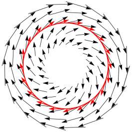
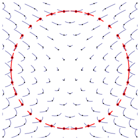
Intrinsic stochastic differential equations as jets



Theory and applications



Institut de Recherche Mathematique Avancees, Strasbourg, March 17, 2017

Scuola Normale Superiore, Pisa, October 10, 2016

2016 Risk & Stochastics Conference, LSE, London, April 21, 2016.

Damiano Brigo
Dept. of Mathematics, Imperial College, London
Joint work with *John Armstrong*
Dept. of Mathematics, King's College, London

—

http://arxiv.org/abs/1602.03931 **http://arxiv.org/abs/1610.03887**

I would like to dedicate this talk to
Giovanni Battista Di Masi (1944-2016),
who passed away on April 4.

PhD at Brown University, Author in signal processing, stochastic control & filtering, probability, stochastic analysis, statistics.

Professor of Probability & Mathematical Statistics, later Head of the Department of Mathematics at the University of Padua and Assessor at the Padua Local Administration

Gianni was my Laurea dissertation supervisor (1990) and he was present at my PhD viva in Amsterdam. He taught me stochastic calculus, nonlinear filtering, and much more beyond mere science.

Agenda I

1. The traditional view of SDEs: Ito and Stratonovich
   - SDEs and stochastic integrals
   - Probability and geometry
2. Itô SDEs on manifolds: 2-Jets
   - Drawing and simulating SDEs as “fields of curves”
   - Coordinate-free converging difference scheme as SDE
   - Coordinate free Itô SDE as 2-jet scheme limit
   - Coordinate-free Itô formula and stochastic analysis
   - Generalizations and other results
3. Applications: Optimal approximation of SDEs on submanifolds
   - Stratonovich projection
   - Ito-vector projection
   - Ito jet projection
4. Conclusions and References
5. Bonus material

SDEs: Brownian Motion as the randomness driver

$$\underbrace{dX_t}_{\substack{\text{Change in } X \\ \text{between } t \text{ and } t+dt}} = \underbrace{a(X_t)}_{\text{"MEAN } \textit{CHANGE}\text{"}} \; dt + \underbrace{b(X_t)}_{\substack{\text{Amplitude of} \\ \text{random shock}}} \underbrace{dW_t}_{\substack{\text{Random} \\ \text{shock}}}$$

W is Brownian motion or Wiener process

Independent stationary increments, $W_{t+\Delta_1 t} - W_t$ indep of $W_t - W_{t-\Delta_2 t}$, continous paths, $W_0 = 0$. This implies Gaussian $\Delta W_t \sim \mathcal{N}(0, \Delta t)$.

These properties can coexist but W's paths have unbounded variation - rough paths - nowhere differentiable. *So what does dW really mean?*

Quadratic variation (nested dyadic grids) $0 = t_0^n < t_1^n < \ldots < t_n^n = T$,

$$\lim_n \sum_{i=0}^{n-1} (W_{t_{i+1}^n} - W_{t_i^n})^2 = T, \quad or \quad \text{"}dW_t dW_t = dt'' \quad (\text{"}dt\, dW_t = 0, \;\; dt\, dt = 0'')$$

Classic theory of Stochastic Differential Equations

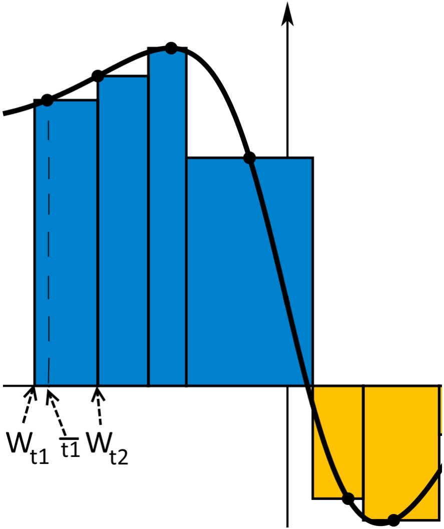
$dX_t = a(X_t)dt + b(X_t)dW_t, \quad X_0.$ dW not a real differential. So?

- Write it as

$$X_t = X_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW_s.$$

- Now the matter is defining the stochastic integral driven by dW
- Since W has unbounded variation, we cannot define this as an ordinary Stiltjes integral on the paths.

The stochastic integral as a Stiltjes integral?



In a Stiltjes integral one has $\int_0^T b(X_s)dW_s =$

$$= \lim_n \sum_{i=1}^{n} b(X(\bar{t}_i))(W_{t_{i+1}} - W_{t_i})$$

for ANY choice $\bar{t}_i \in [t_i, t_{i+1})$.

However, for Brownian motion this does not work since W has unbounded variation.

Add an extra specification: *we need to explicitly decide which point $\bar{t}_i$ is considered.*

- $X_t = X_0 + \int_0^t a(X_s)ds + \boxed{\int_0^t b(X_s)dW_s}$??
- Traditionally, 2 main definitions of stochastic integrals with $L^2(\mathbb{P})$-convergence: Initial point $\bar{t}_i = t_i$ vs mid point $\bar{t}_i = \frac{t_i + t_{i+1}}{2}$

$$\int_0^T b(X_s)dW_s = \lim_n \sum_{i=1}^n b(X(t_i))(W_{t_{i+1}} - W_{t_i}) \quad \text{(Itô)}$$

$$\int_0^T b(X_s)\circ dW_s = \lim_n \sum_{i=1}^n b\left(X\left(\frac{t_i + t_{i+1}}{2}\right)\right)(W_{t_{i+1}} - W_{t_i})\text{(Stratonovich)}$$

(Str more general def. has $[b(X(t_i)) + b(X(t_{i+1}))]/2$ in front of dW) where it is understood that as n tends to infinity the mesh size of the partition $\{[0, t_1), [t_1, t_2), \ldots, [t_{n-1}, t_n = T]\}$ of $[0, T]$ tends to 0.

- Stratonovich integral looks into the future, Ito does not.

Battle of the integrals: Ito or Stratonovich?

Itô (-Doeblin) integral:

- **The good:** “Does not look into the future” (social sciences)
- Itô integral is martingale: split in “local mean” & “volatility” above
- Good probabilistically, many important results in probability theory.
- **The bad:** due to $dWdW = dt \neq 0$ does not satisfy chain rule!

$$dX_t = a(X_t)dt + b(X_t)dW_t,$$

$$\text{Itô's formula: } df(X_t) = ((\nabla f)(X_t))^T\, dX_t + \frac{1}{2}(dX_t)^T(Hf(X_t))(dX_t)$$

- What does it mean as a change of coordinates/variables?
- **The ugly:** Given finite variation noises $W^n \to W$ a.s. uniformly in t-bounded intervals, solutions in dW^n do not converge to Itô SDE sol. Bad for engineering / physical systems with external noise

Battle of the integrals: Ito or Stratonovich?

Fisk ([12])-Stratonovich ([29]) (-McShane [23]) Integral.

