Dr John Armstrong

Department of Mathematics

Faculty of Natural and Mathematical Sciences

2026.06.17

Dr John Armstrong

Department of Mathematics

FU Berlin26

Itô Stochastic Differentials

Itô Stochastic Differentials

- What is the differential calculus corresponding to Itô calculus?
- Motivation: Differentials in classical calculus
- Definitions and key results
- Application: coordinate free finance
- Application: SDEs on a manifold
- Joint work with Andrei Ionescu

History

Defining SDEs on manifolds

- Itô: "Stochastic Differential Equations in a Differentiable Manifold" (1950) - Lipschitz charts
- Ellworthy - approach from Stratonvich calculus
- Hsu - use Witney embedding theorem to consider manifolds in $\mathbb{R}^n$
- Laurent Schwarz/Michel Emery - theory of Schwarz morphisms, second order tangent bundle
- Others...

History

Defining SDEs on manifolds

- Itô: "Stochastic Differential Equations in a Differentiable Manifold" (1950) - Lipschitz charts
- Ellworthy - approach from Stratonvich calculus
- Hsu - use Witney embedding theorem to consider manifolds in $\mathbb{R}^n$
- Laurent Schwarz/Michel Emery - theory of Schwarz morphisms, second order tangent bundle
- Others...

Stochastic differentials

- Itô (1974): "Stochastic differentials" - theory of discrete differences, rather than a continuous differential
- Laurent Schwarz: there is nothing "ponctuel" in the object we call dX_t, believed a differential would be defined rigorously
- Michel Emery: the "existence of the [stochastic differential] is metaphysical and one is free not to believe in it"

Properties of the stochastic differential

We want to find a large space of processes S^{CBP} and operators $d^P(\cdot)_t$ acting on processes $X \in S^{\mathrm{CBP}}$.

- The operators must be local and forward looking.
- Large means that S^{CBP} contains all smooth diffusions

The following theorems must hold

Theorem

(The fundamental theorem of calculus)
Let $X \in S_t^{\mathrm{CBP}}$ and H be an adapted and sample-continuous process, then

$$d^p \left(\int_0^{\cdot} H_u dX_u \right)_t = H_t d^p (X)_t$$

Properties of the stochastic differential

We want to find a large space of processes S^{CBP} and operators $d^P(\cdot)_t$ acting on processes $X \in S^{\mathrm{CBP}}$.

- The operators must be local and forward looking.
- Large means that S^{CBP} contains all smooth diffusions

The following theorems must hold

Theorem

(The fundamental theorem of calculus)
Let $X \in S_t^{\mathrm{CBP}}$ and H be an adapted and sample-continuous process, then

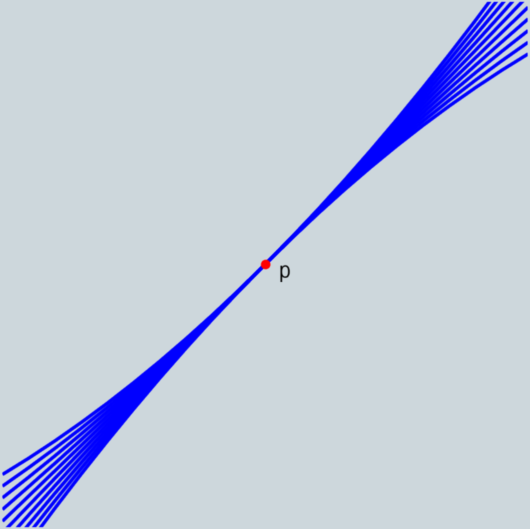
$$d^p \left(\int_0^{\cdot} H_u dX_u \right)_t = H_t d^p (X)_t$$

Theorem

(Uniqueness of solutions)
If $d^p(X)_t = 0$ for all $t \in [0, T]$, then $X_t = X_0$ for all $t \in [0, T]$

Tangent vectors

A tangent vector at a point p in a manfiold M is the equivalence class of smooth curves $\gamma : [-\epsilon, \epsilon] \to M$ with $\gamma(0) = p$ that are equal up to $o(t)$



Definition

Let $(X_t)_{t\geq 0}$ be a continuous time stochastic process and $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be an increasing function.

