Stochastic Filtering by Projection

The Example of the Quadratic Sensor

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

GSI2013

Motivation

Estimate the current state of a stochastic system from imperfect measurements

Motivation

Estimate the current state of a stochastic system from imperfect measurements

- Estimate the position of a car

Motivation

Estimate the current state of a stochastic system from imperfect measurements

- Estimate the position of a car
- Estimate the volatility of a stock from option prices

Motivation

Estimate the current state of a stochastic system from imperfect measurements

- Estimate the position of a car
- Estimate the volatility of a stock from option prices
- Applications in weather forecasting, oil extraction ...

Motivation

Estimate the current state of a stochastic system from imperfect measurements

- Estimate the position of a car
- Estimate the volatility of a stock from option prices
- Applications in weather forecasting, oil extraction ...

The calculation should be performed online.

Mathematical formulation

$$\begin{aligned} dX_t &= f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \\ dY_t &= b_t(X_t)\,dt + dV_t, \quad Y_0 = 0\ . \end{aligned}$$

- X_t is a process representing the state.
- Y_t is a process representing the measurement.
- W_t and V_t are independent Wiener processes.

Mathematical formulation

$$\begin{aligned} dX_t &= f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \\ dY_t &= b_t(X_t)\,dt + dV_t, \quad Y_0 = 0\ . \end{aligned}$$

- X_t is a process representing the state.
- Y_t is a process representing the measurement.
- W_t and V_t are independent Wiener processes.

Question

What is the probability distribution for X_t given the values of Y_t up to time t?

The Kushner–Stratonovich equation

With sufficient regularity and bounds, one can show that the probability density p_t satisfies:

$$\mathrm{d}p_t = \mathcal{L}_t^* p_t \mathrm{d}t + p_t[b_t - E_{p_t}\{b_t\}][\mathrm{d}Y_t - E_{p_t}\{b_t\}\mathrm{d}t] \ .$$

where:

- $$\mathcal{L}^* = -f_t \frac{\partial}{\partial x} + \frac{1}{2} a_t \frac{\partial}{\partial x^2}$$
  is the backward diffusion operator
- $a_t^T a = \sigma$ and a is a square root of σ.
- E_{p_t} denotes expectation with respect to p_t.

The Kushner–Stratonovich equation

With sufficient regularity and bounds, one can show that the probability density p_t satisfies:

$$\mathrm{d}p_t = \mathcal{L}_t^* p_t \mathrm{d}t + p_t[b_t - E_{p_t}\{b_t\}][\mathrm{d}Y_t - E_{p_t}\{b_t\}\mathrm{d}t] \; .$$

where:

- $$\mathcal{L}^* = -f_t \frac{\partial}{\partial x} + \frac{1}{2} a_t \frac{\partial}{\partial x^2}$$
  is the backward diffusion operator
- $a_t^T a = \sigma$ and a is a square root of σ.
- E_{p_t} denotes expectation with respect to p_t.

Question

How can we efficiently approximate solutions to the infinite dimensional Kushner–Stratonovich equation?

The geometric idea

- Choose a submanifold of the space of probability distributions so that points in the manifold can approximate p_t well.

The geometric idea

- Choose a submanifold of the space of probability distributions so that points in the manifold can approximate p_t well.
- View the partial differential equation as defining a stochastic vector field.

The geometric idea

- Choose a submanifold of the space of probability distributions so that points in the manifold can approximate p_t well.
- View the partial differential equation as defining a stochastic vector field.
- Use projection to restrict the vector field to the tangent space.

The geometric idea

- Choose a submanifold of the space of probability distributions so that points in the manifold can approximate p_t well.
- View the partial differential equation as defining a stochastic vector field.
- Use projection to restrict the vector field to the tangent space.
- Solve the resulting finite dimensional stochastic differential equation.

The linear problem

If:

- the coefficient functions a, b and f in the problem are all linear
- p_0, which represents the prior probability distribution for the state, is a Gaussian

then

The linear problem

If:

- the coefficient functions a, b and f in the problem are all linear
- p_0, which represents the prior probability distribution for the state, is a Gaussian

then

- p_t is always a Gaussian
- The mean and standard deviation of p_t follow a finite dimensional SDE.

This is called the *Kalman filter*.

The linear problem

If:

- the coefficient functions a, b and f in the problem are all linear
- p_0, which represents the prior probability distribution for the state, is a Gaussian

then

- p_t is always a Gaussian
- The mean and standard deviation of p_t follow a finite dimensional SDE.

This is called the *Kalman filter*.
One can linearize any filtering problem at each point in time to obtain the *Extended Kalman filter*.

