Ito projection and the optimal Gaussian filter

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

November 2, 2018

Introduction

Plan:

- Outline of Stochastic Filtering
- Projecting ODEs and SDEs
- Ito SDEs on manifolds
- Introducing the Ito projection
- Numerical results
- Geometric interpretation of SDE and projection

General filtering problem

$$\begin{aligned} \mathrm{d}X_t &= f(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t \\ \mathrm{d}Y_t &= b(X_t, t)\,\mathrm{d}t + \mathrm{d}V_t \end{aligned}$$

Q: We have a prior distribution p_0 for X. What is p_t?
A: (Ignoring all technicalities) The *Zakai equation*

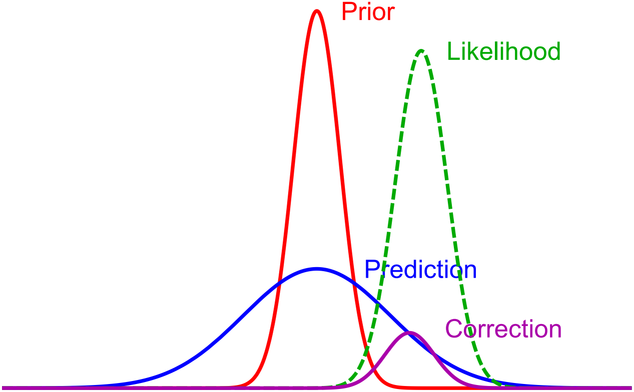
$$\mathrm{d}p = \mathcal{L}^* p \;\mathrm{d}t + p b^T \mathrm{d}Y_t$$

where p is the *likelihood* a.k.a. the unnormalized density.
Alternatively, the Kushner-Stratonovich equation:

$$\mathrm{d}p = \mathcal{L}^* p + p(b - E_p(b))(\mathrm{d}Y_t - E_p(b)\,\mathrm{d}t)$$

Justification

$$\mathrm{d}p = \mathcal{L}^* p \; \mathrm{d}t + p b^T \mathrm{d}Y_t$$
$$= \text{prediction} + \text{correction}$$



Note that for linear filter, Gaussian stays Gaussian

Problem

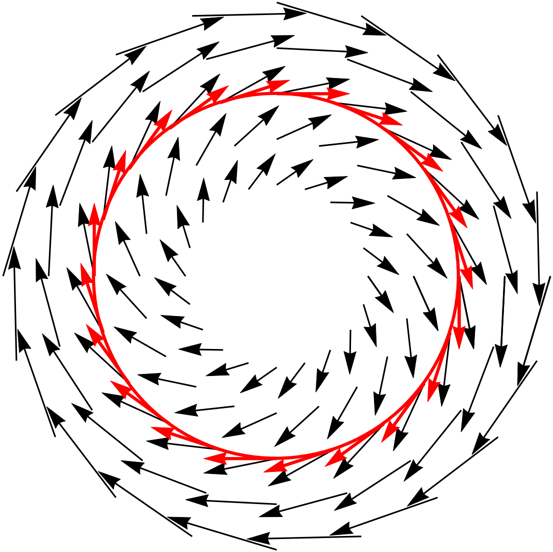
Solution approaches:

- Finite difference methods
- Spectral methods
- Monte Carlo (particle filters)
- ...

Solution goals:

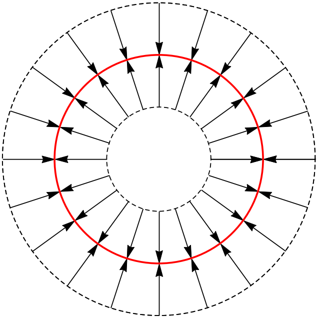
- Moderate dimensions
- Moderate accuracy
- Rapid calculation

Idea: Projection

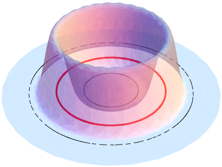


Stochastic projection: Very naive version

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



Π



ρ

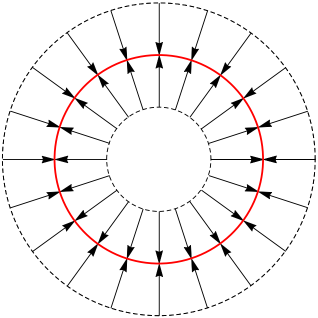
Projected equation?

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

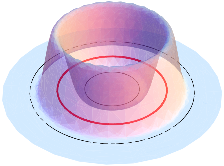
Stochastic projection: Stratonovich version

Fix? Use Stratonovich SDE:

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



Π



ρ

Projected equation?

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

The space of densities?