$X_h = o_p(f(h))$ if

$$\forall \epsilon > 0\, \delta > 0\, \exists \eta > 0 \text{ s.t. } h \in (0, \eta) \implies \mathbb{P}\left(\left|\frac{X_h}{f(h)} > \epsilon\right|\right) < \delta$$

$X_h = \mathcal{O}_p(f(h))$ if

$$\forall \epsilon > 0\, \exists K > 0, \eta > 0 \text{ s.t. } h \in (0, \eta) \implies \mathbb{P}\left(\left|\frac{X_h}{f(h)} > K\right|\right) < \epsilon$$

Chords Bounded in Probability

Notation: For an $\mathbb{R}^n$-valued continuous time stochastic process X_s,

$$X_{t,s} := X_s - X_t$$

$$X^*_{t,s} := \max_{1 \leq i \leq n} \sup_{u \in t,s} |X^i_{t,u}|$$

Chords Bounded in Probability

Notation: For an $\mathbb{R}^n$-valued continuous time stochastic process X_s,

$$X_{t,s} := X_s - X_t$$

$$X^*_{t,s} := \max_{1 \leq i \leq n} \sup_{u \in t,s} |X^i_{t,u}|$$

Definition

Let X be a continuous $\mathbb{R}^n$-valued semi-martingale with canonical decomposition

$$X = X_0 + A + M$$

into finite variation and local-martingale parts with

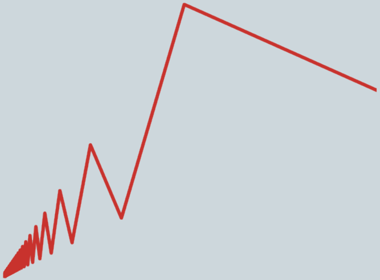
$$A = A^1 - A^2$$

being the canonical decomposition of A as a difference of non-decreasing processes. Then X has *chords bounded in probability*, $X \in S^{\mathrm{CBP}}_t(\mathbb{R}^n)$ if

$$A^1_{t,t+h} = \mathcal{O}_p(h), \quad A^2_{t,t+h} = \mathcal{O}_p(h) \quad \text{and} \quad (M^*_{t,t+h})^2 = \mathcal{O}_p(h)$$

Remarks

- "Chords Bounded in Probability" will be our notion of "differentiable" at t
- The curve below has chords bounded in probability but is not right differentiable at $t = 0$
- The use of convergence in probability rather than convergence in mean ensures locality. It also maximizes the size of S_t^{CBP}



Total variation and quadratic variation

Definition

Let $X = X_0 + A + M$ be a continuous semimartingale. The total variation is defined to equal

$$TV(A)_{t,t+h} := \underset{\|\pi\|\to 0}{\mathbb{P}\text{-lim}} \sum_{(u,v)\in\pi} \|A_{u,v}\|$$

the quadratic variation is defined to equal

$$[M]_{t,t+h} := \underset{\|\pi\|\to 0}{\mathbb{P}\text{-lim}} \sum_{(u,v)\in\pi} \|M_{u,v}\|^2$$

Total variation and quadratic variation

Definition

Let $X = X_0 + A + M$ be a continuous semimartingale. The total variation is defined to equal

$$TV(A)_{t,t+h} := \underset{\|\pi\|\to 0}{\mathbb{P}\text{-lim}} \sum_{(u,v)\in\pi} \|A_{u,v}\|$$

the quadratic variation is defined to equal

$$[M]_{t,t+h} := \underset{\|\pi\|\to 0}{\mathbb{P}\text{-lim}} \sum_{(u,v)\in\pi} \|M_{u,v}\|^2$$

Definition

If $X = X_0 + A + M$, we say X has *variations bounded in probability*, $X \in S^{\mathrm{VBP}}$, if

$$TV(A)_{t,t+h} = \mathcal{O}_p(h) \quad \text{and} \quad [M]_{t,t+h} = \mathcal{O}_p(h)$$

Bridging integration and differentiation

Total variation and quadratic variation are the limits of sums, so are similar to integrals. The definition of chords bounded in probability does not include sums.