Two important families

For multi modal problems, project onto one of the following families:

Two important families

For multi modal problems, project onto one of the following families:

- A mixture of m Gaussian distributions:

$$p_t(x) = \sum_i \lambda_i e^{(x-\mu_i)/2\sigma_i^2}$$

  - $\lambda_i \geq 0$. $\sum_i \lambda_i = 1$.
  - Gives rise to a $3m - 1$ dimensional family.

Two important families

For multi modal problems, project onto one of the following families:

- A mixture of m Gaussian distributions:

$$p_t(x) = \sum_i \lambda_i e^{(x-\mu_i)/2\sigma_i^2}$$

  - $\lambda_i \geq 0$. $\sum_i \lambda_i = 1$.
  - Gives rise to a $3m-1$ dimensional family.
- The exponential family

$$p_t(x) = \exp(a_0 + a_1 x + a_2 x^2 + \dots a_{2n} x^{2n})$$

  - $a_{2n} < 0$
  - Gives rise to a $2n$ dimensional family.

Choice of metric for the projection

Need to choose a Hilbert space structure on the space of probability distributions (more precisely some enveloping space).

Choice of metric for the projection

Need to choose a Hilbert space structure on the space of probability distributions (more precisely some enveloping space).

- The Hellinger metric.
  - Theoretical advantage of coordinate independence
  - Works well with exponential families (Brigo)
  - Meaningful for problems where density p does not exist.
  - Requires numerical approximation of integrals to implement.

Choice of metric for the projection

Need to choose a Hilbert space structure on the space of probability distributions (more precisely some enveloping space).

- The Hellinger metric.
  - Theoretical advantage of coordinate independence
  - Works well with exponential families (Brigo)
  - Meaningful for problems where density p does not exist.
  - Requires numerical approximation of integrals to implement.
- The direct L^2 metric.
  - Works well with mixture families.
  - All integrals that occur can be calculated analytically.

Understanding stochastic differential equations

A stochastic differential equation such as:

$$dX_t = f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t$$

is shorthand for an integral equation such as:

$$X_T = \int_0^T f_t(X_t)\,dt + \int_0^T \sigma_t(X_t)\,dW_t$$

where the right hand integral is defined by the Ito integral:

$$\int_0^T f(t)\,dW_t = \lim_{n\to\infty} \sum_{i=1}^{\infty} f(t_i)(W_{t_{i+1}} - W_{t_i}).$$

The Stratonovich integral

- Take the Ito integral:

$$\int_0^T f(t)\, dW_t = \lim_{n\to\infty} \sum_{i=1}^{\infty} f(t_i)(W_{t_{i+1}} - W_{t_i}).$$

and change the point where you evaluate the integrand

$$\int_0^T f(t) \circ\, dW_t = \lim_{n\to\infty} \sum_{i=1}^{\infty} f(\frac{t_i + t_{i+1}}{2})(W_{t_{i+1}} - W_{t_i}).$$

to get the *Stratonvich integral*. Hence you can define Stratonovich SDE's.

The Stratonovich integral

- Take the Ito integral:

$$\int_0^T f(t)\, dW_t = \lim_{n\to\infty} \sum_{i=1}^{\infty} f(t_i)(W_{t_{i+1}} - W_{t_i}).$$

  and change the point where you evaluate the integrand

$$\int_0^T f(t) \circ dW_t = \lim_{n\to\infty} \sum_{i=1}^{\infty} f(\frac{t_i + t_{i+1}}{2})(W_{t_{i+1}} - W_{t_i}).$$

  to get the *Stratonvich integral*. Hence you can define Stratonovich SDE's.
- The difference between the two integrals is an ordinary integral. This allows you to convert between the two formulations.
- Ito SDE's model causality more naturally
- Stratonovich SDE's transform like vector fields.

A recipe for projecting SDE's

To project an SDE onto a submanifold parameterized by $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$:

- Write the SDE as an SDE with vector coefficients in *Stratonovich form*.
- Project all the coefficients onto the tangent space.
- Equate both sides of the projected equations to get an SDE for the θ_i.

A recipe for projecting SDE's

To project an SDE onto a submanifold parameterized by $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$:

- Write the SDE as an SDE with vector coefficients in *Stratonovich form*.
- Project all the coefficients onto the tangent space.
- Equate both sides of the projected equations to get an SDE for the θ_i.

Since Stratonovich SDE's transform like vector fields, this recipe is invariant of the parameterization.