- **The good:** satisfies chain rule (same as ordinary differential eq./vector fields), good for basic geometry

$$dX_t = a(X_t)dt + b(X_t) \circ dW_t, \quad df(X_t) = ((\nabla f)(X_t))^T \circ dX_t$$

- E.g. the above SDE dX stays in a manifold M if $a(X)$ & $b(X)$ are in the tangent space of the manifold. If not project on tangent space and you have approximated original SDE with SDE on M.
- Now if $W^n \to W$, the solution uder W^n converges to the Stratonovich SDE solution (Wong Zakai)
- **The bad**: Looks into the future.
- **The ugly**: Cannot interpret SDE dt term as local mean (no martingale property but... median?). Not good probabilistically.

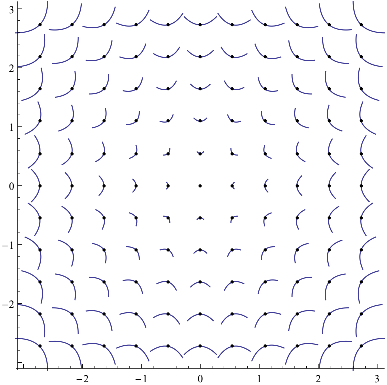
- **In a nutshell:** Itô ok probabilistically, Stratonovich geometrically.
- **This talk**: Let's make Ito SDEs good for **geometry** too.
- Natural question. **Ito-Stratonovich Transformation**: given Ito SDE, by suitably changing $a(X_t)$ one obtains a Strat. SDE with the same solution X. Why not use that back & forth?
- Because e.g. optimality of projection on submanifolds for dimension reduction depends on choice of calculus (later)
- **History**: Itô integral in the 40's-50's. Itô dominates among mathematicians, except for geometry. Stratonovich fared better with physicists & engineers, due to Wong Zakai & symmetry.
- Difficult infancy for symmetric integral. Donald Fisk paper rejected by Annals in mid 60's. In 1967 Skorokhod (1930-2011) [28] reviewed Stratonovich's 1966 book quite critically (euphemism).
- We now introduce **Itô calculus on manifolds** using jets. Previous approaches: Schwartz morphism, see Emery [11], & Itô bundle, see Gliklikh [15].

Jets and SDEs

For all $x \in \mathbb{R}^n$ consider smooth curve $\gamma_x : \mathbb{R} \to \mathbb{R}^n, \quad \gamma_x(0) = x$

Example: γ_x^E on $\mathbb{R}^2$ as follows (zero 3d-on derivatives):

$$\gamma^E_{(x_1,x_2)}(t) = (x_1, x_2) + \underbrace{t(-x_2, x_1)}_{\textit{circular counterclock}} + \underbrace{3t^2(x_1, x_2)}_{\textit{radially outward}}$$

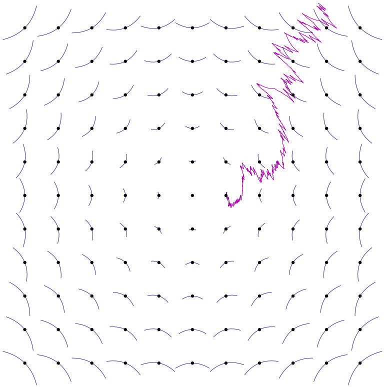


Jets and SDEs

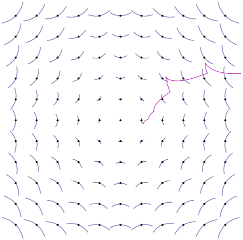
Given such a γ, a starting X_0 ($X_0 = (1,0)$ in our example), a W_t & time step δt define discrete time stochastic process:

$$X_0 := x_0, \quad X_{t+\delta t} := \gamma^E_{X_t}(W_{t+\delta t} - W_t)$$

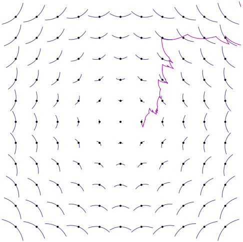
We have connected the points using the curves in $\gamma^E_{X_t}$: follow $s \mapsto \gamma_{X_t}(s)$ from $s = 0$ to $s = \mathcal{N}(0, \delta t)$, all $\mathcal{N}$ indepedent



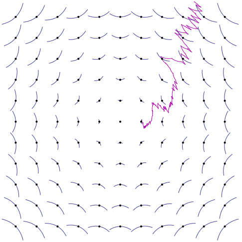
Jets and SDEs



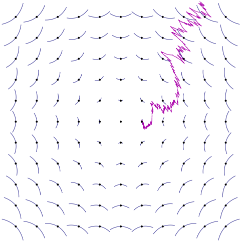
$\delta t = 0.2 \times 2^{-5}$



$\delta t = 0.2 \times 2^{-7}$



$\delta t = 0.2 \times 2^{-9}$



$\delta t = 0.2 \times 2^{-11}$

Jets and SDEs

$$X_{t+\delta t} := \gamma_{X_t}(\delta W_t) \quad \text{reads:}$$

"Follow the curve γ starting from X for a parameter increment of $\delta W_t := W_{t+\delta t} - W_t$". In this description,

Not using the $\mathbb{R}^n$ vector space structure. Intrinsic.

These discrete time stochastic processes converge in some sense to a limit as the time step tends to zero for γ such as γ^E with sufficiently good regularity. Write the limit equation as

$$\text{Coordinate free SDE:} \quad X_t \curvearrowright \gamma_{X_t}(\mathrm{d}W_t), \quad X_0 = x_0. \tag{1}$$

How can the scheme limit be made precise and how does it relate to classic stochastic calculus?

Jets and SDEs

In a coordinate system, consider the Taylor expansion of γ_x.

$$\gamma_x(t) = x + \gamma_x'(0)t + \frac{1}{2}\gamma_x''(0)t^2 + R_x t^3, \quad R_x = \frac{1}{6}\gamma_x'''(\xi), \quad \xi \in [0, t],$$

where $R_x t^3$ is the remainder term in Lagrange form. Substituting this Taylor expansion in our scheme $X_{t+\delta t} = \gamma_{X_t}(W_{t+\delta t} - W_t)$ we obtain

$$\delta X_t = \gamma_{X_t}'(0)\delta W_t + \frac{1}{2}\gamma_{X_t}''(0)(\delta W_t)^2 + R_{X_t}(\delta W_t)^3, \quad X_0 = x_0. \tag{2}$$

Properties of Brownian motion such as "$(dW)^2 = dt$" and "$(dW)^3 = 0$" suggest we replace $(\delta W_t)^2$ with δt and $(\delta W_t)^3$ with 0. We obtain:

$$\delta \bar{X}_t = \underbrace{\gamma_{\bar{X}_t}'(0)}_{=:b(\bar{X}_t)}\delta W_t + \underbrace{\frac{1}{2}\gamma_{\bar{X}_t}''(0)}_{=:a(\bar{X}_t)}\delta t, \quad \bar{X}_0 = x_0.$$

Jets and SDEs

$$\delta \bar{X}_t = a(\bar{X}_t)\delta t + b(\bar{X}_t)\delta W_t. \tag{3}$$

This is the Euler scheme & under suitable assumptions converges in $L^2(\mathbb{P})$ to the solution to the Itô stochastic differential equation:

$$\mathrm{d}\tilde{X}_t = a(\tilde{X}_t)\,\mathrm{d}t + b(\tilde{X}_t)\mathrm{d}W_t, \quad \tilde{X}_0 = x_0. \tag{4}$$

More precisely, assume in the given coordinate system $\gamma_x(t)$ is smoothly varying in x with first & second t derivatives at 0 satisfying Lipschitz conditions in x. Assume that the third t-derivative at $t = 0$ is uniformly bounded in x. **Theorem: (Armstrong & B. 2016).** The following 3 schemes have as same $L^2(\mathbb{P})$ limit the classic Ito SDE $\tilde{X}$.