We need a Hilbert space to define projection. Obvious choices are:

- The space of densities with the L^2 metric: $\mathcal{P} \subseteq L^2(\mathbb{R}^n)$

$$\langle p, q \rangle_{L^2} = \int p(x)q(x)\,\mathrm{d}x$$

- The space of densities with the Hellinger metric: $\mathcal{P}'$

$$\langle p, q \rangle_H = \int \sqrt{p(x)q(x)}\,\mathrm{d}x$$

  $\mathcal{P}' \subseteq L^2(\mathbb{R}^n)$ via $p \to \sqrt{p}$

- Hellinger metric is independent of parameterizations of $\mathbb{R}^n$ and exists for all measures, not just densities.
- L^2 metric works well for mixture families (preserves linearity)
- Hellinger metric works well for exponential families (correction step exact)

Stratonovich projection works well

Stratonovich projection of the filtering equation has been tried for the following manifolds in the space of densities:

- Project onto a linear subspace = Galerkin method
- Project onto an exponential family, e.g.

$$p(x) = \exp(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n)$$

$$a_n < 0, \qquad n \text{ even}, \qquad \int p(x) = 1$$

- Project onto a mixture of Gaussians, e.g.

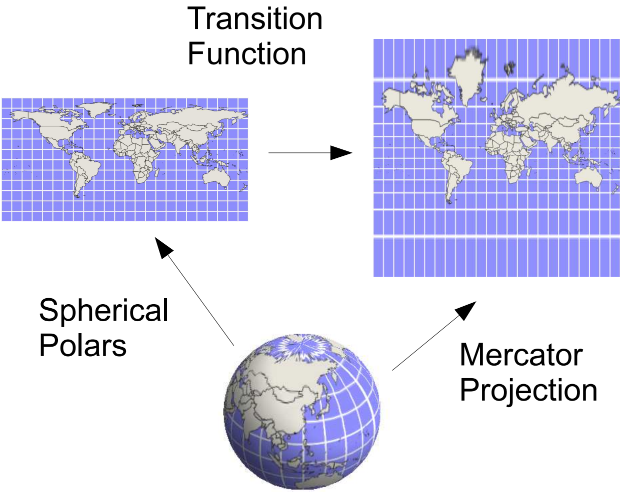
$$p(x) = \sum_i \pi_i N(x, \mu_i, \sigma_i)$$

$$\sum_i \pi_i = 1$$

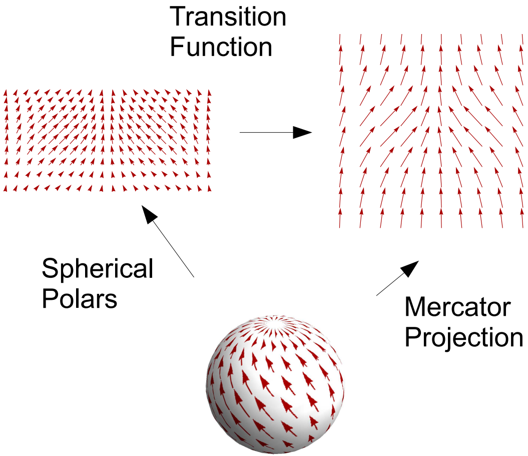
Theoretical Results

- What theoretical results back this idea up?
  - Galerkin method converges in many circumstances
  - Projection onto exponential families is accurate for close to linear problems with small observation noise.
- For ODE's its easy to prove that projection gives "closest" approximation.
- Is the Stratonovich projection the "closest" approximation on M?

Differential geometry 101 - Charts



Differential geometry 101 - Vector Fields



A vector field can be defined as an equivalence class of pairs (chart, vector field on $\mathbb{R}^n$)

Definition of vector fields

- Vector field is equivalence class (ϕ, X) where ϕ is a chart and X is the vector field on $\mathbb{R}^r$.
- We must choose the equivalence class so that the solutions of one ODE are mapped to the solutions of the other ODE by the transition functions.
- So by the chain rule, the correct definition is:

$$(\phi_1, X) \sim (\phi_2, Y)$$

if

$$\begin{aligned} X^i &= \sum_j \frac{\partial \tau^i}{\partial x^j} Y^j \\ &= (\partial_j \tau^i) Y^j \end{aligned}$$

where we're using the Einstein summation convention.

Stochastic differential equations manifolds

- Define an SDE on a manifold as an equivalence classes of

$$(W^t, \phi, a, b)$$

in such a way that the solutions of one SDE:

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

are mapped to the solutions of the other by the transition functions.