Lemma

$S^{\mathrm{VBP}} = S^{\mathrm{CBP}}$

Proof:

$A \in S^{\mathrm{VBP}} \iff A \in S^{\mathrm{CBP}}$

follows from the decomposition

$$A_t = \frac{1}{2}(TV(A)_t + A_t) - \frac{1}{2}(TV(A)_t + A_t).$$

Proving

$$(M^*)^2_{t.t+h} = \mathcal{O}_p(h) \iff [M]_{t,t+h} = \mathcal{O}_p(h)$$

is a convergence-in-probability version of the Birkholder-Davis-Gundy inequality. It follows from the “Good λ estimates” used to prove that inequality.

Processes with differential zero

Having zero differential is defined to mean having "vanishing chords": replace the $\mathcal{O}_p$ in the definition of CBP with o_p.

Definition

Let $X = X_0 + A^1 + A^2 + M$ be a continuous semimartingale we say X has differential zero at time t if

$$A^1_{t,t+h} = o_p(h), \quad A^2_{t,t+h} = o_p(h), \quad \text{and} \quad (M^*_{t,t+h})^2 = o_p(h)$$

We write $d^p(X)_t = 0$

Lemma

$d^p(X)_t = 0$ *iff*

$$TV(A)_{t,t+h} = o_p(h) \quad \textit{and} \quad [M]_{t,t+h} = o_p(h)$$

Definition of the Itô stochastic differential

Definition

Let $X, Y \in S_t^{\mathrm{CBP}}(\mathbb{R}^n)$.
Write $X_t \sim Y_t \iff d^p(X - Y)_t = 0$

Lemma

The binary relation $\sim$ is an equivalence relation.

Definition

$d^p(X)_t$ is the equivalence class of X under $\sim$.

$$\mathbb{D}_t(\mathbb{R}^n) := S_t^{\mathrm{CBP}}(\mathbb{R}^n)/\sim$$

Definition

If X and Y in $S^{\mathrm{CBP}}(\mathbb{R}^n)$ and H is an $\mathcal{F}_t$-measurable random variable

$$d^p(X)_t + d^p(Y)_t := d^p(X+Y)_t \in \mathbb{D}(\mathbb{R}^n)$$

$$d^p(X)_t d^p(Y)_t := d^p[X,Y]_t \in \mathbb{D}_t(S^2(\mathbb{R}^n))$$

$$Hd^p(X)_t = d^p(HX)_t$$

If $H = Y_t$ we write $Y_t d^p(X)_t = d^p(Y_t X_s)_{s=t}$ to indicate that s should be viewed as varying and t as fixed.

Algebraic structures

Definition

If X and Y in $S^{\mathrm{CBP}}(\mathbb{R}^n)$ and H is an $\mathcal{F}_t$-measurable random variable

$$d^p(X)_t + d^p(Y)_t := d^p(X+Y)_t \in \mathbb{D}(\mathbb{R}^n)$$

$$d^p(X)_t d^p(Y)_t := d^p[X,Y]_t \in \mathbb{D}_t(S^2(\mathbb{R}^n))$$

$$H d^p(X)_t = d^p(HX)_t$$

If $H = Y_t$ we write $Y_t d^p(X)_t = d^p(Y_t X_s)_{s=t}$ to indicate that s should be viewed as varying and t as fixed.

Example: If W is a 1-dimensional Brownian motion

$d^p W_t d^p W_t = d^p(s)_{s=t} =: d^p t$

Key result 1

Theorem

If $d^p(X)_t = 0$ *for all* $t \in [0, T]$*, then* $X_t = X_0$ *for all* $t \in [0, T]$

Proof: Consider a process of the form $X = A^1$. Take a sequence of partitions of $[0, T]$, along a subsequence obtained by the diagonal argument, we have that the size of the chords tends to zero as the partitions become finer. Use the compactess of [0,T] to find a cover of intervals where all chords have gradient less than ϵ/T. Remove any interval which is contained fully in another. It follows that the value is less than 3ϵ on $[0, T]$.

