(t_0)=x$, the vectors $\dot\gamma(t_0)$ range over all the tangent space T_xM and the diffusors $\ddot\gamma(t_0)$ linearly span the second-order tangent space $\mathbb{T}_xM$.

c) Still for fixed x, when γ ranges over all C^2 curves such that $\gamma(t_0)=x$ and $\dot\gamma(t_0)=0$, the diffusors $\ddot\gamma(t_0)$ are vectors, and they range over the whole tangent space T_xM. Conversely, for a curve γ such that $\gamma(t_0)=x$, the diffusor $\ddot\gamma(t_0)$ is a vector if and only if $\dot\gamma(t_0)=0$.

The push-forward of diffusors is similar to that of vectors: if $\phi : M \to N$ is at least C^2, the linear map $\phi_{*x} : \mathbb{T}_x M \to \mathbb{T}_{\phi(x)} N$ (also denoted by $\mathbb{T}_x\phi$) is defined by $\big(\phi_{*x} L\big) f = L(f\circ\phi)$. For instance, the acceleration $\ddot\gamma(t_0)$ of a curve γ is equal to the push-forward by γ_{*t_0} of the diffusor $f \mapsto f''(t_0)$, which is an element of $\mathbb{T}_{t_0}\mathbb{R}$.

Using the same notation ϕ_{*x} for the push-forward $\mathrm{T}_x\phi$ of vectors and $\mathbb{T}_x\phi$ of diffusors is harmless, because $\mathbb{T}_x\phi$ is equal to $\mathrm{T}_x\phi$ on the tangent space $\mathrm{T}_x M$. In particular, the push-forward of diffusors respects vectors (that is, $\phi_{*x}(T_x M) \subset \mathrm{T}_{\phi(x)} M$). This implies that not every linear map from $\mathbb{T}_x M$ to $\mathbb{T}_{\phi(x)} N$ is of the form ϕ_{*x}. (Besides stability of first-order vectors, another condition must be satisfied; this is postponed until Proposition 5.)

The disjoint union $\mathbb{T}M = \bigcup_{x\in M} \mathbb{T}_x M$ is the *second-order tangent bundle*; it is a vector bundle over M, to which the figurative language of fibre bundles applies ($\mathbb{T}_x M$ is a fibre, its elements are above x). Its sections are the *diffusor fields* (or *second-order vector fields*); the next proposition extends to second order the characterizations we saw earlier for vector fields.

Proposition 2. *Let M be a C^p-manifold with $2 \leqslant p \leqslant \infty$, and L a real function on $M \times \mathrm{C}^p(M)$; call $L(x)$ the map $L(x, .)$ and Lf the function $L(. , f)$. The following three statements are equivalent:*

(i) for each $x \in M$, $L(x)$ belongs to $\mathbb{T}_x M$ and $x \mapsto L(x)$ is of class C^{p-2};

(ii) for every local chart $(x^1, \ldots, x^d)$ with domain D, there exist functions $L^1, \ldots, L^d, L^{11}, L^{12}, \ldots, L^{dd}$ in $\mathrm{C}^{p-2}(D)$, such that, for each $x \in D$ and each $f \in \mathrm{C}^p(M)$,

$$L(x, f) = \sum_{ij} L^{ij}(x)\,\mathrm{D}_{ij} f(x) + \sum_k L^k(x)\,\mathrm{D}_k f(x)\ ;$$

(iii) $f \mapsto Lf$ is a linear map from $\mathrm{C}^p(M)$ to $\mathrm{C}^{p-2}(M)$ and, putting

$$\Gamma(f, g) = L(fg) - fLg - gLf\ ,$$

one has for all n, all $f^1, \ldots, f^n \in \mathrm{C}^p(M)$ and all $\phi \in \mathrm{C}^p(\mathbb{R}^n, \mathbb{R})$,

$$L\big(\phi\circ(f^1, \ldots, f^n)\big) = \sum_{q=1}^n \mathrm{D}_q\phi\circ(f^1, \ldots, f^n)\, Lf^q + \tfrac{1}{2}\sum_{r,s=1}^n \mathrm{D}_{rs}\phi\circ(f^1, \ldots, f^n)\,\Gamma(f^r, f^s)\ .$$

If moreover $p = \infty$, the further two statements below are also equivalent to the preceding three ones:

(iv) $L : \mathrm{C}^\infty(M) \to \mathrm{C}^\infty(M)$ is linear, and satisfies $L(f^3) = 3fL(f^2) - 3f^2Lf$ for all $f \in \mathrm{C}^\infty(M)$;

(v) $L : \mathrm{C}^\infty(M) \to \mathrm{C}^\infty(M)$ *is linear,* $L1 = 0$*, and* $L(f^3)(x) = 0$ *for all* $f \in \mathrm{C}^\infty(M)$ *and* $x \in M$ *such that* $f(x) = 0$*.*

The operator Γ featuring in (iii) is called the squared gradient (or 'carré du champ') associated to L; see § 1.3 of Chapter ???. The reason for this name is, when L is the Laplacian (in $\mathbb{R}^d$ or more generally in a Riemannian manifold M), $\Gamma(f,f)$ is equal to $2\,\|\nabla f\|^2$.

If V and W are vector fields, their product VW, obtained by composing two first-order differential operators, is a second-order one, that is, a diffusor field. Its squared gradient is given by $\Gamma(f,g) = Vf\,Wg + Wf\,Vg$.

Definitions. The *second-order cotangent space* to M at x is the dual $\mathbb{T}^*_x M$ of the second-order tangent space $\mathbb{T}_x M$. Its elements will be called *codiffusors* Another name in the literature is 'second-order forms' or 'forms of order 2'; we shall not use this name for individual codiffusors, but we reserve it for codiffusor fields (to be introduced later).

A fundamental example of a codiffusor is $\mathrm{d}^2 f(x)$, where $f : M \to \mathbb{R}$ is at least C^2; by definition, $\mathrm{d}^2 f(x)$ is the linear form on $\mathbb{T}_x M$ given by $L \mapsto Lf$. It is characterized by its action on the accelerations of curves:

$$<\mathrm{d}^2 f\big(\gamma(t_0)\big), \ddot\gamma(t_0)> = (f\circ\gamma)''(t_0)\ .$$

All codiffusors at x have the form $\mathrm{d}^2 f(x)$ for a suitable f. The notation $\mathrm{d}^2 f$ will be justified after Proposition 6.

Since $\mathrm{T}_x M$ is a linear subspace of $\mathbb{T}_x M$, to every codiffusor $\theta \in \mathbb{T}^*_x M$ one can associate its *restriction to first order* $\mathrm{R}\theta \in \mathrm{T}^*_x M$, defined as the restriction to $\mathrm{T}_x M$ of the map $\theta : \mathbb{T}_x M \to \mathbb{R}$. Thus, R is a canonical map from $\mathbb{T}^*_x M$ to $\mathrm{T}^*_x M$. (It is the adjoint map to the canonical injection $\mathrm{T}_x M \hookrightarrow \mathbb{T}_x M$.) As for an example, one has $\mathrm{R}\big(\mathrm{d}^2 f(x)\big) = \mathrm{d}f(x)$.

Exercise. Show that a codiffusor $\theta \in \mathbb{T}^*_x M$ verifies $\mathrm{R}\theta = 0$ if and only if $<\theta, \ddot\gamma(t_0)> = 0$ for all C^2 curves γ such that $\gamma(t_0) = x$ and $\dot\gamma(t_0) = 0$.

Proposition 3 and definition. *Given two covectors* σ *and* ρ *in* $\mathrm{T}^*_x M$*, there exists a unique codiffusor in* $\mathbb{T}^*_x M$*, denoted by* $\sigma{\cdot}\rho$ *and called the* product *of* σ *and* ρ*, such that for every* C^2 *curve* γ *with* $\gamma(t_0) = x$*, one has*

$$<\sigma{\cdot}\rho, \ddot\gamma(t_0)> = <\sigma, \dot\gamma(t_0)>\, <\rho, \dot\gamma(t_0)>\ .$$

This product : $\mathrm{T}^*_x M \times \mathrm{T}^*_x M \to \mathbb{T}^*_x M$ *is symmetric, bilinear, and verifies* $\mathrm{R}(\sigma{\cdot}\rho) = 0$*. Moreover, for all* f *and* g *in* $\mathrm{C}^2(M)$*,*

$$\mathrm{d}f(x)\cdot\mathrm{d}g(x) = \tfrac{1}{2}\,\big[\mathrm{d}^2(fg)(x) - f(x)\,\mathrm{d}^2 g(x) - g(x)\,\mathrm{d}^2 f(x)\big]\ .$$

Proof. Choose a local chart whose domain contains X; an arbitrary $L \in \mathbb{T}_x M$ operates on functions via the formula $Lf = \sum_{ij} L^{ij}\mathrm{D}_{ij} f(x) + \sum_k L^k \mathrm{D}_k f(x)$, where the coefficients L^{ij} are chosen symmetric in i and j. Call σ_i and ρ_i the coefficients of σ and ρ in the same local chart, and define $\sigma{\cdot}\rho \in \mathbb{T}^*_x M$ by

$$<\sigma\cdot\rho, L> = \sum_{ij} L^{ij}\sigma_i\rho_j \ .$$

It is easy to verify that $\sigma\cdot\rho$ has all the claimed properties; uniqueness stems from the fact that the diffusors $\ddot\gamma(t_0)$ linearly span $\mathbb{T}^*_x M$. Naturally, $\sigma\cdot\rho$ does not depend upon the choice of the local chart, because it is characterized by an intrinsic property. □

The product of covectors has nothing to do with an inner product (as given for instance by a Riemannian structure); it is not scalar, but valued in $\mathbb{T}_x M$. It is used to express codiffusors in local coordinates:

Exercise. Fix a local chart around x. Every $\theta \in \mathbb{T}^*_x M$ can be written

$$\sum_{ij} \theta_{ij}\, \mathrm{d}x^i(x)\cdot \mathrm{d}x^j(x) + \sum_k \theta_k\, \mathrm{d}^2 x^k(x) \ ,$$

where θ_{ij} and θ_k are real coefficients. One can always choose them such that $\theta_{ij} = \theta_{ji}$; in that case, they are uniquely determined by θ. The duality between codiffusors and diffusors is expressed by

$$<\theta, L> = \sum_{ij} \theta_{ij} L^{ij} + \sum_k \theta_k L^k \ ,$$

provided at least one of the systems of coefficients (L^{ij}) and (θ_{ij}) is symmetric. When $\theta = \mathrm{d}^2 f(x)$, the coefficients are $\theta_{ij} = \mathrm{D}_{ij} f(x)$ and $\theta_k = \mathrm{D}_k f(x)$. Non-symmetric coefficients for $\theta = \sigma\cdot\rho$ are $\theta_{ij} = \sigma_i\rho_j$ and $\theta_k = 0$.

Exercise. A codiffusor $\theta \in \mathbb{T}^*_x M$ is said to be *positive* if $<\theta, \ddot\gamma(t_0)> \geqslant 0$ for all C^2 curves γ such that $\gamma(t_0) = x$. How is positivity expressed in local coordinates? Show that if θ is positive, $\mathrm{R}\theta = 0$, and that $\sigma\cdot\sigma$ is positive for any covector $\sigma \in \mathrm{T}^*_x M$. Conversely, every positive codiffusor is a sum of at most d such "squares".

The *second-order cotangent bundle* $\mathbb{T}^* M$ is the disjoint union $\bigcup_{x\in M} \mathbb{T}^*_x M$ of all second-order cotangent spaces, which are its fibres. The sections of this bundle, that is, *codiffusor fields*, are also called *second-order forms*. They should not be confused with the 2-forms (which are forms of *degree* 2), ubiquitous in the geometric literature, and which we shall never encounter in these lectures. In local coordinates, a codiffusor field is expressed as $\sum_{ij} \theta_{ij}\, \mathrm{d}x^i\cdot \mathrm{d}x^j + \sum_k \theta_k\, \mathrm{d}^2 x^k$, where θ_{ij} and θ_k are functions defined on the domain of the chart.

The pull-back of codiffusors and of codiffusor fields is similar to that of covectors and of covector fields: given $\phi : M \to N$ which is at least C^2, the pull-back $\phi^*_x : \mathbb{T}^*_{\phi(x)} N \to \mathbb{T}^*_x M$ is defined as the adjoint of the push-forward $\phi_{*x} : \mathbb{T}_x M \to \mathbb{T}_{\phi(x)} N$; it is also denoted by $\mathbb{T}^*_x \phi$. If θ is a codiffusor field on N, its pull-back $\phi^*\theta$ is the codiffusor field on M whose value at x is $\phi^*_x\, \theta\big(\phi(x)\big)$.

(But diffusor fields cannot be pushed forward.) The pull-back of codiffusors and of codiffusor fields has functorial properties:

$$\phi^*(\mathrm{d}^2 f) = \mathrm{d}^2(f\circ\phi) \ ;$$
$$\phi^*(\mathrm{R}\theta) = \mathrm{R}(\phi^*\theta) \ ;$$
$$\phi^*(\sigma{\cdot}\rho) = (\phi^*\sigma)\cdot(\phi^*\rho) \ .$$

Complementing these formulae with the exact signification of each symbol is left as an exercise (for instance, both R are not defined on the same space). Their proofs are sketched below; we give only the skeleton of the arguments, the flesh is to be supplied by the reader.

$$<\phi^*(\mathrm{d}^2 f), L> = <\mathrm{d}^2 f, \phi_* L> = (\phi_* L) f = L(f\circ\phi) = <\mathrm{d}^2(f\circ\phi), L> \ ;$$
$$<\phi^*(\mathrm{R}\theta), V> = <\mathrm{R}\theta, \phi_* V> = <\theta, \phi_* V> = <\phi^*\theta, V> = <\mathrm{R}\phi^*\theta, V> \ ;$$
$$\begin{aligned} <\phi^*(\sigma{\cdot}\rho), \ddot\gamma> &= <\sigma{\cdot}\rho, \phi_*\ddot\gamma> = <\sigma{\cdot}\rho, (\phi\circ\gamma)^{\ddot{}}> \\ &= <\sigma, (\phi\circ\gamma)^{\dot{}}><\rho, (\phi\circ\gamma)^{\dot{}}> = <\sigma, \phi_*\dot\gamma><\rho, \phi_*\dot\gamma> \\ &= <\phi^*\sigma, \dot\gamma><\phi^*\rho, \dot\gamma> = <\phi^*\sigma\cdot\phi^*\rho \, , \, \ddot\gamma> \ . \end{aligned}$$

Conversely, the pull-back maps are characterized by the first of these three properties (this is easy), but also by the other two, as we shall see in Proposition 4. Skipping the rest of this section, Propositions 4 and 5, is harmless; what will be needed later is only the definition of Schwartz morphisms (property (i) in Proposition 5).