- Coordinate free γ_x scheme: $X_{t+\delta t} := \gamma_{X_t}(W_{t+\delta t} - W_t), \quad X_0$
- 2-jet scheme: $\delta \hat{X}_t = \gamma'_{\hat{X}_t}(0)\delta W_t + \frac{1}{2}\gamma''_{\hat{X}_t}(0)(\delta W_t)^2, \quad X_0$
- The classic Euler scheme: $\delta \bar{X}_t = \gamma'_{\bar{X}_t}(0)\delta W_t + \frac{1}{2}\gamma''_{\bar{X}_t}(0)\delta t, \quad X_0$

Jets and SDEs

Are the 2-jet scheme and its limit coordinate free?

$$\delta\hat{X}_t = b(\hat{X}_t)\delta W_t + a(\hat{X}_t)(\delta W_t)^2,\ x_0 \quad \to \quad \mathrm{d}\tilde{X}_t = \underbrace{a(\tilde{X}_t)}_{\frac{1}{2}\gamma''_{\tilde{X}_t}(0)}\,\mathrm{d}t + \underbrace{b(\tilde{X}_t)}_{\gamma'_{\tilde{X}_t}(0)}\mathrm{d}W_t, x_0$$

Coefficients a & b of Itô SDE only depend on first two derivatives of γ.

Curves $\gamma_1 \sim \gamma_2$ have the same ***k*-jet** if their Taylor expansions are equal up to order t^k in one (all) coordinate system. k-jet can be defined as equivalence class $j_2(\gamma_1) := \tilde{\gamma}_1$.

Given our convergence results, showing that the limit of our scheme depends only on the two-jet, we may rewrite $X_t \rightsquigarrow \gamma_{X_t}(\mathrm{d}W_t),\ X_0$ as:

$$\text{Coordinate-free 2-jet SDE:} \qquad X_t \rightsquigarrow j_2(\gamma_{X_t})(\mathrm{d}W_t), \quad X_0 = x_0. \quad (5)$$

Itô's formula via 2-jets

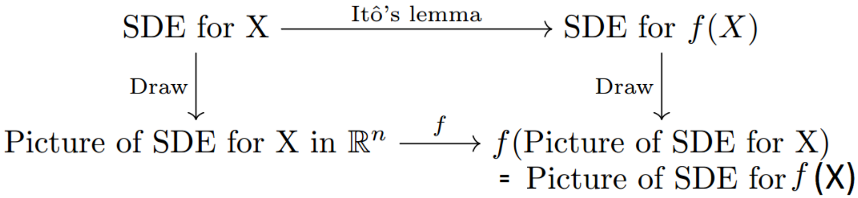
Lemma (Itô's lemma — coordinate free formulation)

If the process X_t satisfies $X_t \rightsquigarrow j_2(\gamma_{X_t})(\mathrm{d}W_t)$
then $f(X_t)$ satisfies $f(X)_t \rightsquigarrow j_2(f \circ \gamma_{X_t})(\mathrm{d}W_t)$.

Itô's formula: the transformation rule for jets under a change of coordinates is the composition of functions.

We have illustrated a way of drawing an SDE on a rubber sheet such that if sheet is stretched, diagram transforms as per Itô's lemma.

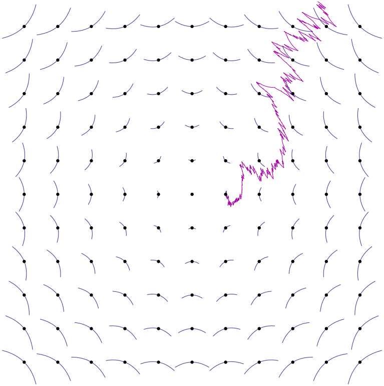
Or: the following diagram commutes



Itô's formula via 2-jets

Since we now understand the geometric content of Itô's lemma, we can draw a picture to illustrate it. Consider the transformation

$$(\theta, s) = \phi(x_1, x_2) = \left(\arctan(x_2/x_1), \log(\sqrt{x_1^2 + x_2^2})\right) \quad (\phi(z) = i \log(z))$$

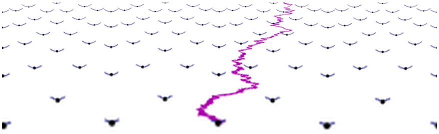


applied to our γ^E (left) process.

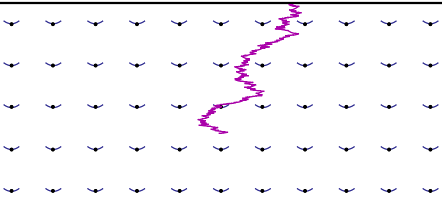
1st apply ϕ to each point (stretch the rubber sheet).

$$\mathrm{d}(\theta, s) = \left(0, \frac{7}{2}\right) \mathrm{d}t + (1, 0)\ \mathrm{d}W_t.$$

Itô's formula via 2-jets



The process $j_2(\phi \circ \gamma^E)$ plotted using image manipulation software



The process $j_2(\phi \circ \gamma^E)$ plotted by applying Itô's lemma

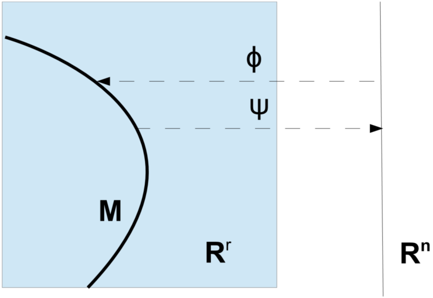
Figure: Two plots of the process $j_2(\phi \circ \gamma^E)$ in the plane (θ, s).

Generalizations and other results

- Can generalize to SDE driven by vector-Brownian motion using jets driven by $\mathbb{R}^m$ parameters.
- In this case only part of the jet information is used for SDE; weak and strong equivalence of SDEs.
- Jet-based definition of backward and fwd diffusion operators
- Fan diagrams and Stratonovich drift $a(X)$ as median.
- Itô - Stratonovich transformation interpreted geometrically as follows: a 2-jet (Itô) can be equivalently represented by subsequent application of two vector flows (Stratonovich) and vice-versa.