Repeat the argument for $X = A^2$. For $X = M$ apply the same argument to the quadratic variation.

Key result 2

Theorem

(The Fundamental Theorem of Calculus)
Let $X \in S_t^{\mathrm{CBP}}$ and H be an adapted and sample-continuous process, then

$$d^p \left(\int_0^s H_u dX_u \right)_{s=t} = H_t d^p (X)_t$$

Proof: The Ito integral preserves the splitting, so just need to show

$$\int_t^{t+h} (H_u - H_t) dA_u = o_p(h), \quad \int_t^{t+h} (H_u - H_t) dM_u = o_p(h)$$

Use a classical argument to show the result for the bounded variation part. Mimic the argument with quadratic variation.

Equivalence of SDEs, chain rule

Corollary

The following are equivalent for a continuous semimartingale $X \in S^{\mathrm{CBP}[0,T]}$ and a continuous function $f(t, X_t)$

- *Y satisfies the Itô stochastic integral equation*

$$dY_t = f(t, X_t)dX_t$$

  with initial condition Y_0

- *Y satisfies the Itô stochastic differential equation*

$$d^p(Y)_t = f(t, X_t)d^p(X)_t$$

  with initial condition Y_0

Equivalence of SDEs, chain rule

Corollary

The following are equivalent for a continuous semimartingale $X \in S^{\mathrm{CBP}[0,T]}$ *and a continuous function* $f(t, X_t)$

- *Y satisfies the Itô stochastic integral equation*

$$dY_t = f(t, X_t) dX_t$$

  with initial condition Y_0

- *Y satisfies the Itô stochastic differential equation*

$$d^p(Y)_t = f(t, X_t) d^p(X)_t$$

  with initial condition Y_0

Corollary

If $f : \mathbb{R} \to \mathbb{R}$ *is* C^2 *and X is 1-dimensional, then*

$$d^p(f(X))_t = f'(X_t) d^p(X)_t + \frac{1}{2} f''(X_t) d^p[X, X]_t$$

Expectation and Covariance of differentials

Definition

Given $\eta \in \mathbb{D}_t(\mathbb{R}^n)$ choose a bounded representative X_t. The conditional expectation of η is

$$\mathbb{E}_t[\eta] := \lim_{h \to 0+} \frac{\mathbb{E}(X_{t+h} - X_t \mid \mathcal{F}_t)}{h}$$

if the limit exists. It is an $\mathcal{F}_t$-measurable random variable. Write $\mathbb{D}_t^1$ for the space of differentials with an expectation.

Definition

Given $\eta, \bar{\eta} \in \mathbb{D}_t(\mathbb{R}^n)$, their covariance, if it exists, is the symmetric $S^2(\mathbb{R}^n)$-valued random variable defined by

$$\mathrm{Cov}_t(\eta, \bar{\eta}) := \mathbb{E}_t(\eta\bar{\eta})$$

$$\mathrm{Var}_t(\eta) := \mathbb{E}_t(\eta^2)$$

Write $\mathbb{D}_t^2$ for the space of differentials with an expectation and a variance.

Change of measure formula

Theorem

Let $\mathbb{Q}$ and $\mathbb{P}$ be equivalent measures with Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$. Define

$$q_t = \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \mathcal{F}_t\right).$$

Suppose that q_t has chords bounded in probability. Then

$$d^{\mathbb{P}} X_t = d^{\mathbb{P}} Y_t \iff d^{\mathbb{Q}} X_t = d^{\mathbb{Q}} Y_t$$

Define

$$\mathcal{L}(q)_t := \log q_t + \int_0^t \frac{1}{2q_s^2} d[q,q]_s.$$

If $\eta, \bar{\eta} \in \mathbb{D}_t^2$ then

$$\mathbb{E}_t^{\mathbb{Q}}(\eta) = \mathbb{E}_t^{\mathbb{P}}(\eta) + \mathrm{Cov}_t^{\mathbb{P}}(d\mathcal{L}(q)_t, \eta)$$

$$\mathrm{Cov}_t^{\mathbb{Q}}(\eta, \bar{\eta}) = \mathrm{Cov}_t^{\mathbb{P}}(\eta, \bar{\eta})$$

A financial application

A complete continuous time market is given by the following data

- Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space
- Random variables represent the payoff of financial contracts at time T.
- The price of a contract $f \in L^0$ at time t is given by $\mathbb{E}_{\mathbb{Q}}(f \mid \mathcal{F}_t)$ for some equivalent measure $\mathbb{Q}$. Contracts without a price or with price $+\infty$ cannot be purchased.