Proposition 4. *Let M and N be two C^p-manifolds (with $2 \leqslant p \leqslant \infty$), x (resp. y) a point in M (resp. N), and $B : \mathbb{T}^*_y N \to \mathbb{T}^*_x M$ a linear map. The following two conditions are equivalent:*

(i) there exists a C^p map $\phi : M \to N$ such that $\phi(x) = y$ and $\mathbb{T}^*_x\phi = B$;

(ii) there exists $b : \mathrm{T}^*_y N \to \mathrm{T}^*_x M$ such that

$$\forall\, \theta \in \mathbb{T}^*_y N \qquad b(\mathrm{R}\theta) = \mathrm{R}(B\theta) \ ;$$
$$\forall\, \sigma \in \mathrm{T}^*_y N \quad \forall\, \rho \in \mathrm{T}^*_y N \qquad B(\sigma{\cdot}\rho) = (b\sigma)\cdot(b\rho) \ .$$

When these conditions are met, $b = \mathrm{T}^*_x\phi$.

Proof (sketchy). We already saw (i) ⇒ (ii).

For the converse, suppose first that $M = \mathbb{R}^d$, $N = \mathbb{R}^e$, $x = 0$ and $y = 0$; call x^i and y^α the coordinates in M and N. Define real coefficients B^α_{ij} and B^α_k by

$$B^\alpha_{ij} = <B\big(\mathrm{d}^2 y^\alpha(0)\big), D_{ij}(0)>$$
$$B^\alpha_k = <B\big(\mathrm{d}^2 y^\alpha(0)\big), D_k(0)>$$

and e functions $\psi^1, \ldots, \psi^e$ on M by

$$\psi^\alpha = \tfrac{1}{2}\sum_{ij} B^\alpha_{ij}\, x^i x^j + \sum_k B^\alpha_k\, x^k \;;$$

last, define $\psi : M \to N$ by $y^\alpha \circ \psi = \psi^\alpha$. It is not difficult to verify that b must be equal to $\mathrm{T}^*_x\psi$ and that ψ satisfies (i).

Suppose now that M (resp. N) is an open neighbourhood of $x = 0 \in \mathbb{R}^d$ (resp. $y = 0 \in \mathbb{R}^e$). Define $\psi : \mathbb{R}^d \to \mathbb{R}^e$ as above. Let $g : M \to [0,1]$ be C^p, equal to 1 in a neighbourhood of x, and supported in a compact $K \subset M$ small enough for $\phi = g\psi$ to take its values in N; this map ϕ satisfies (i).

Last, if M and N are general manifolds, choose a local chart $(x^1, \dots, x^d)$ whose domain D contains x and such that $x^i(x) = 0$ for all i, and a local chart $(y^1, \dots, y^e)$ whose domain E contains y and such that $y^\alpha(y) = 0$ for all α. By the preceding step, there exists $\phi : D \to E$ satisfying (i) and such that $\phi \equiv y$ outside a compact subset $K \subset D$; it now suffices to extend ϕ to the whole of M by setting $\phi \equiv y$ on $M \setminus K$. This proves that (ii) $\Rightarrow$ (i).

We now suppose (i) and (ii) to be satisfied. Since $R : \mathbb{T}^*_y N \to \mathrm{T}^*_y N$ is onto, b is uniquely defined by the identity $bR = RB$; since $\mathrm{T}^*_x\phi$ also satisfies this identity, $b = \mathrm{T}^*_x\phi$. □

Proposition 5 and definition. *Let M and N be two C^p-manifolds (where $2 \leqslant p \leqslant \infty$), x (resp. y) a point in M (resp. N), and $F : \mathbb{T}_x M \to \mathbb{T}_y N$ a linear map. The following are equivalent:*

(i) there exists a C^p map $\phi : M \to N$ such that $\phi(x) = y$ and $\phi_{*x} = F$;

(ii) there exists a linear map $f : \mathrm{T}_x M \to \mathrm{T}_y N$ such that the restriction of F to $\mathrm{T}_x M$ is equal to f, and that, for all σ and ρ in $\mathrm{T}^*_y N$,

$$F^*(\sigma{\cdot}\rho) = f^*\sigma{\cdot}f^*\rho \;,$$

where $F^* : \mathbb{T}^*_y N \to \mathbb{T}^*_x M$ denotes the adjoint of $F : \mathbb{T}_x M \to \mathbb{T}_y N$ and $f^* : \mathrm{T}^*_y N \to \mathrm{T}^*_x M$ the adjoint of $f : \mathrm{T}_x M \to \mathrm{T}_y N$.

When these conditions are met, F is called a Schwartz morphism.

A Schwartz morphism is a bijection between $\mathbb{T}_x M$ and $\mathbb{T}_y N$ if and only if its restriction f to $\mathrm{T}_x M$ is a bijection between $\mathrm{T}_x M$ and $\mathrm{T}_y N$. In that case, its inverse is a Schwartz morphism too.

Proof. Equivalence between (i) and (ii) is straightforward from Proposition 4 by duality.

If a Schwartz morphism $F = \mathbb{T}_x\phi : \mathbb{T}_x M \to \mathbb{T}_y N$ is a bijection, $\mathbb{T}_x M$ and $\mathbb{T}_y N$ have the same dimension, so $\frac{1}{2}d(d{+}1) + d = \frac{1}{2}e(e{+}1) + e$, giving $d = e$, and $\mathrm{T}_x M$ and $\mathrm{T}_y N$ have the same dimension. But F restricted to $\mathrm{T}_x M$ is into, and valued in $\mathrm{T}_y N$; so it is a bijection too.

Conversely, if $f = \mathrm{T}_x\phi$ is a bijection, the inverse mapping theorem says that ϕ is locally invertible, which implies that $\mathbb{T}_x\phi$ is invertible, with inverse the Schwartz morphism $\mathbb{T}_y(\phi^{-1})$. □

By the preceding proposition, the bijective Schwartz morphisms from $\mathbb{T}_0\mathbb{R}^d$ to itself form a group. This group is the structure group needed to make $\mathbb{T}M$ a principal bundle (see Chapter ???).

Real and manifold-valued semimartingales

Conventions. In these lectures, the words 'semimartingale', 'martingale', 'local martingale' always implicitly mean 'continuous semimartingale', 'continuous martingale' or 'continuous local martingale'.

Whenever a stochastic process is involved, we always suppose given a fixed probability space $(\Omega, \mathcal{A}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t)_{t\geqslant 0}$.

Definitions (see Chapter ???). Recall that a *semimartingale* is a real process $X = (X_t)_{t\geqslant 0}$ having a decomposition $X_t = X_0 + M_t + A_t$, where M is a local martingale started from 0 (called the 'martingale part' of X, even when it is not a martingale), and A a process with finite variation (i.e., the difference of two adapted, continuous, increasing processes started from 0).

The *bracket* (or *covariation*) of two semimartingales X and Y is the process $\langle X, Y\rangle$ with finite variation such that

$$\langle X, Y\rangle_t = \lim_{|\sigma|\to 0} \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}) \ ,$$

where the limit is taken (in a suitable sense) along a refining sequence of subdivisions $\sigma = (0 = t_0 < t_1 < \ldots < t_n = t)$ of the interval $[0, t]$. (Notice the typographical difference between this bracket and the pairing bracket $<x^*, x>$ we use for duality in vector spaces.) The bracket $\langle X, Y\rangle$ does not change when processes with finite variation are added to X and Y. The covariation can also be defined using stochastic integrals:

$$\langle X, Y\rangle = XY - X_0Y_0 - \int X\,\mathrm{d}Y - \int Y\,\mathrm{d}X \ .$$

If H (resp. K) is a locally bounded predictable process,

$$\big\langle \textstyle\int H\,\mathrm{d}X, \int K\,\mathrm{d}Y \big\rangle = \displaystyle\int HK\,\mathrm{d}\langle X, Y\rangle \ .$$

(We use the notation $\int X\,\mathrm{d}Y$ to denote the process whose value at time t is $\int_0^t X_s\,\mathrm{d}Y_s$; by convention, all brackets and stochastic integrals are null at time 0.) The bracket is involved in the second-order terms of the change of variable formula: if $X^1, \ldots, X^d$ are semimartingales, and if $f : \mathbb{R}^d \to \mathbb{R}$ is of class C^2, $f(X^1, \ldots, X^d)$ is a semimartingale too, and, more precisely,

$$f(X^1, \ldots, X^d) = f(X_0^1, \ldots, X_0^d) \ + \sum_{k=1}^{d} \int \mathrm{D}_k f(X^1, \ldots, X^d)\,\mathrm{d}X^k + \tfrac{1}{2} \sum_{i,j=1}^{d} \int \mathrm{D}_{ij} f(X^1, \ldots, X^d)\,\mathrm{d}\langle X^i, X^j\rangle \ .$$

Exercise. If Y and Z are two semimartingales, the Stratonovich stochastic integral $\int Y \circ\mathrm{d}Z$, defined by $\int Y \circ\mathrm{d}Z = \int Y\,\mathrm{d}Z + \frac{1}{2}\langle Y, Z\rangle$, has the following property: the change of variable formula can be rewritten as

$$f(X^1,\dots,X^d) = f(X_0^1,\dots,X_0^d) + \sum_{k=1}^{d} \int \mathrm{D}_k f(X^1,\dots,X^d) \circ\mathrm{d}X^k \;.$$

See § 1.7 of Chapter ???. (In my opinion, the name 'integral' is not proper: since some regularity is required from Y, it is not an integral, but an integro-differential operator.)

Definition. A process X with values in a manifold M is called a *semimartingale in* M (or, shortly, a semimartingale if there is no ambiguity) if $f\circ X$ is a (real) semimartingale for every $f\in \mathrm{C}^2(M)$.

This implies in particular that X is continuous, and adapted to the filtration.

Exercise. When $M=\mathbb{R}$, there is no ambiguity: X is a semimartingale in M if and only if X is an ordinary semimartingale.

When $M=\mathbb{R}^d$, X is a semimartingale in M if and only if each component X^i of X is an ordinary semimartingale.

Exercise. If X is a semimartingale in M and N a submanifold of M such that X takes its values in N, then X is a semimartingale in N. (In particular, if M is embedded in $\mathbb{R}^n$, a process X with values in M is a semimartingale if and only if it is a semimartingale in $\mathbb{R}^n$.)

Many interesting manifold-valued processes are diffusions, a particular instance of semimartingales. We shall focus on semimartingales, instead of mere diffusions, because semimartingales lend themselves to general theorems due to their many stability properties (stability by C^2 functions, by stochastic integration, change of time or of probability; whereas, for instance, the image of a diffusion by a map is generally not a diffusion). These general theorems are quite useful, even if one is only interested in applying them to diffusions.

Schwartz' principle

Schwartz' most important contribution to stochastic differential geometry is his second order intrinsic interpretation of the change of variable formula. Let X denote a semimartingale in $\mathbb{R}^d$, or in a manifold with a global chart; call $X^1,\dots,X^d$ the coordinates of X. For an infinitesimal time-increment $\mathrm{d}t$, Schwartz writes

$$f(X_{t+\mathrm{d}t}) - f(X_t) = \sum_k \mathrm{D}_k f(X_t)\,\mathrm{d}X_t^k + \tfrac{1}{2}\sum_{i,j} \mathrm{D}_{ij} f(X_t)\,\mathrm{d}\langle X^i,X^j\rangle_t \;,$$

and interprets the right-hand side as the duality pairing of the codiffusor $\mathrm{d}^2 f\big(X_t(\omega)\big)$ with the diffusor $\mathcal{D}X_t = \sum_k \mathrm{d}X_t^k\,\mathrm{D}_k + \frac{1}{2}\sum_{i,j}\mathrm{d}\langle X^i,X^j\rangle_t\,\mathrm{D}_{ij}$. The former is a bona fide element of $\mathbb{T}^*_{X_t(\omega)}\mathbb{R}^d$, but the latter is an element of $\mathbb{T}_{X_t(\omega)}\mathbb{R}^d$ whose coefficients $\mathrm{d}X_t^k(\omega)$ and $\mathrm{d}\langle X^i, X^j\rangle_t(\omega)$ have no rigorous meaning, because there is no $\int$ sign. The key observation is that, informal as it may be, this diffusor is intrinsic: in a global change of coordinates, that is, after a diffeomorphism, the left-hand side $f\big(X_{t+\mathrm{d}t}(\omega)\big) - f\big(X_t(\omega)\big)$ and the codiffusor $\mathrm{d}^2 f\big(X_t(\omega)\big)$ are invariant, hence this informal diffusor must also be invariant.

The same observation can be made for a semimartingale in a general manifold, with local coordinates instead of global ones. This idea will be made rigorous in Theorem 1, not by giving a precise meaning to $\mathrm{d}X_t^k(\omega)$, $\mathrm{d}\langle X^i, X^j\rangle_t(\omega)$, and $\mathcal{D}X_t(\omega)$, but by supplying the missing $\int$ sign. Indeed, the fact that the intrinsic nature of $\mathcal{D}X_t$ is that of a diffusor, is mathematically expressed by first pairing $\mathcal{D}X_t$ with a codiffusor $\Theta_t \in \mathbb{T}^*_{X_t}M$, then integrating along the random (but not smooth) curve X. (Similarly, the fact that the velocity of a smooth curve is a tangent vector, is expressed by the possibility of integrating cotangent vectors, or a 1-form, along the curve.)

In fact, *codiffusors are the most general objects that can be integrated along manifold-valued semimartingales.* This sentence cannot be given a precise mathematical meaning; it is only a metamathematical statement, but it permeates these lectures from now on. Schwartz has called it the *fundamental principle*; and Meyer has rightly renamed it *Schwartz' principle.*

Definition. Let M be a manifold, X a semimartingale in M, Θ a process with values in $\mathbb{T}^*M$, above X (that is, $\Theta_t(\omega)$ belongs to the fibre $\mathbb{T}^*_{X_t(\omega)}M$ for all t and ω). The process Θ is said to be *locally bounded*[1] if there exist a sequence of compact sets $K_n \subset \mathbb{T}^*M$ and a sequence of stopping times $T_n \nearrow \infty$ such that on the stochastic interval $[\![0, T_n[\![$, Θ takes its values in K_n.

Theorem 1 and definition. *If X is a semimartingale in a manifold M, and Θ a $\mathbb{T}^*M$-valued process above X, locally bounded and predictable, the* stochastic integral of Θ along X *can be defined. It is a real semimartingale, null at time 0, denoted by $\int <\Theta, \mathcal{D}X>$, linear in Θ, and characterized by the following two properties:*

(i) for every $f \in \mathrm{C}^2(M)$,

$$\int <\mathrm{d}^2 f, \mathcal{D}X> = f\circ X - f\circ X_0 \;;$$

(ii) for every real, locally bounded, predictable process H,

$$\int <H\Theta, \mathcal{D}X> = \int H \,\mathrm{d}\big(\textstyle\int <\Theta, \mathcal{D}X>\big) \;.$$

[1] The usual definition of local boundedness involves closed stochastic intervals $[\![0, T_n]\!]$; the property defined here should rather be called 'pre-local'. But we shall apply it to predictable processes only, and it turns out that a predictable process is locally bounded if and only if it is pre-locally bounded.