We now apply jets to optimal approximation of SDEs on submanifolds.

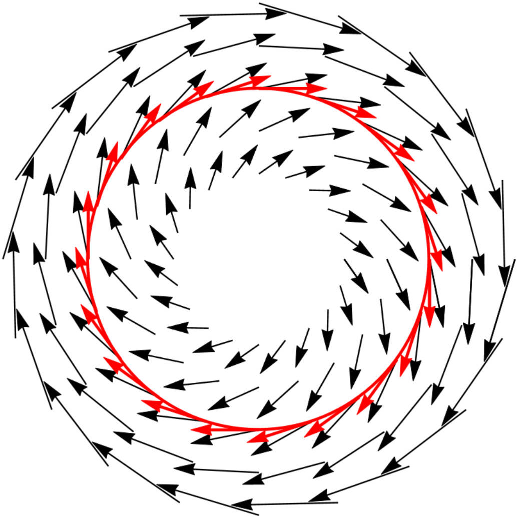
Optimal approximation of SDEs on submanifolds

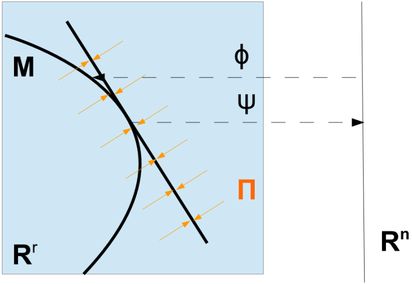


Given SDE X on $\mathbb{R}^r$, with $M \subset \mathbb{R}^r$ an n-dimensional manifold of $\mathbb{R}^r$, and $X_0 \in M$, we wish to find a SDE $\phi(Y)$ in M starting at X_0 whose solution approximates X. Clearly $r > n$.

$$
\begin{aligned}
\text{Approximate} \quad & dX = a(X)dt + b_\alpha(X)dW^\alpha && \text{in } \mathbb{R}^r, \\
\text{with } \phi(Y) \in M, \text{ where} \quad & dY = A(Y)dt + B_\alpha(Y)dW^\alpha && \text{in } \mathbb{R}^n \\
X_0 = \phi(Y_0) \in M, \quad & n - \text{dimensional manifold of} && \mathbb{R}^r
\end{aligned}
$$

Stratonovich projection via tangent space projection Π





Stratonovich projection. Write the Ito SDE for X in Stratonovich form

$$dX = \bar{a}\,dt + b_\alpha \circ dW^\alpha \ \text{ in } \mathbb{R}^r,$$

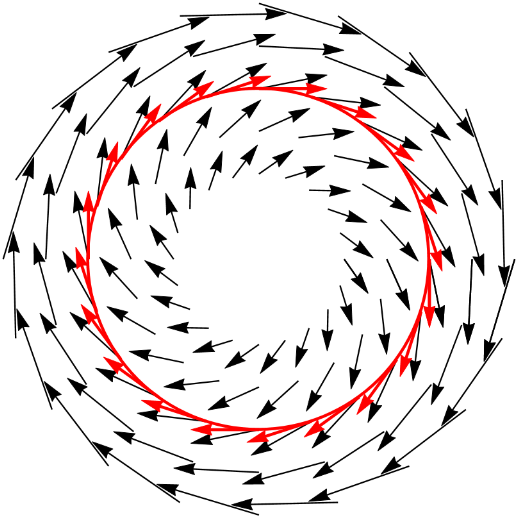
$X_0 \in M$. Apply the tangent space projection to obtain M-SDE $Z = \phi(Y), Z_0 = X_0$,

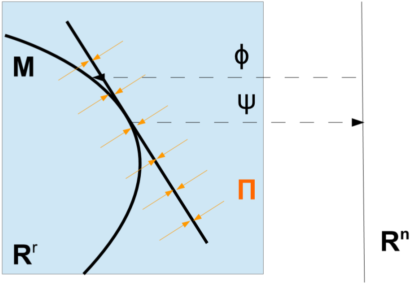
$$dZ = \Pi_Z[\bar{a}]\,dt + \Pi_Z[b_\alpha] \circ dW^\alpha$$

Justification: for $b = 0$ it coincides with optimal ODE projection minimizing leading term of Taylor expansion for $|\phi(Y) - X|^2$.

No optimality (yet) for SDE as a whole: rough paths & $\bar{a}, b$ together. We are investigating potential a.s. optimality (as opposed to mean square optimality of the following projections below)

Ito-vector projection via tangent space projection Π





Itô vector projection. Minimize leading (t-term) coeff. Ito-Taylor expansion of $\mathbb{E}[|{\color{red}X_t} - \phi(Y_t)|^2]$, to get B, **but that coefficient does not vanish**, so the error stays order t. Fixing that B in a neighborhood, get A by minimizing the next leading term coeff (t^2-term). This results also in A minimizing (regardless of B), up to order t, $|\mathbb{E}[{\color{red}X_t} - \phi(Y_t)]|^2$.

$$B_\alpha(Y_t, t) = (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} b_\alpha(\cdot, t) \quad \text{(same as Strat proj.)}$$

$$A(Y_t, t) = (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} \left(a(\cdot, t) - \frac{1}{2} (\nabla_{B_\alpha(Y_t,t)} \phi_*) B_\beta(Y_t, t) g_E^{\alpha\beta} \right).$$

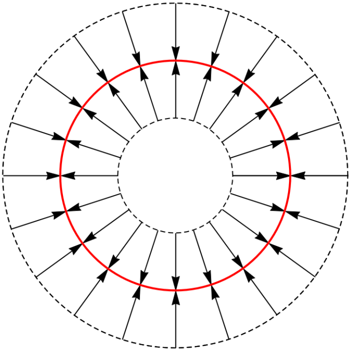
Itô vector projection

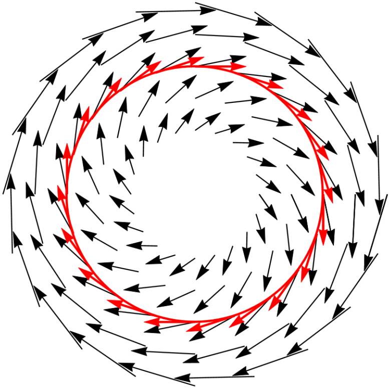
- $g_E^{\alpha,\beta} = 1_{\{\alpha=\beta\}}$, or more generally the symmetric 2-form defining the Euclidean metric on $\mathbb{R}^m$ in the non-orthonormal case.
- In Euclidean space notation (with orthonormal coordinates)

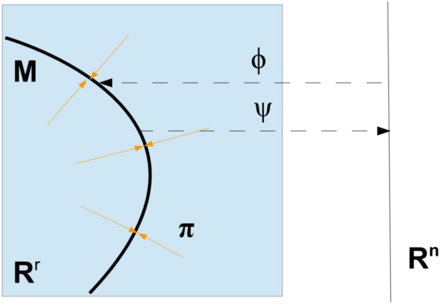
$$(\nabla_{B_\alpha} f_{i,*}) B_\beta g_E^{\alpha,\beta} = \text{Trace}[B^T (H f_i) B]$$

- Drawback: after minimizing the order 1 coefficient of the error $\mathbb{E}[|X_t - \phi(Y_t)|^2]$ in t to get B, we minimize the order 2 coefficient to get A *but without the order 1 coefficient vanishing*. This means we never really get to order 2.
- We also get, as a bonus, A minimizes the weak error $|\mathbb{E}[X_t - \phi(Y_t)]|^2$ up to order t.
- Given these drawbacks, can we find another projection that, differently from the Itô vector projection, is *consistently optimal* up to order t^2?