A financial application

A complete continuous time market is given by the following data

- Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$ be a filtered probability space
- Random variables represent the payoff of financial contracts at time T.
- The price of a contract $f \in L^0$ at time t is given by $\mathbb{E}_{\mathbb{Q}}(f \mid \mathcal{F}_t)$ for some equivalent measure $\mathbb{Q}$. Contracts without a price or with price $+\infty$ cannot be purchased.

Comments

- A financial derivative is a bet on the behaviour of some other financial product, called the underlying.
- In our setup we do not distinguish between derivatives and underlyings, they are all on an equal footing
- Determinisitc random variables correspond to guaranteed cash: we're assuming interest rates are zero for convenience.
- It is a standard result in financial mathematics that prices are given by an expectation in some measure $\mathbb{Q}$ otherwise there will be opportunities for guaranteed risk-free profits.
- Because you expect to receive better returns from risky assets $\mathbb{P} \neq \mathbb{Q}$ in general: what you pay for a risky asset is different from your expected profit..

Two fund theorem

Definition

An *instantaeous portfolio* is a pair (c, η) with $c \in \mathbb{R}$, its cost, and $\eta \in \mathbb{D}^2(\mathbb{R})$ representing its variability which satisfies the pricing condition $\mathbb{E}_t^{\mathbb{Q}}(\eta) = 0$.

Theorem

Any instantaneous portfolio may be written as

$$(c, \eta) = (c, \alpha\mu + \eta^{\perp})$$

where

$$\mu := (d\mathcal{L}(q)_t - d[\mathcal{L}(q), \mathcal{L}(q)]_t)$$

α *is an* $\mathcal{F}_t$*-measurable random variable,* $\mathbb{E}_t^{\mathbb{P}}(\eta^{\perp}) = 0$*,* $\mathrm{Cov}_t^{\mathbb{P}}(\eta^{\perp}, \mu) = 0$*. Hence*

$$\mathrm{Var}_t^{\mathbb{P}}(\eta) = \alpha^2 \mathrm{Var}^{\mathbb{P}}(\mu) + \mathrm{Var}^{\mathbb{P}}(\eta^{\perp}).$$

In particular (c, η) *minimizes risk for a given cost and return if and only if* $\eta^{\perp} = 0$*.*

Two fund theorem

Interpretation: Any efficient instantaneous portfolio is given by a linear combination of the risk-free asset $(1, 0)$ and the *market portfolio* $(0, \mu)$. This is a version of the classical mutual fund theorem from Markowitz's theory, but that theory is a 1-period model which assumes normally distributed assets. This is an infinitessimal result which doesn't assume any dynamics.

Two fund theorem

Interpretation: Any efficient instantaneous portfolio is given by a linear combination of the risk-free asset $(1, 0)$ and the *market portfolio* $(0, \mu)$. This is a version of the classical mutual fund theorem from Markowitz's theory, but that theory is a 1-period model which assumes normally distributed assets. This is an infinitessimal result which doesn't assume any dynamics.

Proof: Treat Cov as analogous to an inner product. The result follows by applying one step of the Gram–Schmidt process

If $\mu = 0$, there is nothing to prove. Otherwise define

$$\eta^{\perp} = \eta - \frac{\mathrm{Cov}(\eta, \mu)}{\mathrm{Var}(\mu, \mu)}\mu.$$

The change of measure formula, implies $\mathbb{E}_t^{\mathbb{P}}(\eta^{\perp}) = \mathbb{E}_t^{\mathbb{Q}}(\eta^{\perp}) = 0$.

Stochastic differentials on manifolds

A tangent vector on a manifold is an equivalence class of curves through a point: you cannot compare tangent vectors at different points.