(In (i), and in the sequel, we use the following notation: if θ is a Borel, locally bounded codiffusor field on M, $\int<\theta, \mathcal{D}X>$ stands for $\int<\theta(X), \mathcal{D}X>$.)

This stochastic integral also has the following further properties:

(iii) *If Ξ satisfies the same requirements as Θ,*

$$\tfrac{1}{2}\left\langle \textstyle\int<\Theta, \mathcal{D}X>, \int<\Xi, \mathcal{D}X>\right\rangle = \int <\mathrm{R}\Theta{\cdot}\mathrm{R}\Xi, \mathcal{D}X> ;$$

and in particular, for f and g in $\mathrm{C}^2(M)$,

$$\tfrac{1}{2}\left\langle f\circ X, g\circ X\right\rangle = \int <\mathrm{d}f{\cdot}\mathrm{d}g, \mathcal{D}X> .$$

(iv) *If T is a stopping time, the stopped stochastic integral is given by*

$$\Big(\int <\Theta, \mathcal{D}X>\Big)^T = \int <\Theta^T, \mathcal{D}X^T> .$$

(v) *If X has finite variation, so has also $\int<\Theta, \mathcal{D}X>$; moreover, this stochastic integral is equal to the pathwise Stieltjes integral $\int<\mathrm{R}\Theta, \mathrm{d}X>$ of the covectors $\mathrm{R}\Theta$ along X.*

(vi) *If $\mathrm{R}\Theta = 0$, $\int<\Theta, \mathcal{D}X>$ has finite variation; if furthermore Θ is positive, $\int<\Theta, \mathcal{D}X>$ is an increasing process.*

(vii) *If $\mathrm{R}\Theta = \mathrm{R}\Xi$, the semimartingales $\int<\Theta, \mathcal{D}X>$ and $\int<\Xi, \mathcal{D}X>$ have the same martingale part.*

Proving the theorem is easy when M is diffeomorphic to an open subset of $\mathbb{R}^d$, that is, when there exist *global* coordinates. It suffices to introduce the d real semimartingales $X^i = x^i \circ X$ and the $d^2 + d$ predictable processes $\Theta_{ij}(t)$ and $\Theta_k(t)$ such that $\Theta_{ij}(t) = \Theta_{ji}(t)$ and

$$\Theta_t = \sum_{ij} \Theta_{ij}(t)\, \mathrm{d}x^i(X_t){\cdot}\mathrm{d}x^j(X_t) + \sum_k \Theta_k(t)\, \mathrm{d}^2 x^k(X_t) .$$

Uniqueness is then established by replacing $\mathrm{d}x^i(X_t){\cdot}\mathrm{d}x^j(X_t)$ with

$$\tfrac{1}{2}\left[\mathrm{d}^2(x^i x^j)(X_t) - x^i(X_t)\,\mathrm{d}^2 x^j(X_t) - x^j(X_t)\,\mathrm{d}^2 x^i(X_t)\right],$$

so as to express Θ_t as a finite sum of processes of the form $H_t\,\mathrm{d}^2 f(X_t)$, and by using linearity. Existence and properties (iii)–(vii) are obtained by verifying that

$$\sum_k \int \Theta_k\, \mathrm{d}X^k + \tfrac{1}{2}\sum_{ij}\int \Theta_{ij}\, \mathrm{d}\langle X^i, X^j\rangle$$

has all the claimed properties of $\int<\Theta, \mathcal{D}X>$.

The general case, when there is no global chart, is less straightforward, but the difficulty is only technical. One can for instance use localization methods;

another possibility is to embed M into some $\mathbb{R}^n$ as a closed submanifold; one can alternatively resort to the theory of semimartingales in open sets. These technical matters are important, but may be ignored in a first approach to the subject; we shall consider this extension as plausible enough and simply admit it.

Exercise. Let X be a semimartingale in a manifold M, and $\phi : M \to N$ a C^2 map; let Θ be a predictable, locally bounded, codiffusor-valued process above the semimartingale $\phi \circ X$. Show that $\phi^*\Theta$ is a locally bounded process above X, and that $\int <\Theta, \mathcal{D}(\phi \circ X)> = \int <\phi^*\Theta, \mathcal{D}X>$.

Exercise. The statement below complements (vii) in Theorem 1, by saying that the geometric nature of the "martingale part $\mathrm{d}\overset{\mathrm{m}}{X}$ of $\mathcal{D}X$" is that of a tangent vector. Prove it in the particular case that M has global coordinates.

If Σ is a locally bounded, predictable, covector-valued process above X, one can define the integral $\int <\Sigma, \mathrm{d}\overset{\mathrm{m}}{X}>$, which is the martingale part of $\int <\Theta, \mathcal{D}X>$ for any Θ such that $\mathrm{R}\Theta = \Sigma$.

Stratonovich stochastic integrals

As explained in Proposition 1.6 of Chapter ???, if X is a semimartingale in $\mathbb{R}^d$, the integral of a closed 1-form σ along the (random, non smooth) curve X can be defined, using Stratonovich stochastic integration. This procedure is intrinsic, and extends to non closed 1-forms and semimartingales in manifolds. It is one of the most popular and useful tools in stochastic differential geometry. We shall present it here as a particular instance of the Schwartz principle: every 1-form (= covector field) σ can canonically be transformed into a second-order form (= codiffusor field) $\mathrm{d}\sigma$, which in turn can be integrated along X to yield the Stratonovich integral of σ along X.

Proposition 6 and definition (symmetric differential of a 1-form). *On a manifold M, let σ be a covector field of class C^1 at least. There exists on M a unique codiffusor field $\mathrm{d}\sigma$, called the* symmetric differential *of σ, such that, for all C^2 curves γ,*

$$<\mathrm{d}\sigma, \ddot{\gamma}(t)> = \frac{\mathrm{d}}{\mathrm{d}t} <\sigma, \dot{\gamma}(t)> .$$

The symmetric differential has the following properties: for all C^1 covector fields σ, all C^2 functions f and all C^1 functions g,

$$\mathrm{R}(\mathrm{d}\sigma) = \sigma \; ; \qquad \mathrm{d}(\mathrm{d}f) = \mathrm{d}^2 f \; ; \qquad \mathrm{d}(g\sigma) = \mathrm{d}g \cdot \sigma + g \, \mathrm{d}\sigma \; .$$

The notation $\mathrm{d}^2 f$ is justified by the formula $\mathrm{d}^2 f = \mathrm{d}(\mathrm{d}f)$. Symmetric differentiation of 1-forms is something completely different from exterior differentiation, which makes 1-forms into skew-symmetric 2-forms (instead of symmetric second-order forms), and which kills $\mathrm{d}f$.

Proof. Uniqueness stems easily from the fact that, for each $x \in M$, the diffusors $\ddot\gamma(0)$ linearly span $\mathbb{T}_x M$ when γ ranges over all C^2 curves such that $\gamma(0) = x$.

Existence is proved in local coordinates: on the domain $D(C)$ of any local chart $C = (x^1, \ldots, x^d)$, every covector field σ can be written as $\sum_i \sigma_i \, \mathrm{d}x^i$, where $\sigma_1, \ldots, \sigma_d$ are C^1 functions on $D(C)$. Define a codiffusor field $(\mathrm{d}\sigma)_C$ on $D(C)$ by the formula $\sum_{ij} \mathrm{D}_i \sigma_j \, \mathrm{d}x^i{\cdot}\mathrm{d}x^j + \sum_k \sigma_k \, \mathrm{d}^2 x^k$. It is easy to verify that $(d\sigma)_C$ satisfies on $D(C)$ the definition of $\mathrm{d}\sigma$, as well as the other three properties; this is left to the reader.

It only remains to show that there exists a global codiffusor field $\mathrm{d}\sigma$ whose restriction to $D(C)$ is $(\mathrm{d}\sigma)_C$ for any local chart C; it suffices to check that if C and C' are two local charts, one has $(\mathrm{d}\sigma)_C = (\mathrm{d}\sigma)_{C'}$ on the intersection of their domains; and this is a consequence of uniqueness applied to the manifold $D(C) \cap D(C')$. □

Exercise. Show that the codiffusor field $\mathrm{d}\sigma$ is characterized by the following two properties: if V and W are vector fields,

$$<\mathrm{d}\sigma, V> = <\sigma, V> \; ;$$
$$<\mathrm{d}\sigma, VW{+}WV> = <\mathrm{d}{<}\sigma, V{>}, W> + <\mathrm{d}{<}\sigma, W{>}, V> \; .$$

Exercise. Symmetric differentiation has the following tensorial property: if $\phi : M \to N$ is C^2 and if σ is a 1-form on N, then $\mathrm{d}(\phi^*\sigma) = \phi^*(\mathrm{d}\sigma)$.

Theorem 2 and definition. *Let M be a manifold of class C^3 at least, X a semimartingale in M, and Σ a semimartingale in $\mathbb{T}M$, above X. The* Stratonovich stochastic integral *of Σ along X is the semimartingale $\int <\Sigma, \circ\mathrm{d}X>$, depending linearly on Σ, and such that:*

(i) if σ is a covector field, C^2 at least,

$$\int <\sigma{\circ}X, \circ\mathrm{d}X> = \int <\mathrm{d}\sigma, \mathcal{D}X> \; ;$$

(ii) if Z is a real semimartingale,

$$\int Z \circ\mathrm{d}\Big(\textstyle\int <\Sigma, \circ\mathrm{d}X>\Big) = \displaystyle\int <Z\Sigma, \circ\mathrm{d}X> \; .$$

This integral also has the following further properties:

(iii) If P satisfies the same requirements as Σ,

$$\tfrac12 \left\langle \textstyle\int <\Sigma, \circ\mathrm{d}X> , \int <P, \circ\mathrm{d}X> \right\rangle = \displaystyle\int <\Sigma{\cdot}P, \mathcal{D}X> \; .$$

(iv) If T is a stopping time, one has

$$\left(\int <\Sigma, \circ\mathrm{d}X>\right)^T = \int <\Sigma^T, \circ\mathrm{d}X^T> \; .$$

For $f \in \mathrm{C}^3(M)$, taking $\sigma = \mathrm{d}f$ in (i) gives the important Stratonovich change of variable of variable formula

$$f \circ X = f(X_0) + \int <\mathrm{d}f, \circ\mathrm{d}X> .$$

Conversely, this formula and property (ii) fully characterize Stratonovich stochastic integrals, giving a convenient definition with no reference to second-order language (this is the definition usually found in the literature).

Exercise. Assuming the existence of a global chart, prove Theorem 2. (Hint: write Σ_t as $\sum_k \Sigma_k(t)\,\mathrm{d}x^k(X_t)$, where $\Sigma_k(t)$ are real semimartingales, and define $\int<\Sigma, \circ\mathrm{d}X>$ as $\sum_k \int \Sigma_k \circ\mathrm{d}(x^k \circ X)$.)

Exercise. The martingale part of the semimartingale $\int<\Sigma, \circ\mathrm{d}X>$ is equal to $\int<\Sigma, \mathrm{d}\overset{\mathrm{m}}{X}>$.

Exercise. Complete and prove the following statement (functoriality of Stratonovich integrals). *Let X be a semimartingale in a manifold M, and $\phi : M \to N$ a C^3 map; let Σ be a semimartingale in T^*N above the semimartingale $\phi \circ X$. Then ...*

Second-order stochastic differential equations

Given two manifolds M and N, it is convenient to call an *ordinary differential equation (ODE) from M to N* any family $e = \big(e(x, y)\big)_{x\in M,\, y\in N}$ such that each $e(x, y)$ is a linear mapping from T_xM to T_yN. Suppose given an ODE e, a differentiable curve $\gamma : I \to M$, where I is an open interval containing 0, and a point y_0 in N. A *solution to the differential equation*

$$\dot c = e(\gamma, c)\,\dot\gamma\ ; \quad c(0) = y_0$$

is any differentiable curve $c : J \to N$, with J an open interval such that $0 \in J \subset I$, verifying $c(0) = y_0$ and $\dot c(t) = e\big(\gamma(t), c(t)\big)\,\dot\gamma(t)$ for all $t \in J$.

In all useful examples of this situation, the map $(x, y) \mapsto e(x, y)$ is locally Lipschitz; this implies existence and uniqueness of the solution (more precisely: of the maximal solution, that which is defined on a maximal interval J; all other solutions are restrictions of the maximal one to sub-intervals of the maximal interval).

This scheme of ODEs is general enough to include many examples of constructions of curves in geometry. Here are some examples.

Take any vector field V on N; the integral curves of V are the N-valued curves c such that $\dot c(t) = V\big(c(t)\big)$ for each t. The problem of constructing the integral curve of V with initial condition $c(0) = y_0$ can be considered as an ODE where $M = \mathbb{R}$, the curve $\gamma : \mathbb{R} \to M$ is the identity, and $e(t, y)$ maps $\mathrm{d}/\mathrm{d}t$ (every vector of $\mathrm{T}_t\mathbb{R}$ is a multiple of that one) to $V(y) \in \mathrm{T}_yN$. The case of a time-dependent vector field $V(t,y)$ reduces just as easily to an ODE, simply by letting $e(t,y)$ map $\mathrm{d}/\mathrm{d}t$ to $V(t,y)$.

Another example is integration of 1-forms along curves. If σ is a 1-form on M, define an ODE e from M to $N=\mathbb{R}$ by taking $e(x,y):\mathrm{T}_xM\to\mathrm{T}_y\mathbb{R}\approx\mathbb{R}$ equal to the covector $\sigma(x)\in\mathrm{T}^*_xM$. If γ is any curve in M, the solution to the ODE e applied to γ with initial condition y_0 is the real function

$$t\mapsto\int_{y_0}^{t}<\sigma\big(\gamma(s)\big),\mathrm{d}\gamma(s)> = \int_{y_0}^{t}<\sigma\big(\gamma(s)\big),\dot\gamma(s)>\mathrm{d}s\ ,$$

that is, the integral of σ along γ.

The horizontal lift of curves to a fibre bundle by a non-linear connection (see Chapter ???) gives a third example of an ODE, but with a little snag: in that case M is the base manifold and N the bundle, and $e(x,y)$ should be the horizontal map from T_xM to T_yN. This does not exactly fit our definition of an ODE, because the horizontal map $h(x,y)$ is not defined for all $(x,y)\in M\times N$, only for those (x,y) such that y belongs to the fibre above x. One way to deal with such situations is to extend h to an e defined on the whole of $M\times N$, such that $e(x,y)=h(x,y)$ if y is above x. If e is regular enough (e can always be chosen as smooth as h; Lipschitz continuity suffices), the solution of the ODE will always remain in the set $\{(x,y):y \text{ is above } x\}$, provided it is started in this set; and it is equal to the horizontal lift of the given curve, regardless of the choice of the extension e of h.

