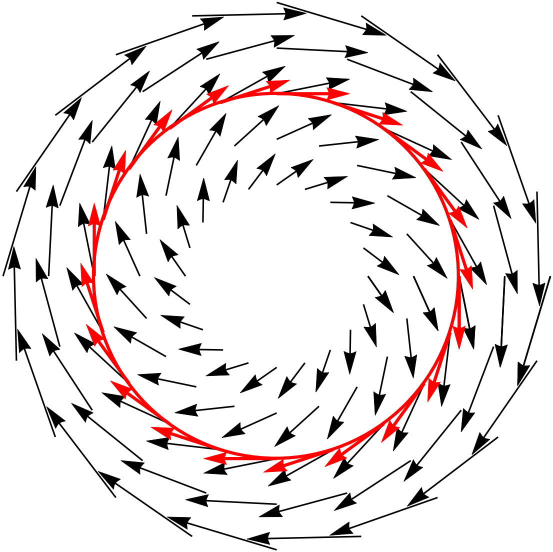
Geometric projection of stochastic differential equations

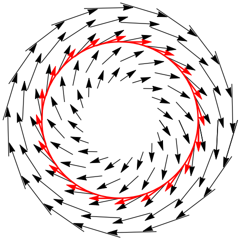
John Armstrong (King's College London)
Damiano Brigo (Imperial)

November 2, 2018

Idea: Projection



Idea: Projection

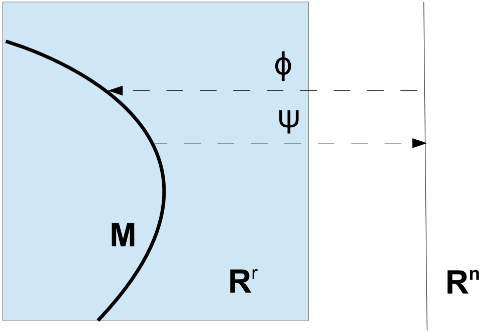


- Projection gives a method of systematically reducing the dimension of an ODE
- Projection onto a linear subspace is the standard numerical method for solving PDEs
- Projecting onto a curved manifold may be more effective if we know the solution is close to this manifold
- e.g. perhaps the known soliton solutions to the KdV equation might give good approximations to solutions to a pertubed KdeV equation?

This talk

- Question: How should the notion of projection be extended to stochastic differential equations?
- Answer:
  - There is a Stratonovich Projection which is best understood using Stratonovich calculus.
  - There is an Extrinsic Ito Projection which is best understood using Ito calculus.
  - There is an Intrinsic Ito Projection which is best understood by using jet bundles.
  - ...
- We will
  - Define these various notions of projection and discuss their motivation and theoretical justifications
  - Describe a geometric formulation of SDEs using 2-jets to understand the Intrinsic Ito projection
  - Look at some numerical results when projection is applied to nonlinear filtering

Setup

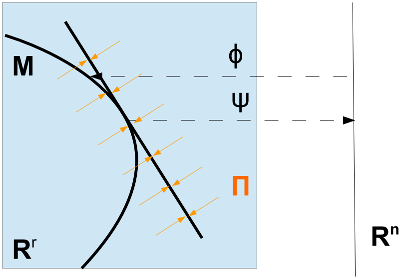


- M is a submanifold of $\mathbb{R}^r$
- $\psi : U \to \mathbb{R}^n$ is a chart for M
- $\phi = \psi^{-1}$
- We have an SDE on $\mathbb{R}^r$

$$\mathrm{d}X_t = a\,\mathrm{d}t + \sum_{\alpha} b_\alpha\,\mathrm{d}W_t^\alpha, \qquad X_0$$

and want to approximate this using an SDE on $\mathbb{R}^n$.

Definition: Stratonovich projection



1. Write the SDE in Stratonovich form

$$\mathrm{d}X_t = \overline{a(X_t)}\,\mathrm{d}t + \sum_\alpha b_\alpha(X_t) \circ \mathrm{d}W_t^\alpha, \qquad X_0$$

2. Apply the projection operator Π to each coefficient to obtain an SDE on M

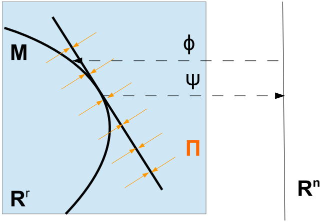
$$\mathrm{d}X_t = \Pi_{X_t}\overline{a(X_t)}\,\mathrm{d}t + \sum_\alpha \Pi_{X_t} b_\alpha(X_t) \circ \mathrm{d}W_t^\alpha, \qquad \psi(X_0)$$

Justifications

What are the justifications for using the Stratonovich projection?