Definition

Let X be a process taking values in an n-dimensional manifold M. The *stochastic tangent* TX_t at time t is the equivalence class of processes X, Y on M satisfying $X \sim Y$ if

$$X_t = Y_t, \qquad d(\phi(X))_t = d(\phi(Y))_t$$

for all smooth maps $\phi : M \to \mathbb{R}^n$.
$\mathbb{D}_t(M)$ is the space of stochastic tangent vectors.

Remark: The chain rule ensures this is equivalent to our old definition on $\mathbb{R}^n$ apart from the addition of a base point - i.e. adding “for all ϕ” makes no difference.

Stochastic differentials on manifolds

A tangent vector on a manifold is an equivalence class of curves through a point: you cannot compare tangent vectors at different points.

Definition

Let X be a process taking values in an n-dimensional manifold M. The *stochastic tangent* TX_t at time t is the equivalence class of processes X, Y on M satisfying $X \sim Y$ if

$$X_t = Y_t, \qquad d(\phi(X))_t = d(\phi(Y))_t$$

for all smooth maps $\phi : M \to \mathbb{R}^n$.
$\mathbb{D}_t(M)$ is the space of stochastic tangent vectors.

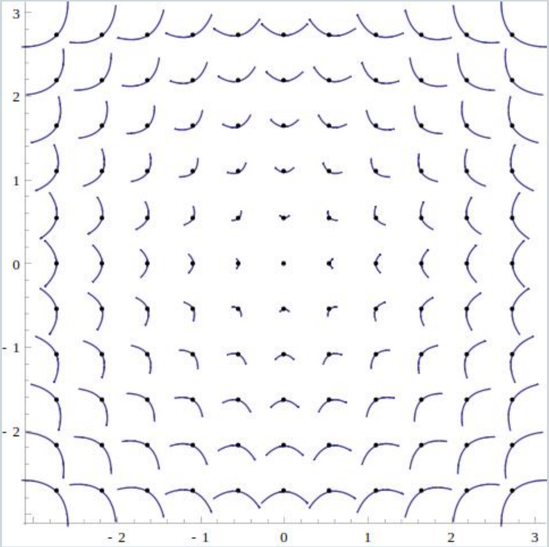
Remark: The chain rule ensures this is equivalent to our old definition on $\mathbb{R}^n$ apart from the addition of a base point - i.e. adding "for all ϕ" makes no difference.

Definition

Let $f : M \to N$ be a smooth map between manfiolds. Define the push forward $f^* : \mathbb{D}_t(M) \to \mathbb{D}_t(N)$ by $f^*(TX_t) = T(f \circ X)_t$.

Stochastic Differential Equations on Manifolds

Let $\gamma : \mathbb{R}^d \times M \to M$ be smooth and satisfy $\gamma(0, p) = p$. The figure illustrates the case when $d = 1$.

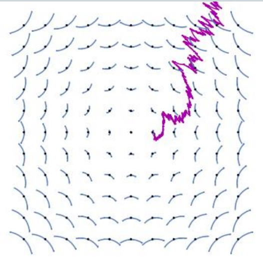


Stochastic Differential Equations on Manifolds

Let $\gamma : \mathbb{R}^d \times M \to M$ be smooth and satisfy $\gamma(0, p) = p$.

Definition

Let W be an $\mathbb{R}^d$-valued process. X solves the SDE (γ, W) if
$TX_t = T\gamma(W_s - W_t, X_t, X_t)_{s=t}$.

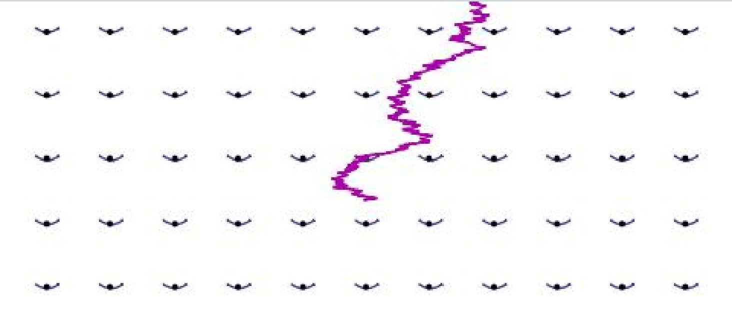


Ito's lemma on manifolds

Classically Ito's lemma relates the coefficients of SDEs generating processes that are related by a smooth map $\phi : M \to N$. It now has the following elegant restatement.