To define stochastic differential equations (SDEs) between manifolds, the first thing that comes to mind is to mimic ODEs, by choosing a family $\big(F(x,y)\big)_{x\in M,\,y\in N}$ such that each $F(x,y)$ is a linear map from $\mathbb{T}_xM$ to $\mathbb{T}_yN$. An SDE driven by an M-valued semimartingale X would then have the form $\mathcal{D}Y=F(X,Y)\,\mathcal{D}X$, yielding the infinitesimal increment $\mathcal{D}Y_t$ as a function of X_t, Y_t and $\mathcal{D}X_t$, linear in $\mathcal{D}X_t$. But this is too general to make sense, as shown by the following simple computation when M and N have global coordinates. Such a general SDE would be given by

$$\begin{cases}\mathrm{d}Y^\gamma_t = \sum\limits_k F^\gamma_k(X_t,Y_t)\,\mathrm{d}X^k_t + \frac12\sum\limits_{ij}F^\gamma_{ij}(X_t,Y_t)\,\mathrm{d}\langle X^i,X^j\rangle_t\\ \frac12\,\mathrm{d}\langle Y^\alpha,Y^\beta\rangle_t = \sum\limits_k F^{\alpha\beta}_k(X_t,Y_t)\,\mathrm{d}X^k_t + \frac12\sum\limits_{ij}F^{\alpha\beta}_{ij}(X_t,Y_t)\,\mathrm{d}\langle X^i,X^j\rangle_t\ ,\end{cases}$$

where the various coefficients $F^{\cdots}_{\cdots}(X_t,Y_t)$ represent the linear map $F(X_t,Y_t)$. But the rules of stochastic calculus applied to the first line yield

$$\mathrm{d}\langle Y^\alpha,Y^\beta\rangle_t = \sum_i F^\alpha_i(X_t,Y_t)\sum_j F^\beta_j(X_t,Y_t)\,\mathrm{d}\langle X^i,X^j\rangle_t\ ;$$

so the compatibility conditions $F^{\alpha\beta}_k=0$ and $F^{\alpha\beta}_{ij}=\frac12[F^\alpha_iF^\beta_j+F^\alpha_jF^\beta_i]$ (if the coefficients are chosen to be symmetric) must be satisfied. These relations turn out to exactly express the fact that $F(X_t,Y_t)$ is a Schwartz morphism from

$\mathbb{T}_{X_t}M$ to $\mathbb{T}_{Y_t}N$ (recall that Schwartz morphisms are defined by property (i) in Proposition 5); this statement is left as an exercise. (Hint: $F_k^\gamma = \mathrm{D}_k(y^\gamma \circ \phi)$; $F_{ij}^\gamma = \mathrm{D}_{ij}(y^\gamma \circ \phi)$.) So the true nature of SDEs between manifolds involves Schwartz morphisms.

Definitions. A *second-order stochastic differential equation* between two manifolds M and N is a family $\big(F(x,y)\big)_{x\in M\ y\in N}$, where each $F(x,y)$ is a Schwartz morphism from $\mathbb{T}_x M$ to $\mathbb{T}_y N$.

Given a second-order SDE F, a semimartingale X in M and an $\mathcal{F}_0$-measurable random variable y_0 in N, a *solution to the second-order SDE*

$$\mathcal{D}Y = F(X,Y)\,\mathcal{D}X\ ; \quad Y_0 = y_0$$

is a semimartingale Y in N (possibly defined only on some stochastic interval $[\![0,\zeta[\![$ where ζ is a predictable stopping time), such that $Y_0 = y_0$ and, for all codiffusor fields θ on N,

$$\int <\theta, \mathcal{D}Y> = \int <F(X,Y)^*\theta, \mathcal{D}X>$$

on the interval $[\![0,\zeta[\![$ where Y is defined.

Observe that it suffices to require the above equality only for the codiffusor fields θ of the form $\mathrm{d}^2 f$ for $f \in \mathrm{C}^2(N)$; that is, Y is characterized by

$$f \circ Y_t = f(y_0) + \int_0^t <F(X_s,Y_s)^*\big(\mathrm{d}^2 f(Y_s)\big), \mathcal{D}X_s> .$$

And on the other hand, if Y is a solution (and if F is locally bounded), one has more generally $\int <\Theta, \mathcal{D}Y> = \int <F(X,Y)^*\Theta, \mathcal{D}X>$ for every locally bounded, predictable, $\mathbb{T}^*N$-valued process Θ above Y.

Theorem 3 (Existence and uniqueness for second-order SDEs. *Let F be a second-order SDE between two manifolds M and N, X a semimartingale in M and y_0 an N-valued, $\mathcal{F}_0$-measurable r.v. Suppose the map $(x,y) \mapsto F(x,y)$ to be locally Lipschitz. There exist a predictable stopping time $\zeta > 0$ and a semimartingale Y in N defined on $[\![0,\zeta[\![$ such that*

(i) on the event $\{\zeta < \infty\}$, the limit $\lim_{t\uparrow\zeta} Y_t$ exists in the one-point compactification $N \cup \{\infty_N\}$ of N and is equal to the point at infinity ∞_N;

(ii) on $[\![0,\zeta[\![$, Y is a solution to the second-order SDE

$$\mathcal{D}Y = F(X,Y)\,\mathcal{D}X\ ; \quad Y_0 = y_0 .$$

Moreover, if Y' is any other solution to this SDE, defined on some predictable interval $[\![0,\zeta'[\![$, then $\zeta' \leqslant \zeta$ and Y' is the restriction of Y to $[\![0,\zeta'[\![$.

This theorem extends to manifolds the existence and uniqueness statement for SDEs driven by vector-valued semimartingales; indeed, one possible proof of the theorem consists in embedding M and N into vector spaces to be brought back to the vector case. Other possible methods are localization, or using the theory of semimartingales in predictable open intervals. A full proof with all technical details is beyond the scope of this “invitation”.

Observe that a consequence of (i) is that $\zeta = \infty$ if N is compact. (More generally, $\zeta = \infty$ on the event that the connected component of N containing y_0 is compact.)

What makes Theorem 3 useful is its generality. Yet, in some applications, one sometimes needs a still more general version (not to be found in the present notes), where F is allowed to depend, not only upon x and y, but also upon t and ω (the dependence on (t,ω) must then be predictable). In that case, $\mathcal{D}Y_t$ is not only a function of X_t, Y_t and $\mathcal{D}X_t$, but may also take into account the past behaviour of X and Y, and also possibly of other processes adapted to the filtration.

Stratonovich stochastic differential equations

Definition. Let M and N be two manifolds (at least C^3) and e an ODE from M to N; suppose the map $(x,y) \mapsto e(x,y)$ to be C^2. Given a semimartingale X in M and an $\mathcal{F}_0$-measurable, N-valued r.v. Y_0, a semimartingale Y (possibly with finite lifetime) in N is called a *solution to the Stratonovich SDE*

$$\circ \mathrm{d}Y = e(X,Y) \circ \mathrm{d}X \ ; \quad Y_0 = y_0$$

if $Y_0 = y_0$ and, for each 1-form σ on N, one has

$$\int <\sigma(Y), \circ \mathrm{d}Y> = \int <e(X,Y)^* \big(\sigma(Y)\big), \circ \mathrm{d}X> .$$

There are two main reason why Stratonovich SDEs are very useful. First, most intrinsic geometric constructions which yield a curve from another one can be seen as ODEs; substituting Stratonovich SDEs to these ODEs gives similar intrinsic constructions of semimartingales from other ones, no extra geometric tool or structure being needed. Second, the semimartingales solving the Stratonovich SDEs associated to these ODEs have properties similar to the properties of the curves solving the ODEs.

These vague assertions, known as the Stratonovich transfer principle, are made more precise in the next statement.

Theorem 4 (Stratonovich transfer principle). *Given a* C^2 *ODE* e *between two* C^3 *manifolds* M *and* N*, there exists a unique second-order SDE* F *from* M *to* N *such that, for every curve* $t \mapsto \big(\gamma(t), c(t)\big)$ *in* $M \times N$ *verifying*

$\dot c(t) = e\big(\gamma(t), c(t)\big)\,\dot\gamma(t)$, one also has $\ddot c(t) = F\big(\gamma(t), c(t)\big)\,\ddot\gamma(t)$. The restriction of $F(x, y)$ to $\mathrm{T}_x M$ is $e(x, y)$.

Moreover, if X (resp. Y) is a semimartingale in M (resp. N), the Stratonovich SDE

$$\circ\mathrm{d}Y = e(X, Y) \circ\mathrm{d}X$$

is satisfied if and only if so is also the second-order SDE

$$\mathcal{D}Y = F(X, Y)\,\mathcal{D}X\ .$$

If σ is a C^2 1-form on the product manifold $M \times N$ such that $<\sigma, \dot z> = 0$ for every curve $z(t) = \big(x(t), y(t)\big)$ in $M \times N$ verifying the ODE $\dot y = e(x, y)\,\dot x$, then $\int <\sigma, \circ\mathrm{d}Z> = 0$ for every semimartingale $Z_t = (X_t, Y_t)$ in $M \times N$ satisfying the associated Stratonovich SDE $\circ\mathrm{d}Y = e(X, Y) \circ\mathrm{d}X$.

The last part of Theorem 4 says that integral invariants (for instance conservation laws) are passed on from ODEs to Stratonovich SDEs. It says in particular that if f is a C^3 function on $M \times N$, and if $f\big(x(t), y(t)\big)$ is constant for any solution $\big(x(t), y(t)\big)$ of the ODE, then $f(X_t, Y_t)$ is constant for any solution (X_t, Y_t) of the Stratonovich SDE. See also Proposition 7 for another conservation property of the Stratonovich transfer principle.

Powerful as it is, the Stratonovich transfer principle has its limitations. In the case when the ODE e depends not only on (x, y), but also on t, one must be careful and replace M with the product $\mathbb{R} \times M$, so as to work with the pair (t, x) in lieu of x (instead of first applying the transfer principle, and only then allowing time-dependence: *this does not work*). This replacement of M with $\mathbb{R} \times M$ can be performed only when time-dependence is smooth enough: *if $e = e(t, x, y)$ is not C^2 in (t, x, y) (or at least C^1 with locally Lipschitz first derivatives), the ODE is not liable to Stratonovich transfer.* Similarly, the case of an ODE which depends on ω (for instance, to take into account the past behaviour of some intervening processes) is outside the scope of the Stratonovich transfer principle.

When applying the Stratonovich transfer principle, *never compute F from e*. The basic idea is to work with first-order vectors and forms, using only ordinary calculus and the Stratonovich chain rule. Second-order objects remain hidden in the background; they are needed only to establish general facts such as the following existence and uniqueness statement. (If you attempt to show by hand existence and uniqueness for a given particular Stratonovich SDE, you will eventually end up with replacing Stratonovich integrals by Itô ones to obtain estimates and perform, for instance, some Picard iteration. This is exactly what F does here once and for all in a general setting.)

Combining Theorems 3 and 4 immediately gives the following corollary:

Corollary (Existence and uniqueness for Stratonovich SDEs). *Given a* C^2 *ODE* e *between two* C^3 *manifolds* M *and* N*, a semimartingale* X *in* M *and an* $\mathcal{F}_0$*-measurable initial condition* y_0 *in* N*, the Stratonovich SDE*

$$\circ\mathrm{d}Y = e(X,Y)\circ\mathrm{d}X\ ;\quad Y_0 = y_0$$

has a unique solution Y *(possibly with finite explosion time).*

Proof of Theorem 4. The proof of existence and uniqueness of F follows the same general scheme as that of $\mathrm{d}\sigma$ in Proposition 6: F is unique because accelerations of curves span the second-order tangent space, and existence is shown locally, in local coordinates. If $(x^i)_{1\leqslant i\leqslant d}$ (resp. $(y^\alpha)_{1\leqslant\alpha\leqslant d'}$) is a local chart with domain $D\subset M$ (resp. $D'\subset N$), for $(x,y)\in M\times N$ the map $e(x,y):\mathrm{T}_xM\to\mathrm{T}_yN$ is given by its coefficients $e_k^\gamma(x,y)$ such that $e(x,y)\mathrm{D}_k=\sum_\gamma e_k^\gamma(x,y)\mathrm{D}_\gamma$; define the map $F(x,y):\mathbb{T}_xM\to\mathbb{T}_yN$ by

$$\begin{aligned} F(x,y)\,\mathrm{D}_{ij} &= \sum_{\alpha\beta} F_{ij}^{\alpha\beta}(x,y)\,\mathrm{D}_{\alpha\beta} + \sum_\gamma F_{ij}^\gamma(x,y)\,\mathrm{D}_\gamma \\ F(x,y)\,\mathrm{D}_k &= \sum_{\alpha\beta} F_k^{\alpha\beta}(x,y)\,\mathrm{D}_{\alpha\beta} + \sum_\gamma F_k^\gamma(x,y)\,\mathrm{D}_\gamma\ , \end{aligned}$$

where

$$\begin{aligned} &F_k^{\alpha\beta}(x,y)=0\ ;\quad F_k^\gamma(x,y)=e_k^\gamma(x,y)\ ;\quad F_{ij}^{\alpha\beta}(x,y)=e_i^\alpha(x,y)e_j^\beta(x,y)\ ;\\ &F_{ij}^\gamma(x,y)=\tfrac12\big[\mathrm{D}_ie_j^\gamma(x,y)+\mathrm{D}_je_i^\gamma(x,y)\\ &\qquad\qquad+\sum_\alpha\big(e_i^\alpha(x,y)\mathrm{D}_\alpha e_j^\gamma(x,y)+e_j^\alpha(x,y)\mathrm{D}_\alpha e_i^\gamma(x,y)\big)\big]\ . \end{aligned}$$

All this is defined only for $(x,y)\in D\times D'$ and depends upon the choice of the charts; but a direct computation shows that $\ddot c(t)=F\big(\gamma(t),c(t)\big)\,\ddot\gamma(t)$ for every curve $t\mapsto\big(\gamma(t),c(t)\big)$ in $D\times D'$ such that $\dot c(t)=e\big(\gamma(t),c(t)\big)\,\dot\gamma(t)$. This intrinsic characterization implies that $F(x,y)$ thus defined remains the same when computed with any other local charts whose domains contain x and y, so all these F can be patched up together to yield a globally defined family $F(x,y)$, of class C^1, meeting the requirement on curves. It only remains to see that each $F(x,y)$ is a Schwartz morphism; this is due to $F_k^{\alpha\beta}=0$ and $F_{ij}^{\alpha\beta}=F_i^\alpha F_j^\beta$.