- It is clearly a well defined SDE. (Contrast with projecting Itô coefficients)
- It is clearly generalizes projection of ODEs - i.e. when $b = 0$ we get ODE projection.
- It gives good numerical results when applied to the filtering problem
- It generalizes the Galerkin method which can be interpreted as projection onto a linear subspace.

A justification for ODE projection



- Consider an ODE on $\mathbb{R}^r$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = a(X), \qquad X_0$$

- Look for an ODE on $\mathbb{R}^n$ of the form

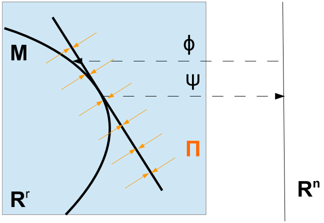
$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x), \qquad \psi(X_0)$$

  such that

$$|\phi(x_t) - X_t|^2$$

  is as small as possible.

A justification for ODE projection

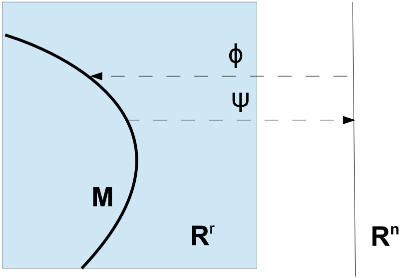


- Compute Taylor expansion to see that leading term is minimized when:

$$a(\psi(x_0)) = \psi_* \Pi_{X_0} A(x_0)$$

- Therefore ODE projection is the unique asymptotically optimal ODE approximating the original ODE at all points on M.
- (Linear projection operator gives solution to a quadratic optimization problem)

Repeat idea for SDEs



Equation in larger space $\mathbb{R}^r$:
$\mathrm{d}X = a(X,t)\,\mathrm{d}t + b(X,t)\,\mathrm{d}W_t$

Equation in chart:
$\mathrm{d}x = A(x,t)\,\mathrm{d}t + B(x,t)\,\mathrm{d}W_t$

We have Itô Taylor series estimates (Kloeden and Platen):

$$E(|X_t - \phi(x_t)|) = |b_0 - \phi_* B_0|\sqrt{t} + O(t)$$

$$|E(X_t - \phi(x_t))| = \left| a_0 - \phi_* A_0 - \frac{1}{2}(\nabla_{B_{\alpha,0}}\phi_*)B_{\beta,0}[W^\alpha, W^\beta] \right| t + O(t^2)$$

Extrinsic Ito Projection

To minimize first estimate:

$$\phi_* B = \Pi b$$

If we define B like this for whole chart, second estimate is minimized when:

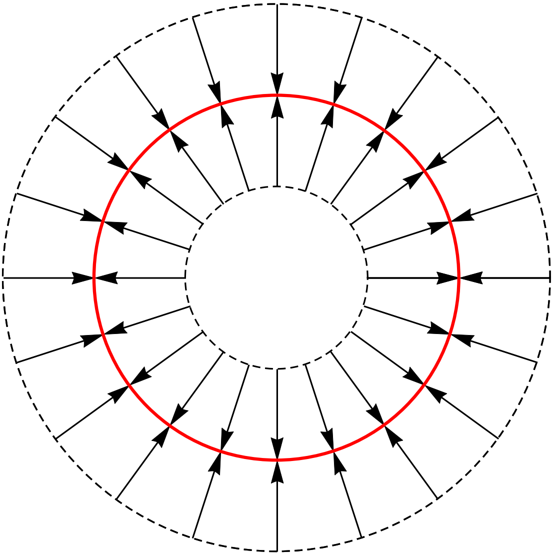
$$\phi_* A = \Pi a - \frac{1}{2} \Pi (\nabla_{B_\alpha} \phi_*) B_\beta [W^\alpha, W^\beta]$$

- Given ϕ, define A and B using these equations
- This defines an SDE on the manifold
- We call this the Extrinsic Itô projection
- It is different from the Stratonovich projection

Discussion

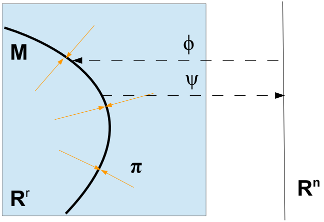
- The Extinsic Itô Projection is optimal in the sense that it asymptotically minimizes two measures of the divergence of the approximation to the SDE from the true solution.
- Measure one is on the expectation of the absolute value. This determines the martingale part of our equation
- Measure two is on the absolute value of the expectation. This determines the bounded variation part of our equation
- The Extrinsic Itô Projection is "greedy" in that it finds the best approximation over short time horizons and hopes they will do well over long time horizons.
- Numerical test on the filtering problem indicate that it slightly outperforms the Stratonovich projection in practice over moderate time horizons.
- Over longer time horizons, it is random which performs better.