Metric projection π vs tangent space projection Π







π is the metric projection $\pi : \mathbb{R}^r \to M$, defined on a tubular neighborhood of M, of which the earlier linear projection Π is the first order component.

Set $\tilde{\pi} = \psi \circ \pi$.

Itô jet projection. Make t coefficient *vanish* and minimize leading t^2 coeff. of Ito-Taylor expansion for the error

$$\mathbb{E}[d_M(\pi(X_t), \phi(Y_t))^2] \text{ or } \mathbb{E}[|\pi(X_t) - \phi(Y_t)|_r^2] \quad \text{for small } t$$

B as before (same as Strat & Ito vector projections) and A

$$A(Y_t, t) = \tilde{\pi}_*(a(\phi(Y_t), t)) + \frac{1}{2}(\nabla_{b_{\alpha(\phi(Y_t),t)}} \tilde{\pi}_*) b_\beta(\phi(Y_t), t) g_E^{\alpha\beta}.$$

Why "jet" projection?

We have called this latest projection "jet projection" because one can check that if the SDE for $X \in \mathbb{R}^r$, $X_0 \in M$, is written as

$$X_t \smile j_2(\gamma_{X_t}(dW_t))$$

then the Ito-jet projection describes in coordinates the SDE

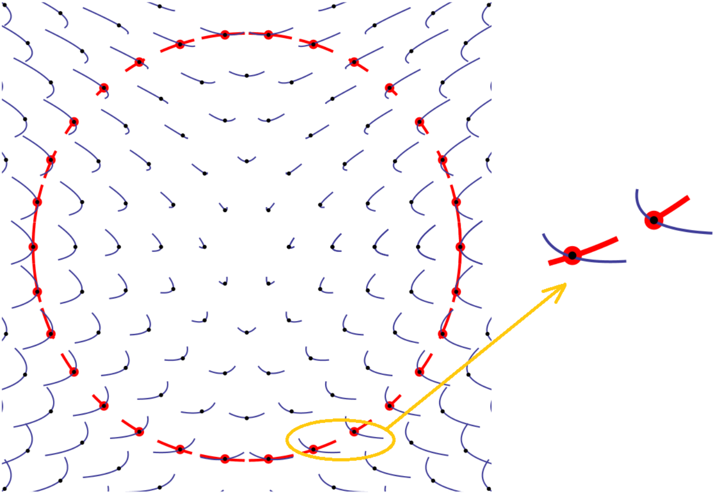
$$Z_t \smile j_2(\pi \circ \gamma_{Z_t}(dW_t)), \quad Z_0 = X_0.$$

We show the example of the cross diffusion in $\mathbb{R}^2$:

$$\begin{aligned} \mathrm{d}X_t &= \sigma Y_t \, dW_t, \\ \mathrm{d}Y_t &= \sigma X_t \, dW_t, \end{aligned} \tag{6}$$

We wish to project this process equation onto the 1-dimensional unit circle M given by $X^2 + Y^2 = 1$. We assume $(X_0, Y_0) \in M$.

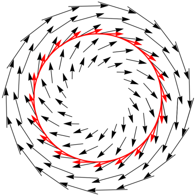
SDE $X_t \rightsquigarrow j_2(\gamma_{X_t}(dW_t))$ has Ito jet projection $Z_t \rightsquigarrow j_2(\pi \circ \gamma_{Z_t}(dW_t))$

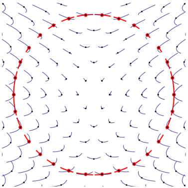


Best probabilistic (mean square) optimality of 3 projections

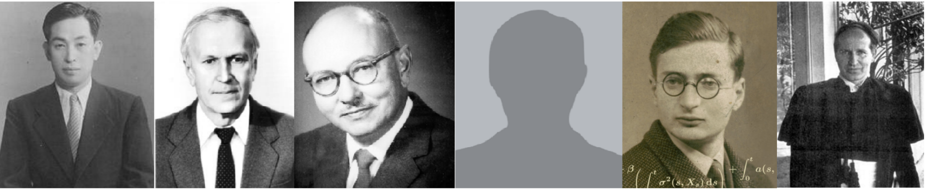
Conclusions

- SDEs: Ito and Stratonovich. Pros and Cons.
- Make Ito SDEs good for geometry: Jet interpretation
- Jet formulation of Ito's formula and other classics
- Investigating relation with Schwartz Morphism [11] & Belopolskaja Dalecky Ito bundle [5, 15] (see paper). Jets more standard?
- At the moment Schwartz Morphism more general, works for semimartingales and the SDE driver itself is in a manifold.
- Optimal SDEs on submanifolds: dimensionality reduction
- 3 types of projections on submanifolds, the best one based on jets
- Applications to signal processing [3] (and finance?)





Thanks



KYIOSI ITO | RUSLAN STRATONOVICH | EDWARD McSHANE | DONALD FISK | WOLFGANG DOEBLIN | CHARLES EHRESMANN

With thanks to IRMA - Univ. of Strasbourg/CNRS for the invitation.

Thank you for your attention.

Questions and comments welcome

References I

[1] J. Armstrong and D. Brigo.
Dimensionality reduction for SDEs with differential geometric projection methods.
Working paper, forthcoming on arxiv.org.

[2] John Armstrong and Damiano Brigo.
Extrinsic projection of Itô SDEs on submanifolds with applications to non-linear filtering.
To appear in: Nielsen, F., Critchley, F., & Dodson, K. (Eds), Computational Information Geometry for Image and Signal Processing, Springer Verlag, 2016.

References II

[3] John Armstrong and Damiano Brigo.
Nonlinear filtering via stochastic PDE projection on mixture manifolds in L^2 direct metric.
Mathematics of Control, Signals and Systems, 28(1):1–33, 2016.

[4] John Armstrong and Damiano Brigo.
Coordinate Free Stochastic Differential Equations as Jets.
http://arxiv.org/abs/1602.03931

[5] Ya. Belopolskaja and Yu. Dalecky.
Stochastic Equations and Differential Geometry.
Mathematics and Its Applications, Vol. 30. Dordrecht, Kluwer Academic Publishers, Boston, London, 1990.

References III

[6] Bjork, T.
Interest Rate Dynamics and Consistent Forward Rate Curves.
Mathematical Finance, Volume 9, Issue 4, 1999.