The second part of the theorem, equivalence between $\circ\mathrm{d}Y=e(X,Y)\circ\mathrm{d}X$ and $\mathcal{D}Y=F(X,Y)\,\mathcal{D}X$, will now be established assuming there exist global coordinates (x^i) and (y^α) on the manifolds. Given any semimartingale (X,Y) in $M\times N$, writing X^i for $x^i\circ X$ and Y^α for $y^\alpha\circ Y$, it suffices to show that

$$\text{for all indices }\gamma,\qquad Y^\gamma-Y_0^\gamma=\int <e(X,Y)^*\big((\mathrm{d}y^\gamma)(Y)\big),\circ\mathrm{d}X>$$

holds if and only if

$$\text{for all indices } \gamma, \qquad Y^\gamma - Y^\gamma_0 = \int {<}F(X,Y)^*\big((\mathrm{d}^2 y^\gamma)(Y)\big), \mathcal{D}X{>}\ ;$$

so it suffices to establish the identity

$$(*) \qquad \int {<}e(X,Y)^*\big((\mathrm{d}y^\gamma)(Y)\big), \circ\mathrm{d}X{>} = \int {<}F(X,Y)^*\big((\mathrm{d}^2 y^\gamma)(Y)\big), \mathcal{D}X{>}\ .$$

The left-hand side of $(*)$ is the real semimartingale $\sum_k \int e^\gamma_k(X,Y) \circ\mathrm{d}X^k$; using the definition of Stratonovich integrals for real semimartingales, it expands as

$$\sum_k \Big[\int e^\gamma_k(X,Y)\,\mathrm{d}X^k + \tfrac12 \big\langle e^\gamma_k(X,Y), X^k\big\rangle\Big]$$
$$= \sum_k \Big[\int e^\gamma_k(X,Y)\,\mathrm{d}X^k + \tfrac12 \sum_i \int \mathrm{D}_i e^\gamma_k(X,Y)\,\mathrm{d}\langle X^i, X^k\rangle$$
$$+ \tfrac12 \sum_\alpha \int \mathrm{D}_\alpha e^\gamma_k(X,Y)\,\mathrm{d}\langle Y^\alpha, X^k\rangle\Big]\ .$$

Since this implies $Y^\gamma = Y^\gamma_0 + \sum_k e^\gamma_k(X,Y)\,\mathrm{d}X^k +$ terms with finite variation, the bracket $\mathrm{d}\langle Y^\alpha, X^k\rangle$ is also equal to $\sum_i e^\alpha_i(X,Y)\,\mathrm{d}\langle X^i, X^k\rangle$; so the left-hand side of $(*)$ is $\sum_k \int e^\gamma_k(X,Y)\,\mathrm{d}X^k + \sum_{ij} \int e^\gamma_{ij}(X,Y)\,\frac12\,\mathrm{d}\langle X^i, X^j\rangle$, where $e^\gamma_{ij}(x,y) = \mathrm{D}_i e^\gamma_j(x,y) + \sum_\alpha e^\alpha_i(x,y)\mathrm{D}_\alpha e^\gamma_j(x,y)$.

We now compute the right-hand side of $(*)$. First, observe that

$$F(X,Y)^*\big((\mathrm{d}^2 y^\gamma)(Y)\big) =$$
$$\sum_k F^\gamma_k(X,Y)\big((\mathrm{d}^2 x^k)(X)\big) + \sum_{ij} F^\gamma_{ij}(X,Y)\big((\mathrm{d}x^i \cdot \mathrm{d}x^j)(X)\big)\ ,$$

wherefrom

$$\int {<}F(X,Y)^*\big((\mathrm{d}^2 y^\gamma)(Y)\big), \mathcal{D}X{>} =$$
$$\sum_k \int F^\gamma_k(X,Y)\,\mathrm{d}X^k + \sum_{ij} \int F^\gamma_{ij}(X,Y)\,\tfrac12\,\mathrm{d}\langle X^i, X^j\rangle\ .$$

The coefficients F^γ_k and F^γ_{ij} were given earlier; they are $F^\gamma_k(x,y) = e^\gamma_k(x,y)$ and $F^\gamma_{ij}(x,y) = \frac12\big(e^\gamma_{ij}(x,y) + e^\gamma_{ji}(x,y)\big)$. Consequently, one has $\int F^\gamma_k(X,Y)\,\mathrm{d}X^k = \int e^\gamma_k(X,Y)\,\mathrm{d}X^k$ and $\int F^\gamma_{ij}(X,Y)\,\mathrm{d}\langle X^i, X^j\rangle = \int e^\gamma_{ij}(X,Y)\,\mathrm{d}\langle X^i, X^j\rangle$; so both sides of $(*)$ are equal.

We now prove the last part of Theorem 4. Let σ be a 1-form on $M \times N$, such that ${<}\sigma, \dot z{>} = 0$ whenever z is a solution to the ODE. For each point (x,y) in $M \times N$, $\sigma(x,y)$ is a linear form on the product $\mathrm{T}_x M \times \mathrm{T}_y N$, that is, an element of $\mathrm{T}^*_x M \oplus \mathrm{T}^*_y N$; in other words, $\sigma(x,y) = \pi(x,y) + \rho(x,y)$ with $\pi(x,y) \in \mathrm{T}^*_x M$ and $\rho(x,y) \in \mathrm{T}^*_y N$. For $z(t) = \big(x(t), y(t)\big)$ any solution of the ODE, write

$$\begin{aligned}
0 &= < \sigma(z), \dot z > = < \pi(x,y), \dot x > + < \rho(x,y), \dot y > \\
&= < \pi(x,y), \dot x > + < \rho(x,y), e(x,y)\dot x > \\
&= < \pi(x,y), \dot x > + < e(x,y)^* \rho(x,y), \dot x > = < \tau(x,y), \dot x > ,
\end{aligned}$$

where $\tau(x,y)$ is defined as $\pi(x,y) + e(x,y)^*\rho(x,y) \in \mathrm{T}^*_x M$. This implies $\tau(x,y) = 0$ for all $(x,y) \in M \times N$; therefore, if $Z_t = (X_t, Y_t)$ is a solution of the Stratonovich SDE,

$$\begin{aligned}
\int < \sigma(Z), \circ\mathrm{d}Z > &= \int < \pi(X,Y), \circ\mathrm{d}X > + \int < \rho(X,Y), \circ\mathrm{d}Y > \\
&= \int < \pi(X,Y), \circ\mathrm{d}X > + \int < e(X,Y)^* \rho(X,Y), \circ\mathrm{d}X > \\
&= \int < \tau(X,Y), \circ\mathrm{d}X > = 0 . \qquad \square
\end{aligned}$$

Examples. Three examples of ODEs were given at the beginning of the previous section.

The first one was a time-dependent vector field $V(x,t)$ on M; we saw that $e(t,x) : \mathrm{T}_t\mathbb{R} \to \mathrm{T}_x M$ given by $e(t,x)\frac{\mathrm{d}}{\mathrm{d}t} = V(x,t)$ is an ODE from $\mathbb{R}$ to M. From the viewpoint of stochastic differential geometry, the Stratonovich SDE associated to this e is not interesting at all. The solutions to the ODE are the integral curves $\gamma(t)$ of the vector field, and the corresponding Stratonovich SDE is $\circ\mathrm{d}X = V(T,X)\circ\mathrm{d}T$, driven by a real semimartingale T, and where the unknown is a semimartingale X in M; its solutions are simply the processes $X_t = \gamma \circ T_t$, whose paths remain on the smooth integral curves γ of V. (The simplest proof of this fact is by identification: just check that $\gamma \circ T_t$ is indeed a solution to the SDE.)

As an exercise, you may compute the second-order SDE $F(t,x) : \mathbb{T}_t\mathbb{R} \to \mathbb{T}_x M$ equivalent to the Stratonovich SDE; but remember that this calculation is never needed, the Stratonovich SDE can be written directly from the ODE. The answer to this exercise is

$$F(t,x)\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big) = V(x,t) , \qquad F(t,x)\Big(\frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big) = V^2(x,t) + \frac{\partial}{\partial t}V(t,x) .$$

Here, $V^2(t,x)$ stands for the second-order differential operator equal to the square of the first-order differential operator $V(t,x)$ when t is kept fixed, and $\frac{\partial}{\partial t}V(t,x)$ means the derivative with respect to t of the vector $V(t,x) \in \mathrm{T}_x M$ for fixed x. This expression of F makes it plain that solving the Stratonovich SDE requires from V more smoothness than solving the ODE.

The second example of an ODE consisted in integrating a 1-form σ along curves. Not surprisingly, the corresponding Stratonovich SDE is just Stratonovich integration of 1-forms along semimartingales. We shall not elaborate further on this; if you are curious enough to compute the corresponding second-order SDE F (let us repeat that this is not necessary!), you will find

$$\forall L \in \mathbb{T}_x M \qquad F(x,y)\, L = <\sigma{\cdot}\sigma, L>\, \mathrm{D}^2 + <\mathrm{d}\sigma, L>\, \mathrm{D} \;\in\; \mathbb{T}_y\mathbb{R}\ .$$

The third example, horizontal lifts of curves, is of everyday use in stochastic differential geometry: the Stratonovich transfer principle is instrumental in the definition of parallel transport of vectors along paths of semimartingales in manifolds, and, more generally, of transport of all kinds of tensors and other geometric objects. Because of the difficulty due to the fact that horizontal lifts are not defined for all pairs (x, y), but only when y is above x, the discussion of these transports is postponed till after Proposition 7.

As a forth example, we shall now revisit the Stratonovich SDE used in Chapter ??? to construct a manifold-valued diffusion with a given infinitesimal generator.

Suppose given $n+1$ smooth vector fields $V_0, \ldots, V_n$ on a manifold N (they are called $X_0, \ldots, X_n$ in Chapter ???). Set $M = \mathbb{R}^{n+1}$, and define an ODE $e(x, y) : \mathrm{T}_x\mathbb{R}^{n+1} \to \mathbb{T}_y N$ by identifying $\mathrm{T}_x\mathbb{R}^{n+1}$ with $\mathbb{R}^{n+1}$ itself, and putting, for $r = (r^0, \ldots, r^n) \in \mathbb{R}^{n+1} \approx \mathrm{T}_x\mathbb{R}^{n+1}$,

$$e(x,y)(r) = r^0 V_0(y) + \ldots + r^n V_n(y)\ .$$

Given a smooth curve $x(t)$ in $\mathbb{R}^{n+1}$, a curve $y(t)$ in N is a solution to the ODE if $\dot y(t) = (x^0)'(t) V_0\big(y(t)\big) + \ldots + (x^n)'(t) V_n\big(y(t)\big)$. The corresponding Stratonovich SDE is $\circ\mathrm{d}Y = V_0(Y) \circ\mathrm{d}X^0 + \ldots + V_n(Y) \circ\mathrm{d}X^n$, driven by an $\mathbb{R}^{n+1}$-valued semimartingale $(X^0, \ldots, X^n)$. This equation means that for every smooth function f on N,

$$f\circ Y = f(Y_0) + \int (V_0 f)(Y) \circ\mathrm{d}X^0 + \ldots + \int (V_n f)(Y) \circ\mathrm{d}X^n\ .$$

These Stratonovich integrals of real semimartingales will now be converted into Itô ones. It is an easy computation:

$$\begin{aligned} \int (V_i f)(Y) \circ\mathrm{d}X^i &= \int (V_i f)(Y)\, \mathrm{d}X^i + \tfrac12 \langle (V_i f)(Y), X^i\rangle \\ &= \int (V_i f)(Y)\, \mathrm{d}X^i + \tfrac12 \big\langle \textstyle\int \sum_j (V_j V_i f)(Y)\, \mathrm{d}X^j, X^i \big\rangle \\ &= \int (V_i f)(Y)\, \mathrm{d}X^i + \tfrac12 \sum_j \int (V_j V_i f)(Y)\, \mathrm{d}\langle X^j, X^i\rangle\ , \end{aligned}$$

which implies

$$f\circ Y = f(Y_0) + \sum_k \int (V_k f)(Y)\, \mathrm{d}X^k + \tfrac12 \sum_{ij} \int (V_i V_j f)(Y)\, \mathrm{d}\langle X^i, X^j\rangle\ .$$

[This formula is also a straightforward consequence of the expression of the Schwartz morphisms $F(x, y)$ that make up the second-order SDE associated to e; in turn, these $F(x,y)$ are easily obtained from the accelerations of the curves solving the ODE:

$$<\mathrm{d}^2 f, \ddot{y}(t)> = \sum_k (V_k f)\big(y(t)\big)\,\big(x^k\big)''(t) + \sum_{ij} (V_i V_j f)\big(y(t)\big)\,\big(x^i\big)'(t)\,\big(x^j\big)'(t)\ .\,]$$

The link with diffusions arises when one takes the driving semimartingale $(X_t^0, X_t^1, \ldots, X_t^n)$ equal to $(t, B_t^1, \ldots, B_t^n)$, where $(B^1, \ldots, B^n)$ is an n-dimensional Brownian motion. Then all brackets $\langle X^i, X^j \rangle$ vanish except $\langle X^i, X^i \rangle_t = t$ for $i \neq 0$. So one ends up with $f \circ Y = f(Y_0) + M + A$, with the martingale part given by $M = \sum_{k=1}^n \int (V_k f)(Y)\,\mathrm{d}B^k$, and, more important, the finite variation part A equal to $\int (Lf)(Y)\,\mathrm{d}t$, where L is the diffusor field defined by $Lf = V_0 f + \frac{1}{2}\sum_{i=1}^n V_i V_i f$. In other words, all solutions to the Stratonovich SDE are diffusions on N with infinitesimal generator L. As every smooth diffusor field admits a decomposition as a sum $V_0 + \frac{1}{2}\sum_{i=1}^n V_i V_i$ for a suitable choice of n and of the vector fields $V_0, V_1, \ldots, V_n$, this Stratonovich SDE is a convenient tool to construct general diffusions. (But one must be aware that, given L, there exists no canonical way of choosing the vector fields V_i, and the structure of the stochastic flow associated to the SDE is strongly influenced these choices.)

Time discretization in Stratonovich integrals and SDEs

Definition. Let M be a manifold of class C^3 (at least). An *interpolation rule* on M is a measurable map $I : M \times M \times [0,1] \to M$ such that

(i) for all x and y in M, $I(x,y,0) = x$ and $I(x,y,1) = y$;
(ii) for all $x \in M$ and $t \in [0,1]$, $I(x,x,t) = x$;
(iii) for x and y close enough (i.e., for (x,y) in some neighbourhood G of the diagonal in M), $t \mapsto I(x,y,t)$ is a C^3 curve;
(iv) uniformly for $(x,y) \in (K \times K) \cap G$, where K is any compact subset of M,

$$\frac{\partial^m}{\partial t^m} I(x,y,t) \in O\big(\|x-y\|^m\big) \qquad \text{for } m \in \{1,2,3\}.$$

Interpolation rules are only required to be measurable because topological obstructions may prevent continuous interpolation rules to exist: think of the case that M is not connected, or not simply connected.

At first sight, (iv) looks meaningless; but it can be understood using local coordinates, or using a global embedding of M into a vector space, and it then turns out to be intrinsic. For instance, choosing K included in the domain of some local chart gives a meaning to (iv) by (locally) replacing M with an open subset of $\mathbb{R}^d$; and it is not difficult to verify that the so-obtained condition is invariant under C^3 diffeomorphisms of $\mathbb{R}^d$.