Geodesic projection map



Let π denote the smooth map defined on a tubular neighbourhood of M that projects $\mathbb{R}^\ell$ onto M along geodesics.

An alternative justification for ODE projection



- Consider an ODE on $\mathbb{R}^r$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = a(X), \qquad X_0$$

- Look for an ODE on $\mathbb{R}^n$ of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x), \qquad \psi(X_0)$$

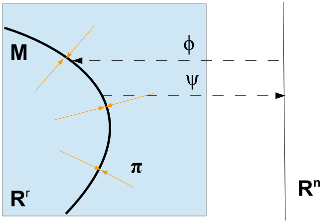
such that

$$\sout{|\phi(x_t) - X_t|^2}$$

$$d(x_t, \psi \circ \pi(X_t))$$

is as small as possible. d is induced Riemannian distance.

Intrinsic Itô projection



Repeating the ideas used to derive the Extrinsic Itô projection:

Definition

The Intrinsic Itô projection is the best approximation to $\pi(X_t)$ in the sense that it asymptotically minimizes both:

$$E(d(x_t, \psi \circ \pi((X_t)))$$

$$d(E(x_t), E(\psi \circ \pi(X_t)))$$

Discussion

All three projections are distinct. Which is better?

Lemma

(Factorizable SDEs) Suppose that S is an SDE for X on $\mathbb{R}^r$ such that $\pi(X)$ solves an SDE S' on M then the Stratonovich and intrinic Itô projections are both equal to S'. However, the extrinsic projection may be different.

Example

The SDE S on $\mathbb{R}^2$

$$\begin{aligned} \mathrm{d}X_t &= \sigma Y_t \,\mathrm{d}W_t \\ \mathrm{d}Y_t &= \sigma X_t \,\mathrm{d}W_t \end{aligned}$$

In polar coordinates, solutions satisfy:

$$\mathrm{d}\theta = -\frac{1}{2}\sigma^2 \sin(4\theta)\,\mathrm{d}t + \sigma \cos(2\theta)\,\mathrm{d}W_t$$

Understanding the Intrinsic Itô projection

Definition

The Intrinsic Itô projection is the best approximation to $\pi(X_t)$ in the sense that it asymptotically minimizes both:

$$E(d(x_t, \psi \circ \pi((X_t)))$$

$$d(E(x_t), E(\psi \circ \pi(X_t)))$$

- For applications, one must calculate this in local coordinates, but the resulting expression is complex
- One can understand this projection more intuitively, and express the answer more elegantly, using the language of 2-jets.

Euler Scheme

- All being well in the limit the Euler scheme

$$\delta X_t = a(X)\,\delta t + b(X)\,\delta W_t$$

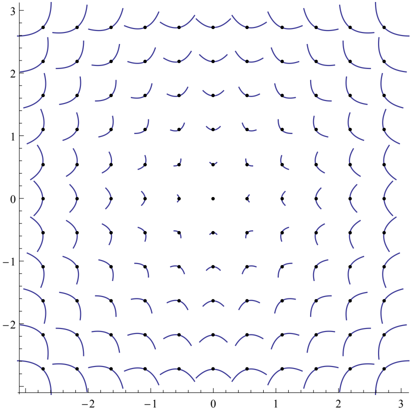
  converges to a solution of the SDE

$$\mathrm{d}X_t = a(X)\,\mathrm{d}t + b(X)\,\mathrm{d}W_t$$

- d, δ, $+$ imply vector space structure
- This is highly coordinate dependent

Curved Scheme

Let γ_x be a choice of curve at each point x of M. $\gamma_x(0) = x$.