[7] D Brigo, B Hanzon, and F LeGland.
A differential geometric approach to nonlinear filtering: The projection filter.
IEEE Transactions on automatic control, 43:247–252, 1998.

[8] D. Brigo, B. Hanzon, and F. LeGland.
Approximate nonlinear filtering by projection on exponential manifolds of densities.
Bernoulli, 5(3):495–534, 06 1999.

References IV

[9] K. D. Elworthy.
Stochastic differential equations on manifolds.
Cambridge University Press, Cambridge, New York, 1982.

[10] K. D. Elworthy.
Geometric aspects of diffusions on manifolds.
In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 277–425. Springer, 1988.

[11] M. Emery.
Stochastic calculus in manifolds.
Springer-Verlag, Heidelberg, 1989.

References V

[12] D. Fisk.
Quasi-martingales and stochastic integrals.
PhD Dissertation, Michigan State University, Department of Statistics, 1963.

[13] Avner Friedman.
Partial differential equations of parabolic type.
Prentice-Hall, reprinted in 2008 by Dover-Publications, Englewood Cliffs N.J, 1964.

[14] Avner Friedman.
Stochastic differential equations and applications, vol. I and II.
Academic Press, reprinted by Dover Publications, New York, 1975.

References VI

[15] Yuri E Gliklikh.
Global and Stochastic Analysis with Applications to Mathematical Physics.
Theoretical and Mathematical Physics. Springer, London, 2011.

[16] I. Gyongy.
A note on Eulers approximations.
Potential Analysis, (8):205–216, 1998.

[17] Steven L Heston.
A closed-form solution for options with stochastic volatility with applications to bond and currency options.
Review of financial studies, 6(2):327–343, 1993.

References VII

[18] K. Itô.
Stochastic differential equations in a differentiable manifold.
Nagoya Math. J., (1):35–47, 1950.

[19] P. E. Kloeden and E. Platen.
Higher-order implicit strong numerical schemes for stochastic differential equations.
Journal of statistical physics, 66(1-2):283–314, 1992.

[20] Peter E. Kloeden and Eckhard Platen.
Numerical solution of stochastic differential equations.
Applications of mathematics. Springer, Berlin, New York, Third printing, 1999.

References VIII

[21] S. Kobayashi and K. Nomizu.
Foundations of differential geometry, volume 1.
New York, 1963.

[22] Graham, R.L., Knuth, D.E., and O. Patashnik (1994).
Concrete Mathematics, 2nd Ed., ADDISON-WESLEY.

[23] E.J. McShane.
Stochastic Calculus and Stochastic Models (2nd Ed.).
Academic Press, New York, 1974.

[24] G. A. Pavliotis.
Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations.
Springer, Heidelberg, 2014.

References IX

[25] K. Pearson.
Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material.
Philosophical Transactions of the Royal Society of London. A, 186:343–414, 1895.

[26] Protter, P. (2004).
Stochastic Integration and Differential Equations. Springer Verlag.

[27] L. C. G. Rogers and D. Williams.
Diffusions, markov processes and martingales, vol 2: Ito calculus, 1987.

References X

[28] A. V. Skorokhod (1967).
R. L. Stratonovich “Conditional Markov processes and their application to the theory of optimal control” (book review), Teor. Veroyatnost. i Primenen., 12:1 (1967), 184187.

[29] R. L. Stratonovich.
A new representation for stochastic integrals and equations.
SIAM Journal on Control, 4(2):362–371, 1966.

[30] N. G. Van Kampen.
Itô vs Stratonovich.
Journal of Statistical Physics, 24(1), 1981.

Bonus material

The following material did not fit the talk for matters of time, but it is here in case I need to discuss it during questions.

2-jets driven by vector Brownian motion

Consider functions $\gamma_x : \mathbb{R}^d \to \mathbb{R}^n$ & as before the coordinate free

$$X_{t+\delta t} := \gamma_{X_t} \left(\delta W_t^1, \ldots, \delta W_t^d \right)$$

Again, the limiting behaviour will only depend upon the 2-jet $j_2(\gamma_x)$ and can still be denoted by $X_t \smile\!\!\nearrow j_2(\gamma_{X_t})(dW_t)$. The scheme still $\xrightarrow{L^2(\mathbb{P})}$ to the classical Itô SDE (see proof in A & B [4]) in coordinates:

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t a(\tilde{X}_s)\, \mathrm{d}s + \sum_{\alpha=1}^{d} \int_0^t b_\alpha(\tilde{X}_s)\, \mathrm{d}W_s^\alpha, \quad t \in [0, T]$$

$$a(x) := \frac{1}{2} \sum_{\alpha=1}^{d} \frac{\partial^2 \gamma_x}{\partial u^\alpha \partial u^\alpha}\bigg|_{u=0}, \quad b_\alpha(x) := \frac{\partial \gamma_x}{\partial u^\alpha}\bigg|_{u=0}.$$

SDEs driven by vector Brownians

We can also write the SDE as

$$\mathrm{d}X_t^i = \frac{1}{2}\partial_\alpha\partial_\beta\gamma^i \mathrm{d}W_t^\alpha \mathrm{d}W_t^\beta + \partial_\alpha\gamma^i\,\mathrm{d}W_t^\alpha = \frac{1}{2}\partial_\alpha\partial_\beta\gamma^i g_E^{\alpha\beta} dt + \partial_\alpha\gamma^i\,\mathrm{d}W_t^\alpha \quad (7)$$

with the convention that $\mathrm{d}W_t^\alpha \mathrm{d}W_t^\beta = g_E^{\alpha\beta}\mathrm{d}t$ where g_E is a Kronecker delta with orthonormal coordinates, or in generalizations the symmetric 2-form defining the Euclidean metric on $\mathbb{R}^d$.

2-jet based definition of the SDE **backward diffusion operator**:

$$\mathcal{L}_{\gamma_x} f := \frac{1}{2}\Delta_E(f \circ \gamma_x) = \frac{1}{2}\partial_\alpha\partial_\beta(f \circ \gamma_x) g_E^{\alpha\beta}. \quad (8)$$

Here Δ_E is the Laplacian defined on $\mathbb{R}^d$. $\mathcal{L}_{\gamma_x}$ acts on functions defined on the state space manifold M. We define $\mathcal{L}^*$ to be its formal adjoint which acts on densities defined on M (Fokker Planck eq).

Weak and Strong Equivalence of SDEs through jets

Both the Itô SDE (7) & the backward diffusion operator use only part of the 2-jet: only the diagonal terms of $\partial_\alpha \partial_\beta \gamma^i$ influence the SDE and even for these terms it is only their sum that is important.

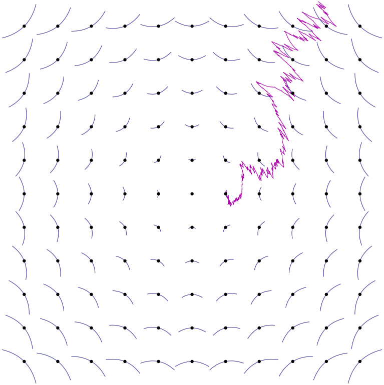
We say that two 2-jets γ^1_x and γ^2_x are *weakly equivalent* if $\mathcal{L}_{\gamma^1_x} = \mathcal{L}_{\gamma^2_x}$.