The projected equations

The end result for the case of L^2 projection is:

$$\mathrm{d}\theta^i = \sum_{j=1}^{m} h^{ij} \left\{ \langle p(\theta), \mathcal{L} v_j \rangle \mathrm{d}t - \langle \gamma^0(p(\theta)), v_j \rangle \mathrm{d}t + \langle \gamma^1(p(\theta)), v_j \rangle \circ \mathrm{d}Y \right\}.$$

Where:

- The $v_j = \frac{\partial p}{\partial \theta_j}$ give a basis for the tangent space
- h_{ij} and h^{ij} are the Riemannian metric tensor $\langle v_i, v_j \rangle$.
- $\gamma^0_t(p) := \frac{1}{2} [|b_t|^2 - E_p\{|b_t|^2\}]$
- $\gamma^1_t(p) := [b_t - E_p\{b_t\}]p$
- $\langle \cdot, \cdot \rangle$ is the L^2 inner product.

Note that the inner products and expectations give rise to integrals. We can compute these analytically for the normal mixture family.

Solving the finite system of SDE's

- Approximate the differential equation as a difference equation and solve numerically.
- This is more delicate for stochastic equations than ordinary ones. See Kloeden and Platen. We use the Stratonovich–Heun cheme.
- Note that the resulting difference equation will depend upon the choice of parameterization of the submanifold. Choose coordinates $\phi : \mathbb{R}^n \longrightarrow \mathcal{M}$ so that ϕ is defined on all of $\mathbb{R}^n$.

The quadratic sensor

$$\mathrm{d}X_t = \mathrm{d}W_t$$
$$\mathrm{d}Y_t = X^2 + \mathrm{d}V_t \ .$$

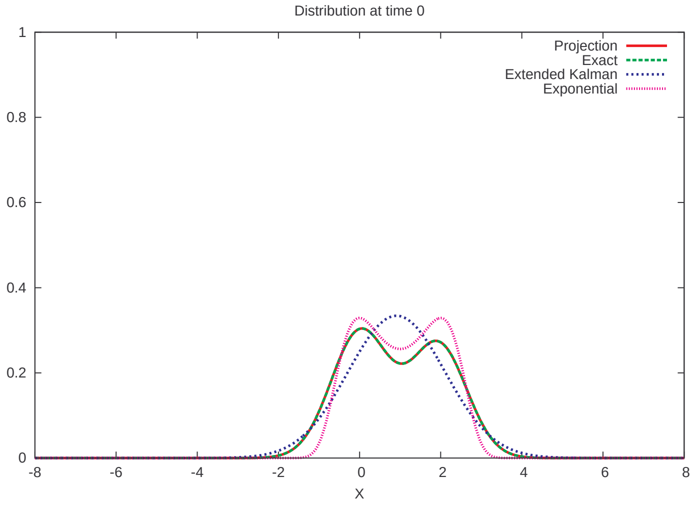
The quadratic sensor

$$\mathrm{d}X_t = \mathrm{d}W_t$$

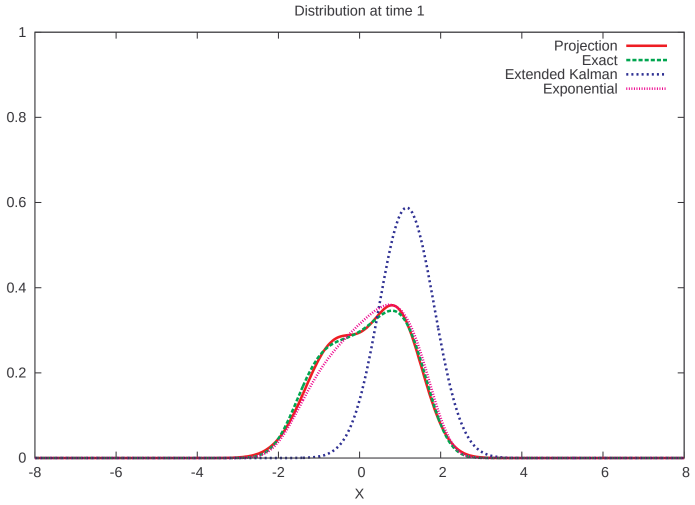
$$\mathrm{d}Y_t = X^2 + \mathrm{d}V_t \ .$$

- We do not receive any information on the sign of X.
- We expect that once X has hit the origin, p will be approximately symmetrical.
- We expect a bimodal distribution