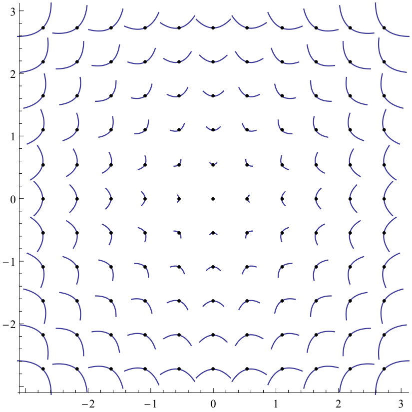


Consider the scheme

$$X_{t+\delta t} = \gamma_{X_t}(\delta W_t) \qquad X_0$$

Concrete example

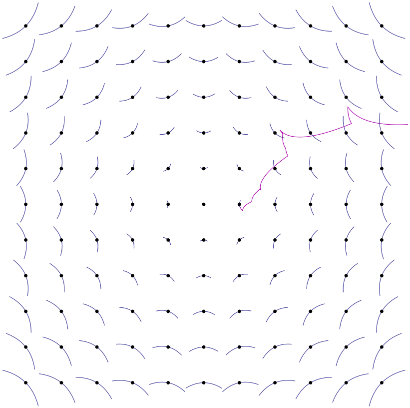
$$\gamma^{E}_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$



- First order term is rotational vector
- Second order term is axial vector

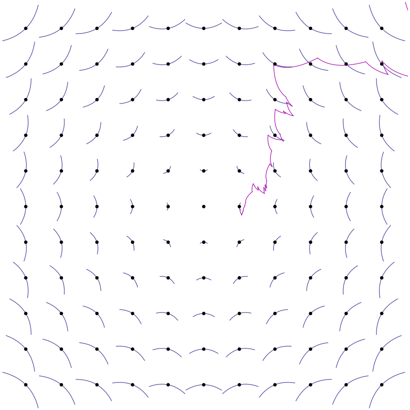
Simulation: Large time step

$$\gamma^E_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$



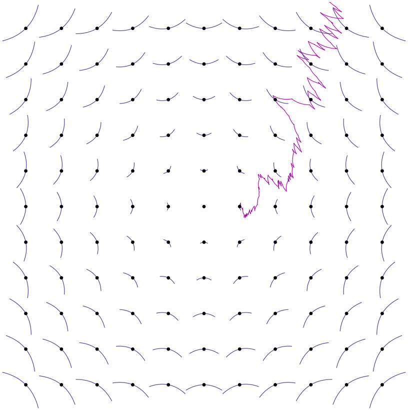
Simulation: Smaller time step

$$\gamma^E_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$



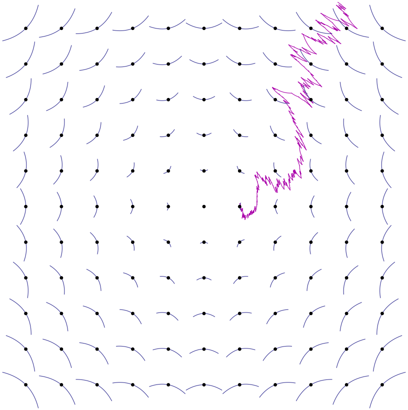
Simulation: Even smaller

$$\gamma^E_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$



Simulation: Convergence

$$\gamma^{E}_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$



Formal argument

Write:

$$\gamma_x(s) = x + \gamma_x'(0)s + \frac{1}{2}\gamma_x''(0)s^2 + O(s^3)$$

Then:

$$\begin{aligned} X_{t+\delta t} &= \gamma_t(\delta W_t) \\ &= X_t + \gamma_{X_t}'(0)\delta W_t + \frac{1}{2}\gamma'' X_t(0)(\delta W_t)^2 + O\left((\delta W_t)^3\right) \end{aligned}$$

Rearranging:

$$\delta X_t = X_{t+\delta t} - X_t = \gamma_{X_t}'(0)\delta W_t + \frac{1}{2}\gamma'' X_t(0)(\delta W_t)^2 + O\left((\delta W_t)^3\right)$$

Taking the limit:

$$\begin{aligned} \mathrm{d}X_t &= b(X)\mathrm{d}W_t + a(X)(\mathrm{d}W_t)^2 + O\left((\mathrm{d}W_t)^3\right) \\ &= b(X)\mathrm{d}W_t + a(X)\mathrm{d}t \end{aligned}$$

where

$$b(X) = \gamma_X'(0)$$
$$a(X) = \gamma_X''(0)/2$$

Comments

- The curved scheme depends only on the 2-jet of the curve
- SDEs driven by 1-d Brownian motion are determined by 2-jets of curves
- The first derivative determines the volatility term
- The second derivative determines the drift term