- So by Ito's lemma, the correct definition is:

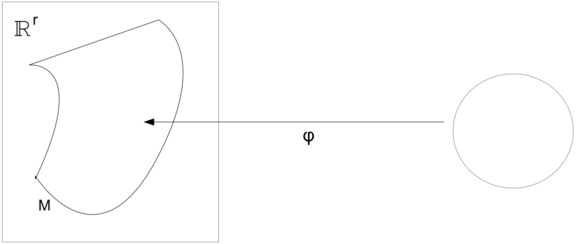
$$(W_t, \phi, a, b) \sim (V_t, \Phi, A, B) \;\text{ if}$$

$$\begin{cases} W_t = V_t \\ A^j = a^i \partial_i \tau^j + \dfrac{1}{2} b^i_\alpha b^k_\beta [W^\alpha, W^\beta]_t \partial_i \partial_k \tau^j \\ B^j_\alpha = b^i_\alpha \partial_j \tau^i \end{cases}$$

where we're using the Einstein summation convention.

Stratonovich approach

- You can use Stratonovich SDE's if you prefer, but your definition of an SDE will be essentially equivalent.
- It is not true that you have to use Stratonovich calculus on manifolds when using the intrinsic approach (i.e. charts)
- Stratonovich calculus allows a crisper definition in the intrinsic approach, a Strat SDE has vector fields as coefficients.
- If you use the extrinsic approach, Stratonovich calculus is intuitive because:
  - An SDE on M is an SDE on $\mathbb{R}^r$ whose solutions starting from a point in M stay on M with probability 1.
  - An SDE on M is an SDE on $\mathbb{R}^r$ whose Strat coefficients at a point $x \in M$ lie in the tangent space $T_x M$.



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:
$\mathrm{d}Y = A(Y,t)\,\mathrm{d}t + b(Y,t)\,\mathrm{d}W_t$

Ito Taylor series estimates:

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

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

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 *Ito projection*
- It is different from the Stratonovich projection

Discussion

- Have we found the "right" estimates to optimize?
- We have two estimates:
  - Estimate one is on the expectation of the absolute value. This determines the martingale part of our equation
  - Estimate two is on the absolute value of the expectation. This determines the bounded variation part of our equation
  - Estimate one determines the short term behaviour
  - Estimate two determines the long term behaviour
- Conjecture that Stratonovich projection arises from estimating errors in

$$(X_t - X_{-t}) - (\phi(Y_t) - \phi(Y_{-t}))$$

  i.e. Stratonovich projection is time symmetric.

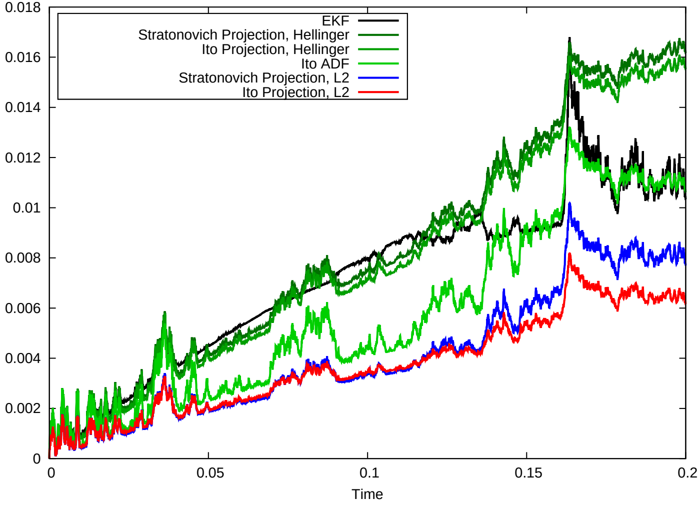
Numerical experiments

Engineers have been using Gaussian approximations to non-linear filters for decades

- Extended Kalman Filter (based on linearization)
- Ito Assumed Density Filter (based on heuristic moment matching arguments)
- Stratonovich Assumed Density Filter
- Stratonovich Projection Filter

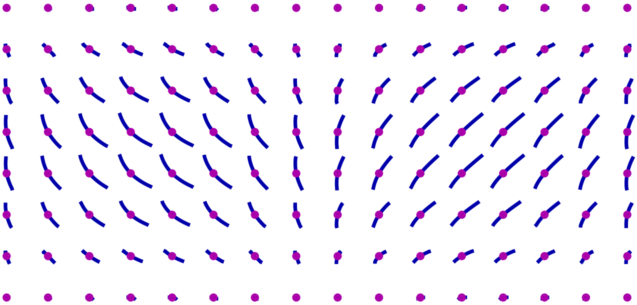
We expect the Ito projection to filter to outperform these filters. At least for short time Ito projection filter should be optimal.