γ^1 and γ^2 are *strongly equivalent* if in addition $j_1(\gamma^1) = j_1(\gamma^2)$.

Strong equivalence means that given the same realization of the driving Brownian motions W^α_t the solutions of the SDEs will be almost surely the same (under assumptions ensuring pathwise uniqueness).

Weak equivalence means that the transition probability distributions are the same even though the dynamics may be different for any specific realisation of the Brownian motions.

Drawing SDEs driven by 2-dimensional Brownians



We saw previously a way to draw a SDE in $\mathbb{R}^2$, $j_2(\gamma^E)$, driven by one-dimensional Brownian motion:

$$d \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = 3 \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} -X_2 \\ X_1 \end{bmatrix} dW_t.$$

How can we draw a SDE driven by 2-dimensional Brownian motion?

Given an SDE in local coordinates $\mathrm{d}X_t = a(X_t)\mathrm{d}t + b_i(X_t)\mathrm{d}W_t^i$ (Einstein summation) with $a \in \mathbb{R}^2$ and $b_1 \in \mathbb{R}^2, b_2 \in R^2$, we can write down a specific representative two jet by

$$\gamma_x(t^1, t^2) = x + a g_{ij}^E t^i t^j + b_i t^i.$$

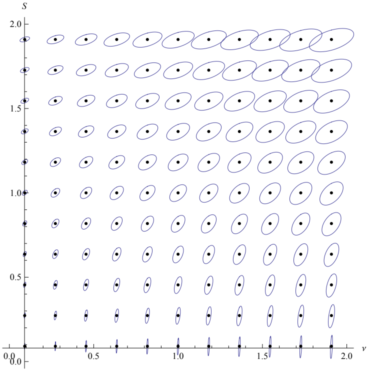
Drawing SDEs driven by 2-dimensional Brownians

$$\gamma_x(t^1, t^2) = x + a g^E_{ij} t^i t^j + b_i t^i.$$

The image of an ϵ ball under γ_x will be an ellipsoid. Moreover, if we know that γ_x is of this form, we can recover the coefficients a and b_i up to weak equivalence just from knowledge of the image of the ϵ ball.

This method of drawing an $\mathbb{R}^2$ SDE driven by 2-dim Brownian motion in local coordinates is to draw the image of an ϵ ball of (t^1, t^2) at each point.

Drawing SDEs driven by 2-dimensional Brownians



For example in this figure we show a plot of the Heston stochastic volatility model with drift (see [17]). Note that as well as plotting the ellipses, the figure indicates the exact point that each ellipse is associated with. The extent to which the centre of the ellipse differs from the associated point is a measure of the drift.

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sqrt{\nu_t} S_t \mathrm{d}W_t^1 \\ \mathrm{d}\nu_t &= \kappa(\theta - \nu_t)\mathrm{d}t + \xi\sqrt{\nu_t}\left(\rho \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2\right) \end{aligned} \tag{9}$$

Parameter values $\xi = 1$, $\theta = 0.4$, $\kappa = 1$, $\mu = 0.1$, $\rho = 0.5$. We have plotted the image of the balls for $\epsilon = 0.05$.

The one-dimensional case: Fan Diagrams

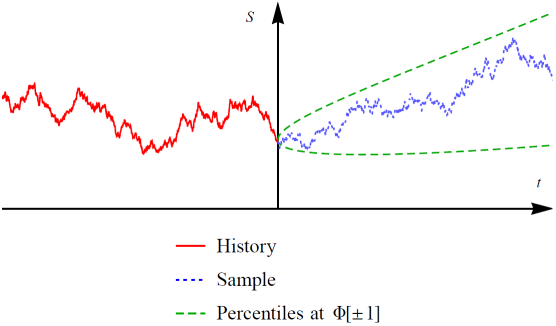
Standard statistical properties of a distribution depend upon the coordinate system.

For example $\mathbb{E}$ of a process in $\mathbb{R}^n$ involves the vector space structure of $\mathbb{R}^n$. If f is a nonlinear coordinate transition map, one has $\mathbb{E}(f(X)) \neq f(\mathbb{E}(X))$.

However, the definition of the α-percentile depends only upon the ordering of $\mathbb{R}$ and not its vector space structure.

As a result, for continuous monotonic f and X with connected state space, the median of $f(X)$ is equal to f applied to the median of X. If f is strictly increasing, the analogous result holds for the α percentile.

The one-dimensional case: Fan Diagrams



This has the implication that the trajectory of the α-percentile of an $\mathbb{R}$ valued stochastic process is invariant under smooth monotonic coordinate changes of $\mathbb{R}$. **In other words, percentiles have a coordinate free interpretation.** How can the trajectories of percentiles be related to the coefficients of the SDE?

The one-dimensional case: Fan Diagrams

Theorem (Armstrong and B.) [4]. For sufficiently small t, the α-th percentile of the solutions to

$$\mathrm{d}X_t = a(X_t,t)\,\mathrm{d}t + b(X_t,t)\mathrm{d}W_t, \qquad X_0 = x_0 \tag{10}$$

is given by: $$x_0 + b_0\sqrt{t}\Phi^{-1}(\alpha) + \left[a_0 - \frac{b_0 b_0'}{2}(1-\Phi^{-1}(\alpha)^2)\right] t + O(t^{3/2})$$

so long as the coefficients of (10) are smooth, the diffusion coefficient b never vanishes, and sufficient conditions for the Lamperti transformed SDE and for $\mathcal{L}^* p = 0$ to have a unique regular solution hold. In this formula a_0 and b_0 denote the values of $a(x_0,0)$ and $b(x_0,0)$ respectively. In particular, **the median process is a straight line up to $O(t^{\frac{3}{2}})$ with tangent given by the drift of the Stratonovich version of the Itô SDE** (10)**.The $\Phi(1)$ and $\Phi(-1)$ percentiles correspond up to $O(t^{\frac{3}{2}})$ to the curves $\gamma_{X_0}(\pm\sqrt{t})$ where γ_{X_0} is any representative of the 2-jet that defines the SDE in Itô form.**

Jets & vector fields: Ito / Str as different coordinates

We have seen that, geometrically, a Str SDE is described by 2 vector fields, while a Ito SDE is described by one 2-jet. We now relate the two.

An alternative way to specify the k-jet of a curve at every point is to choose k vector fields $A_1, \dots, A_k$ on the manifold. One can then define $\Phi^t_{A_i}$ to be the vector flow associated with the vector field A_i. This allows one to define curves at each point x as follows:

$$\gamma_x(t) = \Phi^{t^k}_{A_k}(\Phi^{t^{k-1}}_{A_{k-1}}(\dots(\Phi^t_{A_1}(x))\dots)) \tag{11}$$

where t^k denotes the k-th power of t. We will call this the *vector representation* for a family of k-jets.