ODEs correspond to 1-jets of curves
SDEs correspond to 2-jets of curves

- Rigorous proof of convergence of quadratic scheme can be proved using standard results on Euler scheme

$$
\begin{aligned}
\mathrm{d}X_t &= a(X)\mathrm{d}t + b(X)\mathrm{d}W_t \\
&= a(X)\left(\mathrm{d}(W_t^2) - 2W_t\mathrm{d}(W_t)\right) + b(X)\mathrm{d}W_t \\
&\approx a(X)\left(\delta(W_t^2) - 2W_t\delta(W_t)\right) + b(X)\delta W_t \\
&= a(X)\left((\delta W_t)^2\right) + b(X)\delta W_t
\end{aligned}
$$

- For general curved schemes some analysis needed.

Itô's lemma

Given a family of curves γ_x we will write:

$$X_t \smile j_2\left(\gamma_x(\mathrm{d}W_t)\right)$$

if X_t is the limit of our scheme.
If

$$X_t \smile j_2\left(\gamma_x(\mathrm{d}W_t)\right)$$

and $f : X \to Y$ then:

$$f(X)_t \smile j_2\left(f \circ \gamma_x(\mathrm{d}W_t)\right)$$

Itô's lemma is simply composition of functions.

Usual formulation

$$X_t \smile j_2\left(\gamma_x(\mathrm{d}W_t)\right)$$

Is equivalent to:

$$\mathrm{d}X_t = a(X)\mathrm{d}t + b(X)\mathrm{d}W_t, \quad a(X) = \frac{1}{2}\gamma_X''(0), \quad b(X) = \gamma_X'(0)$$

We calculate the first two derivatives of $f \circ \gamma_X$:

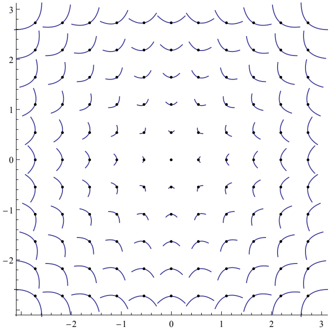
$$(f \circ \gamma_X)'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma_X(t))\frac{\mathrm{d}\gamma_X}{\mathrm{d}t}$$

$$(f \circ \gamma_X)''(t) = \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\gamma_X(t))\frac{\mathrm{d}\gamma_X^i}{\mathrm{d}t}\frac{\mathrm{d}\gamma_X^j}{\mathrm{d}t} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma_X(t))\frac{\mathrm{d}^2\gamma_X}{\mathrm{d}t^2}$$

So $f(X_t) \smile j_2\left(f \circ \gamma_x(\mathrm{d}W_t)\right)$ is equivalent to standard Itô's formula

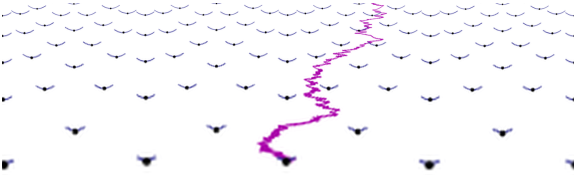
Example

$$\gamma^E_{(x_1,x_2)}(s) = (x_1, x_2) + s(-x_2, x_1) + 3s^2(x_1, x_2)$$

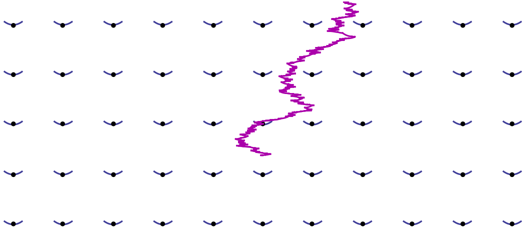


Clearly polar coordinates might be a good idea. So consider the transformation $\phi : \mathbb{R}^2/\{0\} \to [-\pi, \pi] \times \mathbb{R}$ by:

$$\phi(\exp(s)\cos(\theta), \exp(s)\sin(\theta)) = (\theta, s),$$



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

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

Drawing SDEs

The following diagram commutes:

$$
\begin{array}{ccc}
\text{SDE for X} & \xrightarrow{\text{Itô's lemma}} & \text{SDE for } f(X) \\
\text{Draw}\big\downarrow & & \text{Draw}\big\downarrow \\
\text{Picture of SDE for X in } \mathbb{R}^n & \xrightarrow{f} & f(\text{Picture of SDE for X})
\end{array}
$$