On any given manifold, there always exist many interpolation rules. To construct one of them, one may for instance embed M into a Euclidean vector space E by Whitney's theorem, and then define $I(x,y,t)$ as the point of M closest (in E) to the affine interpolation $(1-t)x+ty \in E$. This point is unique when x and y are close enough; when it is not unique, any measurable choice of $I(x,y,t)$ can be used.

Other interpolation rules, the *geodesic interpolations*, are often used. They are defined in such a way that the curve $t \mapsto I(x,y,t)$ is a geodesic when x and y are close enough; they require M to be endowed with an extra structure (a Riemannian metric, or, more generally, a linear connection). We shall come back to this later, when discussing Itô integrals and SDEs.

Definition. By a *subdivision*, we mean a sequence $\Upsilon = (T_n)_{n\geqslant 0}$ of stopping times such that $T_0 = 0$, $T_{n+1} \geqslant T_n$, and $\lim_{n\to\infty} T_n = \infty$ a.s. The *mesh-size* of the subdivision Υ on the interval $[0,t]$ is the r.v. $|\Upsilon|_{[0,t]} = \sup_n [T_{n+1}\wedge t - T_n \wedge t]$.

Theorem 5 (Approximating Stratonovich integrals and SDEs). *Let M and N be two C^3 manifolds, I an interpolation rule on M, and e a C^2 ODE from M to N. Let X be a semimartingale in M and Y a semimartingale in N (with a possibly finite explosion time ζ) such that*

$$\circ\mathrm{d}Y = e(X,Y)\circ\mathrm{d}X \ .$$

For each subdivision $\Upsilon = (T_n)_{n\geqslant 0}$, define a (non adapted) process X^Υ in M by

$$X^\Upsilon_t = I\big(X_{T_n}, X_{T_{n+1}}, \tfrac{t-T_n}{T_{n+1}-T_n}\big) \qquad \text{for } T_n \leqslant t \leqslant T_{n+1},$$

and a (non adapted) process Y^Υ in N as the solution of the ODE

$$\dot Y^\Upsilon = e(X^\Upsilon, Y^\Upsilon)\,\dot X^\Upsilon \ ; \qquad Y^\Upsilon_0 = Y_0$$

(Y^Υ is defined on a random interval $[\![0,\zeta^\Upsilon[\![$, where ζ^Υ is the first time when X^Υ is no longer piecewise C^3 or when Y^Υ tends to ∞_N).

Let $(\Upsilon_k)_{k\in\mathbb{N}}$ be a sequence of subdivisions whose mesh-sizes $|\Upsilon_k|_{[0,t]}$ tend to 0 in probability for each $t>0$. Then $\liminf_{k\to\infty} \zeta^{\Upsilon_k} \geqslant \zeta$, where the $\liminf$ is understood in probability, and Y^{Υ_k} converges to Y on the time interval $[\![0,\zeta[\![$, uniformly on compact time intervals in probability.

The statement of Theorem 5 is obscured by the possibility that Y or Y^Υ is not defined at all times. If you suppose further that $\zeta \equiv \infty$ and that $I(x,y,t)$ is C^3 in t for all x and y (and not only when they are close enough to each other), and that the solutions to the ODE e cannot explode in finite time, then Y^Υ_t is defined for all t, and the piecewise C^2 process Y^{Υ_k} tends to the semimartingale Y uniformly on any time interval $[0,t]$ in probability.

A useful particular case of Theorem 5 is when $N = \mathbb{R}$ and e is the ODE given by $e(x,y) = \sigma(x)$, where σ is a 1-form on M:

Corollary. *Let σ be a* C^2 *1-form on M and $(\Upsilon_k)_{k\in\mathbb{N}}$ a sequence of subdivisions whose mesh-sizes $|\Upsilon_k|_{[0,t]}$ tend to 0 in probability for each $t>0$. The real-valued, non adapted process $\int<\sigma,\dot{X}^{\Upsilon_k}>\mathrm{d}t$ converges to the Stratonovich integral $\int<\sigma,\circ\mathrm{d}X>$ uniformly on compact time-intervals in probability when $k\to\infty$.*

The hypothesis that σ is C^2 is too strong: the corollary remains true when M is C^2 and σ is C^1. The need for more regularity is an artifact due to our itinerary via differential equations, which require more care than mere integrals.

An important consequence of Theorem 5 is a further conservation property of the Stratonovich transfer principle, which complements the one we already saw at the end of Theorem 4.

Proposition 7. *Let e be a* C^2 *ODE between two* C^3 *manifolds M and N, and P a closed subset of $M\times N$ with the property that every* C^1 *curve $z(t)=\big(x(t),y(t)\big)$ in $M\times N$ verifying $\dot{y}=e(x,y)\,\dot{x}$ and $z(0)\in P$, is completely included in P.*

If $Z=(X,Y)$ is any semimartingale in $M\times N$ (with a possibly finite lifetime) such that $\circ\mathrm{d}Y=e(X,Y)\circ\mathrm{d}X$ and $Z_0\in P$, then Z is valued in P.

This applies for instance to the ODEs transforming a curve $x(t)$ in M into its horizontal lift $y(t)$ in some fibre bundle N over M. We saw earlier that such an ODE e is usually not everywhere defined on $M\times N$, but only on the subset $P\subset M\times N$ consisting of all points (x,y) such that y is above x. Proposition 7 legitimates the use of Stratonovich transfer in this situation: extend e to the whole product $M\times N$, apply Theorem 4, and use Proposition 7 to ensure that the Stratonovich horizontal lift of any semimartingale X in M remains above X and therefore does not depend upon the choice of the extension of e.

For instance, the *parallel transport* of vectors or tensors along smooth curves gives rise by Stratonovich transfer to the *Stratonovich parallel transport* of vectors or tensors along paths of semimartingales. This is a most important tool in stochastic differential geometry.

Itô stochastic integrals

We now turn to intrinsic Itô stochastic calculus on manifolds endowed with connections. This topic is less important than Stratonovich stochastic calculus, and can be skipped (or skimmed through) in a first encounter with stochastic differential geometry. But the last section, on lifts of semimartingales, introduces some of the most basic tools in this field and should not be omitted; it is postponed to the end only because it uses connections.

Besides Stratonovich stochastic integration, there is another way to integrate 1-forms along semimartingales, also by converting 1-forms into second-order forms. It consists in a geometric device transforming individual covectors into codiffusors, or, dually, diffusors into tangent vectors. It is pointwise, so it needs less regularity than symmetric differentiation; but its drawback is that it is not canonically associated to the differentiable structure of the manifold: it has to be superimposed as an additional structure.

It turns out that this extra datum is equivalent to a torsion-free linear connection on the tangent bundle (see Chapter ???). We shall just call it a connection, for short, but you should remember that this term has a much broader meaning in other contexts.

Definition. Given a point x in a manifold M, a *connection at* x is a linear map $\Gamma_x : \mathbb{T}_x M \to \mathrm{T}_x M$ whose restriction to the subspace $\mathrm{T}_x M \subset \mathbb{T}_x M$ is identity.

So Γ_x is a projector onto $\mathrm{T}_x M$, characterized by the linear subspace $\operatorname{Ker}\Gamma_x \subset \mathbb{T}_x M$, supplementary to $\mathrm{T}_x M$ in $\mathbb{T}_x M$.

Given a local chart around x, a connection at x is characterized by its coefficients Γ^k_{ij} (called the *Christoffel symbols*), such that $\Gamma^k_{ij} = \Gamma^k_{ji}$ and that

$$\forall\, L = \sum_{ij} L^{ij}\mathrm{D}_{ij} + \sum_k L^k \mathrm{D}_k \in \mathbb{T}_x M \ , \quad \Gamma L = \sum_k \Big(\sum_{ij} \Gamma^k_{ij} L^{ij} + L^k\Big)\mathrm{D}_k \ .$$

Dually, the adjoint $\Gamma^*_x : \mathrm{T}^*_x M \to \mathbb{T}^*_x M$ makes covectors into codiffusors. If $\sigma \in \mathrm{T}^*_x M$ has coefficients σ_k in the local chart, that is, if $\sigma = \sum_k \sigma_k \, \mathrm{d}x^k(x)$,

$$\Gamma^*\sigma = \sum_k \sigma_k \, \Gamma^*\big(\mathrm{d}x^k(x)\big) = \sum_k \sigma_k \Big(\mathrm{d}^2 x^k(x) + \sum_{ij} \Gamma^k_{ij} \, \mathrm{d}x^i(x)\!\cdot\!\mathrm{d}x^j(x)\Big) \ .$$

Definition. A *connection* Γ on a C^p manifold M is a family $(\Gamma_x)_{x\in M}$ of connections at each point of M, such that the map $x \mapsto \Gamma_x$ is of class C^{p-2}.

A connection Γ transforms diffusor fields into vector fields, by acting separately at each $x \in M$. In particular, if V and W are vector fields, VW is a diffusor field and $\Gamma(VW)$ is a vector field. Given Γ, there exists a (unique) linear connection ∇ on the tangent bundle of M (in the sense of Chapter ???) such that $\nabla_V W = \Gamma(VW)$ for all vector fields V and W; and ∇ is torsion-free. Conversely, if ∇ is any torsion-free linear connection on the tangent bundle, there exists a unique connection Γ such that $\Gamma(VW) = \nabla_V W$ for all vector fields V and W. Moreover, ∇ and Γ have the same Christoffel symbols Γ^k_{ij}.

An important example is the flat connection. If M is a vector space (or an affine space), every differential operator $L \in \mathbb{T}_x M$ can be canonically written as the sum of its first order part $V \in \mathrm{T}_x M$ and its purely second-order part: use global, linear (or affine) coordinates on M and check that the decomposition is invariant under linear (or affine) changes of coordinates. The flat connection on M (also called the canonical connection) is the projection $L \mapsto V$.

Another important example is the connection Γ constructed as follows on a manifold M embedded (or immersed) in a Euclidean vector space E. Denote by $i : M \hookrightarrow E$ the immersion. The push-forward i_*L of any $L \in \mathbb{T}_xM$ is an element of $\mathbb{T}_{ix}E$; the flat connection Γ_E on E transforms i_*L into the vector $\Gamma_E i_*L \in \mathrm{T}_{ix}E$; then this vector can be orthogonally projected onto the subspace $\mathrm{T}_{ix}(iM) \subset \mathrm{T}_{ix}E$ using the Euclidean structure on $\mathrm{T}_{ix}E \approx E$; last, this element of $\mathrm{T}_{ix}(iM) \subset \mathrm{T}_{ix}E$ can be pulled back to a vector $V \in \mathrm{T}_xM$ because the linear map $i_* : \mathrm{T}_xM \to \mathrm{T}_{ix}E$ is one-to-one with range $\mathrm{T}_{ix}(iM)$. The connection Γ is defined by $\Gamma L = V$; we shall call it the connection associated to the immersion.

It is a particular instance of the Levi-Civita connection defined on a Riemannian manifold; this will be defined later.

Using Whitney's embedding theorem and the above construction one sees that every manifold can be endowed with a connection.

On a manifold M endowed with a connection Γ, a particular set of curves, the *geodesics*, can be defined. The traditional definition, $\nabla_{\dot\gamma}\dot\gamma = 0$, is in terms of covariant derivatives (see Chapter ???); in the present second-order framework, a curve $t \mapsto \gamma(t)$ in (M, Γ) is a geodesic if and only if $\Gamma\ddot\gamma(t) = 0$ for all t.

If M is an affine space with the flat connection, the geodesics are the uniforms motions in M (straight lines run at constant speed).

If M is a submanifold of a Euclidean vector space E, endowed with the connection Γ associated to the immersion $M \hookrightarrow E$, an M-valued curve γ is a geodesic of (M, Γ) if and only if for each t the second derivative $\gamma''(t)$ (this is a vector in E) is orthogonal to the tangent space $\mathrm{T}_{\gamma(t)}M$, considered as a subspace of E.

Exercise. Let M be endowed with a connection Γ. If $\gamma^i(t)$ denote the coordinates of a curve γ in a local chart, γ is a geodesic if and only if

$$(\gamma^k)''(t) = -\sum_{ij} \Gamma^k_{ij}\big(\gamma(t)\big)\,(\gamma^i)'(t)\,(\gamma^j)'(t)\ ,$$

where Γ^k_{ij} are the Christoffel symbols of Γ.

Given any tangent vector $V \in \mathrm{T}_xM$, there exists a unique maximal geodesic γ such that $\gamma(0) = x$ and $\dot\gamma(0) = V$.

If Γ is a connection, Γ^* dually transforms 1-forms into second-order forms. One has $<\Gamma^*\sigma, L> = <\sigma, \Gamma L>$ for all diffusor fields L and 1-forms σ (by definition of the adjoint); and $\mathrm{R}\Gamma^*\sigma = \sigma$ because Γ preserves each tangent vector.

Definition. Let M be a manifold endowed with a connection Γ, X a semimartingale in M, and Σ a T^*M-valued, locally bounded, predictable process above X. The stochastic integral $\int <\Gamma^*\Sigma, \mathcal{D}X>$ is called the *Itô integral of* Σ *along* X, and denoted by $\int <\Sigma, \Gamma\mathcal{D}X>$.

Some properties of this integral are summarized in the following theorem. We shall not prove it; the case when M admits global coordinates is left as an exercise to the reader. (Hint: formally, $\Gamma\mathcal{D}X_t$ is the infinitesimal tangent vector $\sum_k\bigl(\sum_{ij}\Gamma^k_{ij}(X_t)\,\frac{1}{2}\,\mathrm{d}\langle X^i,X^j\rangle_t+\mathrm{d}X^k_t\bigr)\mathrm{D}_k$, which "belongs" to $\mathrm{T}_{X_t}M$.)