Theorem (Armstrong B. (2016)). All k-jets of curves can be represented this way via vector fields flows.

Jets & vector fields: Ito / Str as different coordinates

Corollary (Ito Stratonovich transformation as correspondence between 2-jets and two vector fields.)

Suppose that a family of 2-jets of curves is given in the vector representation as

$$\gamma_x(t) = \Phi_A^{t^2}(\Phi_B^t(x))$$

for vector fields A and B. Choose a coordinate chart and let A^i, B^i be the components of the vector fields in this chart. Then the corresponding standard representation for the family of 2-jets is:

$$\gamma_x(t) = x + a(x)t^2 + b(x)t$$

with

$$a^i = A^i + \frac{1}{2}\frac{\partial B^i}{\partial x^j}B^j, \quad b^i = B^i.$$

Jets & vector fields: Ito / Str as different coordinates

Geometric interpretation of the Ito-Stratonovich transformation: switching between 2-jets and pairs of vector fields.

Despite Itô's 1950 paper [18] on SDEs on manifolds based on using Itô's lemma to change coordinates, a few authors have even asserted that stochastic differential geometry *requires* Stratonovich calculus.

From an extrinsic perspective (i.e. manifolds embedded in $\mathbb{R}^n$ instead of charts) Stratonovich may appear necessary since an SDE remains on a submanifold a.s. if Str-drift and Str-diffusion vector fields are tangent to the manifold.

It is easy to write down the Stratonovich SDE induced on a submanifold from a Str SDE on $\mathbb{R}^n$. However, this is simply a consequence of the curvature of the 2-jet following the curvature of the manifold, so the Itô/2-jet interpretation works as well.

From our point of view we consider these two calculi as different coordinate systems for the same underlying coordinate-free SDE.

Many notions in probability are not coordinate free however (the expected value $\mathbb{E}$ for example, but see also our earlier discussion on the assumed density principle).

Jets & vector fields: Ito / Str as different coordinates

One should choose the most convenient coordinate system for the problem at hand along the properties we highlighted in the introduction (Wong Zakai convergence, martingale, anticipative features, etc).

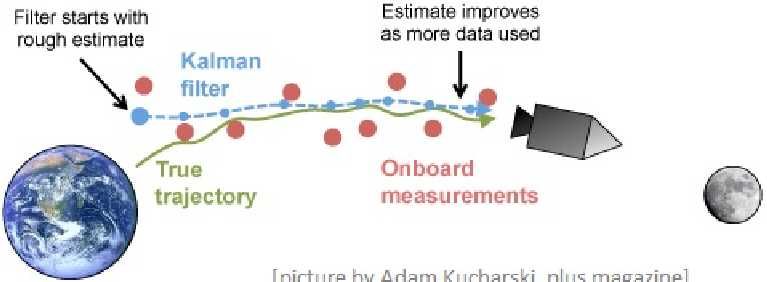
The most important difference between Stratonovich & Itô arises during the modelling process. It is when choosing what equation to write down in the first place that the choice is most telling.

The modelling process is not a strictly mathematical process: it relies upon the modellers intuition.

So fortunately “The ultimate goal of mathematics is to eliminate all need for intelligent thought” [1] does not seem to apply here.

[1] (Graham, Knuth and Patashnik [22])

Filtering problem (e.g. Apollo 11)



[picture by Adam Kucharski, plus magazine]

Signal and observation equations:

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dB_t, \quad dR_t = h(S_t, t)dt + \epsilon dV_t$$

where B, V are independent Brownian motions (noises).

Given observations R from 0 to t, estimate S_t. Full solution:
$p_t(\xi)d\xi = \mathbb{P}\{S_t \in d\xi | \sigma(R_s, s \in [0, t])\}$. Point estimate: mean p_t.

Filtering problem and SPDE projection to SDE

p_t is our previous X but now in an infinite dimensional function space (typically $\sqrt{p}$ or p in L^2) that plays the role of our former $\mathbb{R}^r$.

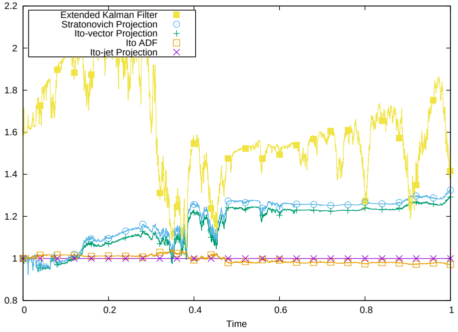
p_t follows a SPDE (Kushner-Stratonovich or Zakai) and we can use our three above projections to estimate an optimal finite dimensional approximation Y of $p = X$ according to different criteria.

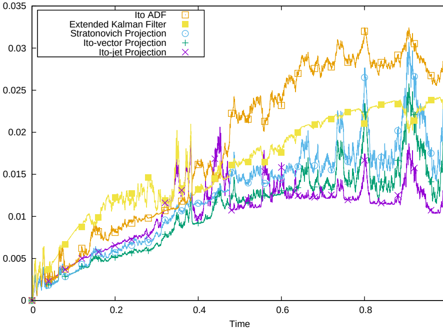
In B., Hanzon & LeGland [7, 8], Armstrong & B. [3] we study projections on $M =$ Gaussians (here), exponential families, mixtures.

Filtering numerical example: Cubic sensor

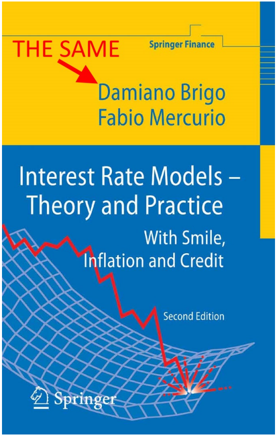
$$dS_t = dB_t, \quad dR_t = (S_t + \alpha S_t^3)dt + dV_t$$

Hellinger rel. residuals: $\frac{\|\sqrt{p(t)} - \sqrt{p_{\mathcal{N}}(t)}\|_2}{\|\sqrt{p(t)} - \sqrt{p_{\mathcal{N},Ito-jet}(t)}\|_2}$; L^2 resid.: $\|p(t) - p_{\mathcal{N}}(t)\|_2$





Consistency rate dynamics - curve parameterization



In finance, we use short rate interest rate model $r_t = \varphi(t, X_t)$, where X follows a driving SDE in $\mathbb{R}^r$. X is chosen based on history and derivatives prices (calibration). Spot rate at t for maturity T is

$$R(t,T) = \frac{1}{T-t} \ln \left(\mathbb{E}_t \left[\exp \left(- \int_t^T r_s ds \right) \right] \right)$$

Practitoners wish curve $T \mapsto R(t,T)$ to have a particular parametric shape, $R(t,T) = R(T; \theta(t))$, $\theta \in \mathbb{R}^n$.

Use the projection framework to try and optimally approximate the correct dynamics of dR coming from dX with one on “manifold” $R(\theta)$. Related work was done in the 90’s by Bjork [6] but looking for exact results rather than optimal approximations.