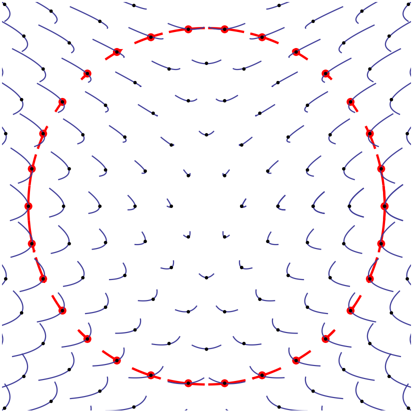
Intrinsic Itô Projection: 2-jet formulation

If original SDE is:

$$X_t \smile j_2\left(\gamma_x(\mathrm{d}W_t)\right)$$

then intrinsic Itô projection is:

$$x_t \smile j_2\left(\pi \circ \gamma_x(\mathrm{d}W_t)\right)$$



Local coordinate formulation

Calculate Taylor series for π to second order to compute:

$$\mathrm{d}x = A\,\mathrm{d}t + B_\alpha\,\mathrm{d}W_t^\alpha, \qquad x_0$$

where:

$$B_\alpha^i = (\pi_*)^i_\beta b^\beta_\alpha$$

and:

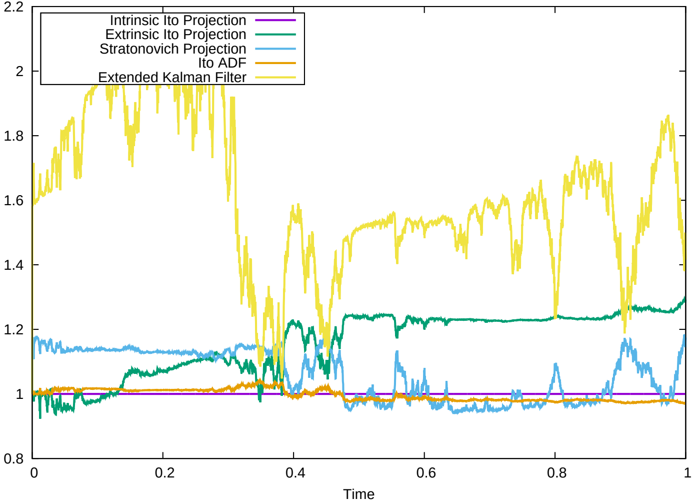
$$\begin{aligned}
A^i = {} & (\pi_*)^i_\alpha a^\alpha + \\
& \left(-\frac{1}{2}\frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} (\pi_*)^a_\gamma (\pi_*)^\alpha_\delta (\pi_*)^\beta_\epsilon \right. \\
& \left. + \frac{\partial^2 \phi^\epsilon}{\partial x^\alpha \partial x^\beta} (\pi_*)^\beta_\delta h^{a\alpha} - \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} (\pi_*)^\beta_\epsilon (\pi_*)^\eta_\gamma (\pi_*)^\zeta_\delta h_{\eta\zeta} h^{a\alpha} \right) \\
& \times b^\delta_\kappa b^\epsilon_\iota [W^\kappa, W^\iota]_t .
\end{aligned}$$

Numerical example

- The linear filtering problem has solutions given by Gaussian distributions
- Maybe approximately linear filtering problems can be well approximated by Gaussian distributions?
- Heuristic algorithms:
  - Extended Kalman Filter
  - Itô Assumed Density Filter
  - Stratonovich Assumed Density Filter
  - Stratonovich Projection Filter
- Algortihms based on optimization arguments:
  - Extrinsic Itô Projection Filter
  - Intrinsic Itô Projection Filter

Relative performance (Hellinger Residuals)

All projections performed w.r.t. the Hellinger metric.



Summary - projection methods

	Extrinsic Ito	Intrinsic Ito	Stratonovich
Optimal?	Yes	Yes	
Factorizable SDE	Surprising	Expected	Expected
Aesthetics		Elegant	
Practice	Best short term	Best medium term	Acceptable

- Note that our notion of optimal is based on expectation of squared residuals
- Other "risk measures" could be used

Summary - 2 jets

- 2-jets allow you to draw pictures of SDEs
- They provide an intuitive and elegant reformulation of Itô's lemma
- They provide an alternative route to coordinate free stochastic differential geometry to operator opproaches