Theorem 6. *The Itô integral is characterized by the following properties: it depends linearly upon* Σ, *and*

(i) for every $f\in\mathrm{C}^2(M)$, *one has the intrinsic Itô formula*

$$f\circ X=f(X_0)+\int<\mathrm{d}f,\Gamma\mathcal{D}X>+\int<\mathrm{Hess}\,f,\mathcal{D}X>\ ,$$

where $\mathrm{Hess}\,f$ *denotes the second-order form* $\mathrm{d}^2f-\Gamma^*\mathrm{d}f$ *(which verifies* $\mathrm{R}\,\mathrm{Hess}\,f=\mathrm{d}f-\mathrm{d}f=0$, *so* $\int<\mathrm{Hess}\,f,\mathcal{D}X>$ *has finite variation);*

(ii) if H *is any predictable, locally bounded, real-valued process,*

$$\int<H\Sigma,\Gamma\mathcal{D}X>=\int H\,\mathrm{d}\bigl(\textstyle\int<\Sigma,\Gamma\mathcal{D}X>\bigr)\ .$$

Furthermore, one also has:

(iii) if P *verifies the same hypothesis as* Σ,

$$\tfrac{1}{2}\,\bigl\langle\textstyle\int<\Sigma,\Gamma\mathcal{D}X>\,,\int<P,\Gamma\mathcal{D}X>\bigr\rangle=\displaystyle\int<\Sigma{\cdot}P,\mathcal{D}X>\ ;$$

(iv) if T *is a stopping time,*

$$\Bigl(\int<\Sigma,\Gamma\mathcal{D}X>\Bigr)^T=\int<\Sigma^T,\Gamma\mathcal{D}X^T>\ ;$$

(v) if Σ *is a* T^*M*-valued semimartingale above* X, *both the Stratonovich and Itô integrals of* Σ *along* X *are defined, and their difference has finite variation. (Equivalently: both integrals have the same martingale part, namely, the local martingale* $\int<\Sigma,\mathrm{d}\overset{\mathrm{m}}{X}>$*.)*

Definition. Let M be endowed with a connection Γ. A semimartingale X in M is a *Γ-martingale* if for each function $f\in\mathrm{C}^2(M)$, the Itô integral $\int<\mathrm{d}f,\Gamma\mathcal{D}X>$ is a (real) local martingale.

When X is a Γ-martingale, $\int<\Sigma,\Gamma\mathcal{D}X>$ is also a local martingale for any T^*M-valued, locally bounded, predictable processes Σ above X.

If M is an affine space and Γ the flat connection, X is a Γ-martingale if and only if it is a local martingale (in the usual sense).

If i is an immersion of M in a Euclidean space E and if Γ is the connection associated to this immersion, a semimartingale X in M is a Γ-martingale if and only if in the canonical decomposition $iX=iX_0+M+A$ of iX in E, the part with finite variation $\mathrm{d}A_t$ remains orthogonal to $\mathrm{T}_{iX_t}M$ in the following sense: A has the form $\int H\,\mathrm{d}B$, where B is a real-valued, increasing process, and the E-valued process H verifies $H_t(\omega)\perp\mathrm{T}_{iX_t(\omega)}M$ for all (t,ω).

Exercise. Let M be endowed with a connection $\mathit{\Gamma}$; suppose M admits global coordinates. An M-valued process X is a $\mathit{\Gamma}$-martingale if and only if its coordinates X^i verify

$$X^k = X^k_0 + M^k - \tfrac{1}{2}\sum_{ij}\int \mathit{\Gamma}^k_{ij}(X)\,\mathrm{d}\langle M^i, M^j\rangle\ ,$$

where $M^1, \dots, M^d$ are local martingales and $\mathit{\Gamma}^k_{ij}$ are the Christoffel symbols of $\mathit{\Gamma}$.

Exercise (Convex functions on M). Let $f \in \mathrm{C}^2(M)$ and M be endowed with a connection $\mathit{\Gamma}$. Show that the codiffusor $\operatorname{Hess} f(x)$ is positive for all $x \in M$ if and only if, for every geodesic γ, the function $f \circ \gamma$ is convex (on the interval where γ is defined). When these conditions are met, $f \circ X$ is a local submartingale for every $\mathit{\Gamma}$-martingale X.

Itô stochastic differential equations

Definition. Let X (resp. Y) be a semimartingale in a C^2 manifold M (resp. N) endowed with a connection $\mathit{\Gamma}_M$ (resp. $\mathit{\Gamma}_N$). Let E be a locally bounded, predictable process such that $E(t,\omega)$ is a linear map from $\mathrm{T}_{X_t(\omega)}M$ to $\mathrm{T}_{Y_t(\omega)}N$. One says that Y *is a solution to the Itô stochastic differential equation*

$$\mathit{\Gamma}_N \mathcal{D}Y = E\,\mathit{\Gamma}_M \mathcal{D}X$$

if for every locally bounded, predictable, T^*N-valued process $\mathit{\Sigma}$ above Y, the Itô stochastic integral $\int <\mathit{\Sigma}, \mathit{\Gamma}_N\mathcal{D}Y>$ of $\mathit{\Sigma}$ along Y equals the Itô stochastic integral $\int <E^*\mathit{\Sigma}, \mathit{\Gamma}_M\mathcal{D}X>$ of $E^*\mathit{\Sigma}$ along X. In fact, it suffices to have it for $\mathit{\Sigma} = (\mathrm{d}f)(Y)$, where f ranges over $\mathrm{C}^2(N)$.

Comparing this definition with that of $\mathit{\Gamma}$-martingales, one immediately obtains the following preservation property:

Proposition 8 (Itô SDEs preserve martingales). *With the same notations as in the preceding definition, suppose Y to be a solution to the Itô SDE $\mathit{\Gamma}_N\mathcal{D}Y = E\,\mathit{\Gamma}_M\mathcal{D}X$. If X is a $\mathit{\Gamma}_M$-martingale, Y is a $\mathit{\Gamma}_N$-martingale.*

Exactly like Stratonovich SDEs, Itô ones can be considered a particular instance of second-order SDEs owing to a general transfer principle:

Theorem 7 (Itô transfer principle). *a) Let x (resp. y) be a point in a C^2 manifold M (resp. N) endowed with a connection $\mathit{\Gamma}_M$ (resp. $\mathit{\Gamma}_N$); let e be a linear map from T_xM to T_yN. There exists a unique Schwartz morphism $f : \mathbb{T}_xM \to \mathbb{T}_yN$ such that*

- *the restriction of f to T_xM is e;*

- *the following diagram commutes:*

$$\begin{array}{ccc} \mathbb{T}_x M & \xrightarrow{f} & \mathbb{T}_y N \\ \big\downarrow{\scriptstyle \mathit{\Gamma}_M} & & \big\downarrow{\scriptstyle \mathit{\Gamma}_N} \\ \mathrm{T}_x M & \xrightarrow{e} & \mathrm{T}_y N \end{array} .$$

b) Let $e(x,y)$ *be an ODE from* M *to* N*; for each* $(x,y) \in M \times N$*, denote by* $f(x,y)$ *the Schwartz morphism associated to* $e(x,y)$ *by a). The Itô SDE*

$$\mathit{\Gamma}_N \mathcal{D}Y = e(X,Y)\, \mathit{\Gamma}_M \mathcal{D}X$$

is equivalent to the second-order SDE

$$\mathcal{D}Y = f(X,Y)\, \mathcal{D}X \ .$$

The Itô transfer principle is far less used (and much less known) than the Stratonovich one. Its main weakness is that it does not enjoy the same powerful conservation properties as the Stratonovich transfer principle. But one encounters it from times to times; as for an example, the so-called 'damped stochastic parallel transport' is an instance of an Itô SDE. Another drawback of the Itô transfer is that it needs both M and N to be endowed with connections.

On the other hand, it necessitates less regularity than the Stratonovich transfer, and it extends to equations where e (and f) depend not only upon the current position (X,Y) of the driving process and the solution, but also on their past, and, more generally, upon (t,ω) in a predictable way.

The proof of Theorem 7 is left to the reader. Part a) can be established in local coordinates: if e is given by its coefficients e_i^α, the Schwartz morphism f can be obtained by

$$f_k^\gamma = e_k^\gamma \ ; \qquad f_{ij}^\gamma = \sum_k e_k^\gamma \mathit{\Gamma}_{ij}^k - \sum_{\alpha\beta} e_i^\alpha e_j^\beta \mathit{\Gamma}_{\alpha\beta}^\gamma \ ;$$
$$f_k^{\alpha\beta} = 0 \ ; \qquad f_{ij}^{\alpha\beta} = e_i^\alpha e_j^\beta \ .$$

Part b) is esentially the following formal computation

$$\mathit{\Gamma}_N \mathcal{D}Y - e\, \mathit{\Gamma}_M \mathcal{D}X = \mathit{\Gamma}_N \mathcal{D}Y - \mathit{\Gamma}_N f\, \mathcal{D}X = \mathit{\Gamma}_N(\mathcal{D}Y - f\, \mathcal{D}X) \ ,$$

which must be made rigorous by integrating 1-forms on the left.

Corollary (Existence and uniqueness in Itô SDEs). *Suppose* e *to be an ODE from* M *to* N *such that the map* $(x,y) \mapsto e(x,y)$ *is locally Lipschitz. Given a driving semimartingale* X *in* M *and an initial r.v.* y_0 *in* N*, the Itô SDE*

$$\mathit{\Gamma}_N \mathcal{D}Y = e(X,Y)\, \mathit{\Gamma}_M \mathcal{D}X \ ; \qquad Y_0 = y_0$$

has a unique solution Y *(up to a possibly finite explosion time).*

This is an immediate consequence of Theorems 7 and 3, provided the second-order SDE f obtained from e by the Itô transfer principle is shown to be locally Lipschitz. This can be seen on the explicit formulae giving f from e in the sketched proof of Theorem 7.

Time discretization in Itô integrals and SDEs

Definition. Let M be endowed with a connection $\mathit{\Gamma}$. A *geodesic interpolation rule* is an interpolation rule I such that, for all (x,y) in some neighbourhood of the diagonal in $M \times M$, the curve $t \mapsto I(x,y,t)$ is a geodesic.

On any given $(M,\mathit{\Gamma})$, a geodesic interpolation rule always exists; moreover it is essentially unique, in the sense that any two agree on some neighbourhood of the diagonal. (When x and y are close to each other, there may exist many geodesics linking them, but there exists only one "small" one; and the definition of an interpolation rule forces $t \mapsto I(x,y,t)$ to be that one.)

Theorem 8 (Approximating Itô integrals and SDEs). *Let e be a locally Lipschitz ODE between two manifolds M and N respectively endowed with connections $\mathit{\Gamma}_M$ and $\mathit{\Gamma}_N$; let I be a geodesic interpolation rule on M. Suppose given a semimartingale X in M and a semimartingale Y in N (with a possibly finite life time ζ), verifying the Itô SDE*

$$\mathit{\Gamma}_N \mathcal{D}Y = e(X,Y)\,\mathit{\Gamma}_M \mathcal{D}X \ .$$

For each subdivision $\Upsilon = (T_n)_{n\geqslant 0}$, define a (non adapted) process X^Υ in M by

$$X^\Upsilon_t = I\big(X_{T_n}, X_{T_{n+1}}, \tfrac{t-T_n}{T_{n+1}-T_n}\big) \qquad \text{for } T_n \leqslant t \leqslant T_{n+1},$$

and a (non adapted) process Y^Υ in N as follows: on the interval $[T_n, T_{n+1}]$, Y^Υ_t is the geodesic curve with initial position $Y^\Upsilon_{T_n}$ given by the preceding step (given by $Y^\Upsilon_0 = Y_0$ for $n=0$), and with initial velocity

$$\dot Y^\Upsilon_{T_n} = e(X^\Upsilon_{T_n}, Y^\Upsilon_{T_n})\,\dot X^\Upsilon_{T_n} \ ,$$

where the dots denotes time-derivatives on the right, that is, at time T_n+. The process Y^Υ is defined on a random interval $[\![0,\zeta^\Upsilon[\![$, where ζ^Υ is the first time when X^Υ is no longer piecewise geodesic or when Y^Υ no longer exists.

Let $(\Upsilon_k)_{k\in\mathbb{N}}$ be a sequence of subdivisions whose mesh-sizes $|\Upsilon_k|_{[0,t]}$ tend to 0 in probability for each $t>0$. Then $\liminf_{k\to\infty} \zeta^{\Upsilon_k} \geqslant \zeta$, where the $\liminf$ is taken in probability, and Y^{Υ_k} converges to Y on the time interval $[\![0,\zeta[\![$, uniformly on compact time intervals in probability.

Exactly as with the Stratonovich approximation procedure, Theorem 8 is instrumental in establishing conservation properties for Itô SDEs:

Corollary. *Let e be an ODE between two manifolds M and N endowed with connections, and P a closed subset of $M \times N$ with the following property: for all $(x_0, y_0) \in P$ and all $V \in \mathrm{T}_{x_0}M$, the curve $z(t) = \big(x(t), y(t)\big)$ in $M \times N$ is completely included in P, where $x(t)$ is the geodesic in M with initial conditions $x(0) = x_0$ and $\dot{x}(0) = V$, and $y(t)$ is the geodesic in N with initial conditions $y(0) = y_0$ and $\dot{y}(0) = e(x_0, y_0)\, V$.*

If $Z = (X, Y)$ is any semimartingale in $M \times N$ (possibly with finite lifetime) such that $\mathit{\Gamma}_N \mathcal{D}Y = e(X, Y)\, \mathit{\Gamma}_M \mathcal{D}X$ and $Z_0 \in P$, then Z lives in P.

Indeed, for each Υ, the process (X^Υ, Y^Υ) constructed in Theorem 8 takes its values in P; in the limit, so does also (X, Y). (This proof by time discretization requires M and N to be C^3; a direct proof which needs less regularity is also possible.)

An example where this corollary applies is the horizontal lift of a curve from the base M to a fibre bundle N over M, given a non-linear connection in the sense of Chapter ???. We already saw after Proposition 7 how this kind of conservation property can be used to apply the theory of ODEs and their Stratonovich transfer to situations where $e(x, y)$ is not defined everywhere, but only on a good subset P (in the present case, $(x, y) \in P$ if and only if y is above x). The same arguments apply to the Itô transfer; all we need is a connection $\mathit{\Gamma}_M$ on M and one $\mathit{\Gamma}_N$ on the bundle N, such that the projection on M of any $\mathit{\Gamma}_N$-geodesic of N is a $\mathit{\Gamma}_M$-geodesic of M.

Such connections exist and are well known to geometers; for instance, if M is endowed with a connection, if $N = \mathrm{T}M$ and if the ODE is the parallel transport of vectors along curves (see Chapter ???), two connections on N are particularly interesting.

The first one is called the horizontal connection on $\mathrm{T}M$; the parallel transport of tangent vectors to M along geodesics of M yields geodesics for this horizontal connection, and the Itô transfer principle gives in this case the same result (the stochastic parallel transport along semimartingales) as the Stratonovich transfer principle.

The second interesting connection on $\mathrm{T}M$ is called the complete connection. All Jacobi fields along geodesics of M are geodesics for this connection, and the Itô transfer gives in that case a different result, called the damped stochastic parallel transport along semimartingales.

Another corollary of Theorem 8 compares the Stratonovich and Itô SDEs associated to a given ODE. We just saw that, with a suitable choice of the connection of $\mathrm{T}M$, the Itô stochastic parallel transport is the same as the Stratonovich one. The argument is quite general; comparing Theorems 5 and 8 immediately gives:

Corollary. *Let e be an ODE between two manifolds M and N, each of which is endowed with a connection. Suppose that for any geodesic $x(t)$ in M and any point y_0 of N, the solution $y(t)$ to the ODE*

$$\dot{y} = e(x, y)\, \dot{x}\ ; \qquad y(0) = y_0$$

is a geodesic in N*. The Stratonovich and Itô SDEs*

$$\circ \mathrm{d}Y = e(X,Y) \circ \mathrm{d}X \qquad \textit{and} \qquad \mathit{\Gamma}_N \mathcal{D}Y = e(X,Y)\, \mathit{\Gamma}_M \mathcal{D}X$$

are equivalent.