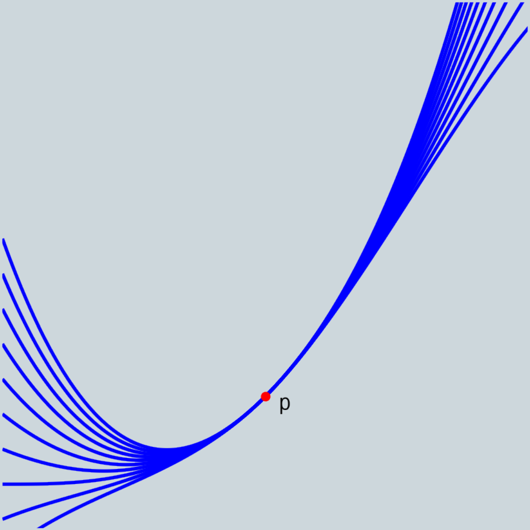
Theorem

ϕ maps the solutions of (γ, W) to the solutions of $(\phi \circ \gamma, W)$.



2-jets

(γ, W) has the same solutons as $\gamma', W)$ if $\gamma(d,p) = \gamma'(h,p) + o_p(|h|^2)$. $j_2(\gamma)$ is the equivalence class.

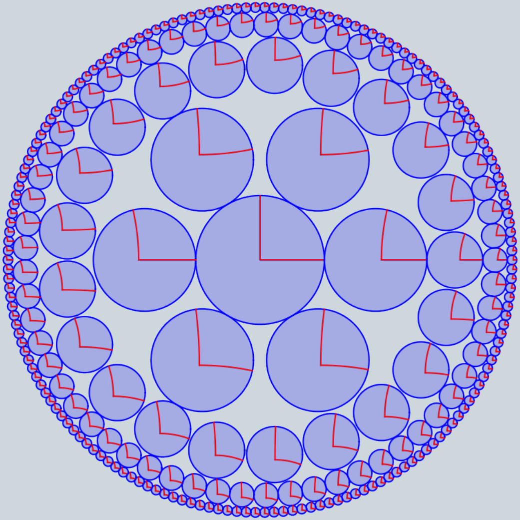


Jets and differentials

- The coefficients of an SDE on M driven by Euclidean Brownian motion are given by 2-jets $j_2(\gamma)$ of maps from $\mathbb{R}^d$ to M.
- Curvature of $j_2(\gamma)$ causes drift.
- Differentials and jets transform by function composition.
- Jets and differentials have dual roles: jets are coefficients, differentials are drivers.

Brownian motion on hyperbolic space

Let $\gamma : \mathbb{R}^n \times M \to M$ be a smoothly varying choice of geodesic coordinates at each point $p \in M$. (γ, W) is Brownian motion



Brownian motion on a closed surface

The problem is that such a γ only exists if M is *parallelizable*: i.e. if you can smoothly choose linearly independent vectors at every point.

By the hairy ball theorem, the sphere is not parallelizable. Brownian motion on the sphere cannot be written as (γ, W) where W is Euclidean Brownian motion. It is not the solution of a smooth classical Itô SDE.

There is no canoncal mapping between a Euclidean Brownian motion and a Brownian motion on the sphere.

Brownian motion

Brownian motion on (M, g) is any process which satsifies $\mathbb{E}_t^g(TX_t) = 0$ and $\mathrm{Var}_t(TX_t) = g$.

Summary

- Ito stochastic differentials are real
- Stochastic differential equations can be written in integral or differential form
- Stochasic ifferentials have the operations you expect: addition, scalar multiplication and a symmetic product.
- There is pushforward operation for any smooth functon. In local coordinates it is given by the chain rule.
- There is change of measure formula.
- Some ideas which are hard to articulate using integrals can be expressed using differentials.