Residuals for cubic sensor, L^2 metric



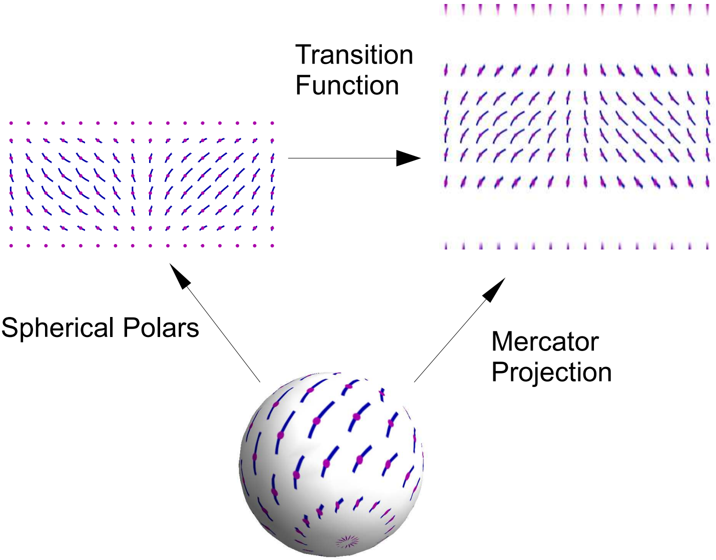
How to draw Ito SDE's with 1-d noise

$$\begin{aligned} & \mathrm{d}X = a\,\mathrm{d}t + b\,\mathrm{d}W_t \\ \Leftrightarrow\quad & \mathrm{d}X = a\,(\mathrm{d}W_t)^2 + b\,\mathrm{d}W_t \\ \Rightarrow\quad & \delta X \approx a(\delta W_t)^2 + b\,\delta W_t \end{aligned}$$



The coefficients of an SDE can be thought of as the *2-jet* of a path $\gamma : \mathbb{R} \to M$.

2-jets of paths obey Ito's lemma



Remark: Ito's lemma

Associate a path γ_x starting at x with every point $x \in \mathbb{R}^r$.
Consider numerical scheme:

$$\begin{aligned} \delta X &= \gamma_x(\delta W_t) \\ &= b\,\delta W_t + a\,\delta W_t^2 + O(\delta W_t^3) \\ &= b\,\delta W_t + a\,\delta t + O(\delta W_t^3) \end{aligned}$$

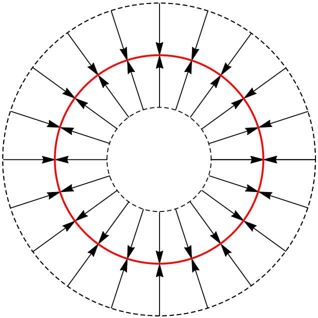
- In the limit as $\delta t \to 0$ we obtain SDE.
- Conclusion: 2-jet of path at every point $\equiv$ SDE
- Consider $g : \mathbb{R}^r \to R^s$. $g \circ \gamma$ is a path at every point in $\mathbb{R}^r$. Induced SDE is SDE for $g(X)$.
- Therefore transformation law of SDE = Transformation law of 2-jets.
- Ito's lemma can be interpreted as the transformation law of 2-jets of paths.

Coordinate free definition of SDE on a manifold

- One can define an Ito SDE in terms of 2-jets of paths
- Very clean in case of 1-d noise as we've seen
- Some redundancy for higher dimensional noise since $(\mathrm{d}W_t^1)^2 = (\mathrm{d}W_t^2)^2 = \mathrm{d}t$.
- Working with Ito formulation of SDE's on manifolds has numerous advantages (e.g. Taylor series, Martingale properties etc.). 2-jets allow this formulation to be handled in a coordinate free manner.
- There is no need to use Stratonovich formulation of SDE's just because one wishes to use coordinate free formulations.

Intrinsic projection

Let π be smooth projection defined on a tubular neighbourhood of M:



- Consider 2-jets of paths $\gamma_x : \mathbb{R} \to \mathbb{R}^r$ that define the SDE on $\mathbb{R}^r$
- At a point $x \in M$ the map $\pi \circ \gamma_x$ defines the *instrinsic* Ito projection

Conclusions

- Ito projection gives the optimal lower dimensional approximation to an SDE over short time horizons.
- Numerical experiments confirm that Ito projection outperforms known approximation methods over short time horizons.
- Stratonovich projection lacks such a convincing optimality property, but in practice it is close to the Ito projection so still performs well.
- Only shown results for L^2 projection onto Gaussian family. But projecting onto manifolds has been shown to be effective for a number of much more interesting statistical families and it generalizes the Galerkin method.