Lift of a semimartingale to the tangent space

Let M be endowed with a connection $\mathit{\Gamma}$. Recall from Chapter ??? that the (ordinary) parallel transport of a vector along a smooth curve γ in M is the solution above γ to the ODE from M to $N = \mathrm{T}M$

$$\dot{v}(t) = h\big(\gamma(t), v(t)\big)\ ; \qquad v(0) = v_0 \in \mathrm{T}_{\gamma(0)}M\ ,$$

where $h(x,y) : \mathrm{T}_x M \to \mathrm{T}_y N$, defined for y above x, is the horizontal lift associated to $\mathit{\Gamma}$. (In other words, $\mathit{\Gamma}\big(v'(t)\big) = 0$, where $v'(t) \in \mathbb{T}_{\gamma(t)}M$ is the diffusor such that $<\mathrm{d}^2 f, v'(t)> = \frac{\mathrm{d}}{\mathrm{d}t} <\mathrm{d}f, v(t)>$ for all f.)

The stochastic parallel transport of vectors is obtained from the ordinary one by Stratonovich transfer (or, as mentioned in the preceding section, also by Itô transfer). Given a semimartingale X in M, any solution to the Stratonovich SDE associated by transfer to the above ODE is a semimartingale in $\mathrm{T}M$ above X. Given an initial condition $u \in \mathrm{T}_{X_0(\omega)}M$, the value at time t of the solution started from u will be denoted by $/\!/_t u$; this defines a family of maps $/\!/_t(\omega) : \mathrm{T}_{X_0(\omega)}M \to \mathrm{T}_{X_t(\omega)}M$. These maps are defined for all t, and they are linear bijections; in particular, given any $\mathcal{F}_0$-measurable basis $(u_1, \ldots, u_d)$ of the vector space $\mathrm{T}_{X_0}M$, the image of this basis by $/\!/_t$ is an $\mathcal{F}_t$-measurable basis $U(t) = \big(U_1(t), \ldots, U_d(t)\big)$ of $\mathrm{T}_{X_t}M$. (This is called the *parallel transport of frames.*) Call $\big(\rho^1(t), \ldots, \rho^d(t)\big)$ the dual frame to $U(t)$, that is, $\rho^i(t) \in \mathrm{T}^*_{X_t}M$ assigns to each vector of $\mathrm{T}_{X_t}M$ its i-th coordinate relative to the basis $U(t)$. Since the semimartingale ρ^i is above X in T^*M, the Stratonovich integral

$$Y^i = \int <\rho^i, \circ \mathrm{d}X>$$

can be defined. (As explained in the preceding section, it can also be computed as an Itô integral: $Y^i = \int <\rho^i, \mathit{\Gamma}\mathcal{D}X>$.) The process $\widetilde{X}$, taking values in the (random) vector space $\mathrm{T}_{X_0}M$ and defined by

$$\widetilde{X} = \sum_i Y^i u_i\ ,$$

is easily seen not to depend upon the initial choice of the frame $U(0) = (u_1, \ldots, u_d)$; it is called the *lift of* X *to the tangent space* $\mathrm{T}_{X_0}M$. (This is the stochastic analogue of the *rolling without slipping* procedure: when X is a smooth curve, $\widetilde{X}_t$ can be interpreted as the contact point of the flat space $\mathrm{T}_{X_0}M$ with M when the latter “moves” without slipping so as to remain tangent at X_t to the former.)

Proposition 9. *Let M be endowed with a connection Γ.*

(i) A smooth curve γ in M is a geodesic if and only if its lift to $\mathrm{T}_{\gamma(0)}M$ is a uniform motion.

(ii) A semimartingale X in M is a Γ-martingale if and only if its lift to the vector space $\mathrm{T}_{X_0}M$ is a local martingale.

The proof of (i) is an exercise in (ordinary) differential geometry; it uses only the definitions of geodesics and lifts. (ii) can be derived from Proposition 8, because $\widetilde{X}$ is also obtained from X by an Itô procedure; one can also get (ii) as a consequence of (i) and of the corollary of Theorem 8, because in a vector space with its flat connection, the geodesics are the uniform motions.

The most basic objects in stochastic differential geometry are *Brownian motions with values in Riemannian manifolds.* This is where the theory really begins, where the vocabulary and the basic tools introduced in the present notes are put into use—and also where this *Invitation* ends, just after defining these objects.

A manifold is called *Riemannian* if

1) each tangent space $\mathrm{T}_x M$ is endowed with a Euclidean structure (which depends smoothly upon x);
2) M is endowed with a connection, which is *metric*: the parallel transport $/\!/_t : \mathrm{T}_{\gamma(0)}M \to \mathrm{T}_{\gamma(t)}M$ along any smooth curve γ (and also, by the transfer principle, along any semimartingale) is an isometry for those Euclidean structures.

The first theorem in any book of Riemannian geometry says that 2) is a consequence of 1): given a Euclidean structure on each tangent space $\mathrm{T}_x M$, depending smoothly upon x, there exists a unique connection (in our sense, that is, a linear, torsion-free connection in the sense of Chapter ???) which is metric.

A most important object in Riemannian geometry (whether stochastic or not) is the *Laplacian.* This diffusor field Δ on M has many equivalent definitions; one of them considers $\Delta f(x)$ as the trace (for the Euclidean structure on $\mathrm{T}_x M$) of the bilinear form on $\mathrm{T}_x M$

$$(V, W) \mapsto (\nabla \mathrm{d}f)(V, W) = <\mathrm{d}^2 f, VW> - <\mathrm{d}f, \Gamma(VW)>$$

(Another characterization of Δ, known to probabilists only, is the equivalence between (i) and (iii) in Proposition 10!)

The vector space $\mathrm{T}_x M$ can be identified with its own dual space using its Euclidean structure; this yields a scalar product on covectors which we denote by $\mathrm{T}^*_x M \times \mathrm{T}^*_x M \ni (\sigma, \tau) \mapsto <\sigma|\tau> \in \mathbb{R}$.

Proposition 10 and definition. *Let X be a semimartingale in a Riemannian manifold M. The following three statements are equivalent; when they hold, X is called a* Brownian motion *in M:*

(i) X is a diffusion with generator $\frac{1}{2}\Delta$, that is,

$$\forall f \in \mathrm{C}^2(M) \quad f\circ X_t - f\circ X_0 - \tfrac{1}{2}\int_0^t \Delta f(X_s)\,\mathrm{d}s \quad \textit{is a local martingale;}$$

(ii) X is a Γ-martingale, and, for all f and g,

$$\langle f\circ X, g\circ X\rangle_t = \int_0^t <\mathrm{d}f|\mathrm{d}g>(X_s)\,\mathrm{d}s$$

(more generally, for all 1-forms σ and τ,

$$\int_0^t <\sigma{\cdot}\tau, \mathcal{D}X_s> = \int_0^t <\sigma|\tau>(X_s)\,\mathrm{d}s \textit{);}$$

(iii) the lift of X to the Euclidean space $\mathrm{T}_{X_0}M$ is a Brownian motion.

Bibliography

For a first contact with manifold-valued stochastic processes, several books can be recommended to readers already familiar with stochastic calculus; my favorite are:

K.D. Elworthy. Stochastic Differential Equations on Manifolds. *London Math. Soc. Lecture Notes Series 70*, Cambridge University Press 1982.

E.P. Hsu. Stochastic Analysis in Manifolds. *Graduate Studies in Mathematics*, American Math. Soc. 2002.

The next three books provide an introduction to stochastic calculus in general *and* in differentiable manifolds:

W. Hackenbroch and A. Thalmaier. Stochastische Analysis. Eine Einführung in die Theorie der stetigen Semimartingale. Teubner, 1994.

N. Ikeda and S. Watanabe. Stochastic Differential Equations and Diffusion Processes. North-Holland, 1981.

L.C.G. Rogers and D. Williams. Diffusion, Markov Processes and Martingales. Volume 2: Itô Calculus. John Wiley & Sons, 1987.

Among these five books, only Hackenbroch-Thalmaier provides an introduction to Schwartz' second-order language. Here is a sample of references where continuous semimartingales in manifolds are considered from that viewpoint. (In some of them, second order does not come in the same way as in Schwartz, but via Itô bundles or Itô stochastic integrals.)

M. Arnaudon. Connexions et martingales dans les groupes de Lie. *Séminaire de Probabilités XXVI, LNM 1526,* Springer, 1992.

M. Arnaudon. Caractéristiques locales des semi-martingales et changements de probabilités. Ann. Inst. Henri Poincaré 29, 251–267, 1993.

M. Arnaudon. Semi-martingales dans les espaces homogènes. Ann. Inst. Henri Poincaré 29, 269–288, 1993.

M. Arnaudon. Dédoublement des variétés à bord et des semi-martingales. Stochastics and Stochastics Reports 44, 43–63, 1993.

M. Arnaudon. Propriétés asymptotiques des semi-martingales à valeurs dans les variétés à bord continu. *Séminaire de Probabilités XXVII, LNM 1557,* Springer, 1993.

M. Arnaudon. Differentiable and analytic families of continuous martingales in manifolds with connection. Probability Theory and Related Fields 108, 219–257, 1997.

M. Arnaudon and A. Thalmaier. Stability of stochastic differential equations in manifolds. *Séminaire de Probabilités XXXII, LNM 1686,* Springer, 1998.

M. Arnaudon and A. Thalmaier. Complete lifts of connections and stochastic Jacobi fields. J. Math. Pures et Appliquées 77, 283–315, 1998.

M. Arnaudon, X.-M. Li and A. Thalmaier. Manifold-valued martingales, change of probabilities, and smoothness of finely harmonic maps. Ann. Inst. H. Poincaré Probab. Statist. 35, 765–791, 1999.

M. Arnaudon and A. Thalmaier. Horizontal martingales in vector bundles. Séminaire de Probabilités XXXVI, LNM 1801, Springer 2003.

Y. Belopolskaya and Y. Dalecky. Stochastic Equations and Differential Geometry. Kluwer Academic Publisher, 1990.

J.-M. Bismut. Mécanique Aléatoire. *LNM 866,* Springer, 1981.

P. J. Catuogno. On stochastic parallel transport and prolongation of connections. Rev. Un. Mat. Argentina 41, 107–118, 1999.

P. J. Catuogno. Stochastic parallel transport and connections of H^2M. Arch. Math. (Brno) 35, 305–315, 1999.

P. J. Catuogno and Paulo C. Ruffino. Product of harmonic maps is harmonic: a stochastic approach. *Séminaire de Probabilités XL,* to appear.

R.W.R. Darling. Martingales on Manifolds and Geometric Itô Calculus. Ph. D. Thesis, University of Warwick, 1982.

R.W.R. Darling. Approximating Itô integrals of differential forms and geodesic deviation. Z. Wahrscheinlichkeitstheorie verw. Gebiete 65, 563–572, 1984.

T.E. Duncan. Stochastic integrals in Riemannian manifolds. J. Multivariate Anal. 6, 397–413, 1976.

T.E. Duncan. Some geometric methods for stochastic integration in manifolds. *Geometry and Identification 63–72, Lie Groups: Hist., Frontiers and Appl. Ser B, 1,* Math. Sci. Press, Brookline, 1983.

M. Émery. Stochastic Calculus in Manifolds. With an appendix by P. A. Meyer. *Universitext,* Springer, 1989.

M. Émery. On two transfer principles in stochastic differential geometry. *Séminaire de Probabilités XXIV, LNM 1426,* Springer, 1990. (Corrigendum in *Séminaire de Probabilités XXVI, LNM 1526,* Springer, 1992.)

M. Émery. Martingales continues dans les variétés différentiables. *Lectures on probability theory and statistics (Saint-Flour, 1998),* Lecture Notes in Math. 1738, Springer, 2000.

Y. Gliklikh. Global analysis in mathematical physics. Geometric and stochastic methods. Springer, 1997.

A. Grorud and M. Pontier. Calcul anticipatif d'ordre deux. Stochastics and Stochastics Reports 42, 1–23, 1993.

A. Grorud and M. Pontier. Calcul stochastique d'ordre deux et équation différentielle anticipative sur une variété. Japan J. Math. 21, 441–470, 1995.

A. Grorud and M. Pontier. Équation différentielle anticipative sur une variété et approximations. Mathematics and Computers in Simulation 38, 51–61, 1995.

P. A. Meyer. Géométrie stochastique sans larmes. *Séminaire de Probabilités XV, LNM 850,* Springer, 1981.

P. A. Meyer. A differential geometric formalism for the Itô calculus. *Stochastic Integrals, Proceedings of the L.M.S. Durham Symposium, LNM 851,* Springer, 1981.

P. A. Meyer. Géométrie différentielle stochastique (bis). *Séminaire de Probabilités XVI bis, LNM 921,* Springer, 1982.

P. A. Meyer. Géométrie différentielle stochastique. *Colloque en l'honneur de Laurent Schwartz, Volume 1, Astérisque 131,* Société Mathématique de France, 1985.

P. A. Meyer. Qu'est-ce qu'une différentielle d'ordre n ? Exposition. Math. 7, 249–264, 1989.

J.R. Norris. A complete differential formalism for stochastic calculus in manifolds. *Séminaire de Probabilités XXVI, LNM 1526,* Springer, 1992.

J. Picard. Martingales on Riemannian manifolds with prescribed limits. J. Funct. Anal. 99, 223–261, 1991.

J. Picard. Barycentres et martingales sur une variété. Ann. Inst. Henri Poincaré 30, 647–702, 1994.

L. Schwartz. Semi-martingales sur des variétés et martingales conformes sur des variétés analytiques complexes. *LNM 780,* Springer, 1980.

L. Schwartz. Géométrie différentielle du 2^{e} ordre, semimartingales et équations différentielles stochastiques sur une variété différentielle. *Séminaire de Probabilités XVI bis, LNM 921,* Springer, 1982.

L. Schwartz. Semimartingales and their Stochastic Calculus on Manifolds. Presses de l'Université de Montréal, 1984.

L. Schwartz. Construction directe d'une diffusion sur une variété. *Séminaire de Probabilités XIX, LNM 1123,* Springer, 1985.

L. Schwartz. Compléments sur les martingales conformes. Osaka J. Math. 23, 77–116, 1986.

L. Schwartz. Calcul infinitésimal stochastique. *Analyse mathématique et applications. Contributions en l'honneur de Jacques-Louis Lions.* Gauthier-Villars, 1988.

A. Thalmaier. Martingales on Riemannian manifolds and the nonlinear heat equation. *Stochastic Analysis and Applications. Proceedings of the Fifth Gregynog Symposium.* World Scientific, 1996.

A. Thalmaier. Brownian motion and the formation of singularities in the heat flow for harmonic maps. Probability Theory and Related Fields 105, 335–367, 1996.