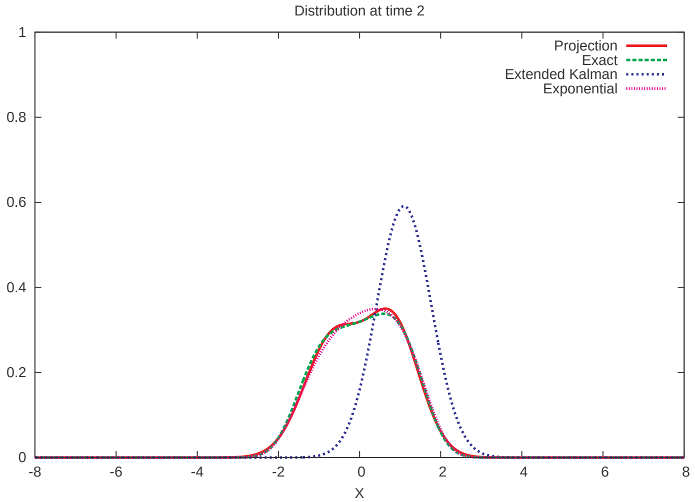
Simulation for the Quadratic Sensor



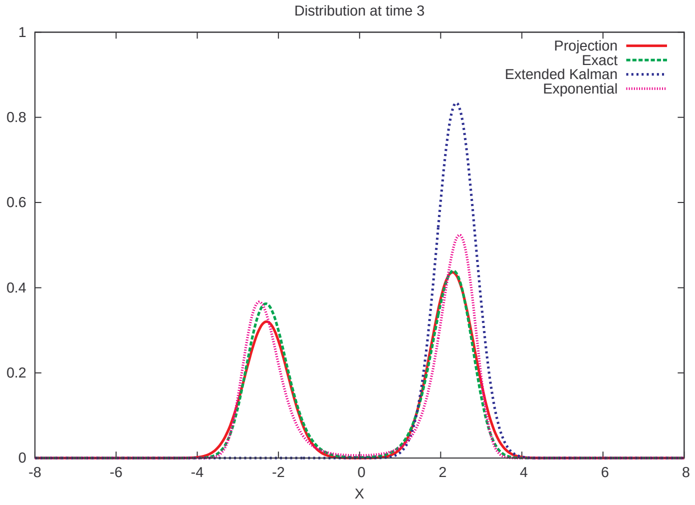
Simulation for the Quadratic Sensor



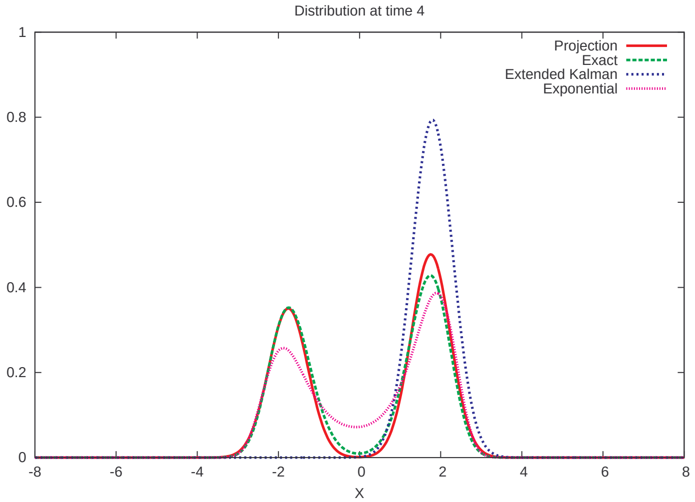
Simulation for the Quadratic Sensor



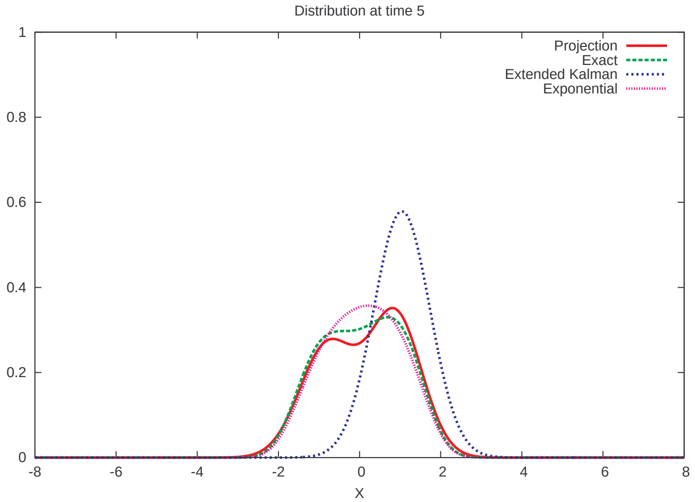
Simulation for the Quadratic Sensor



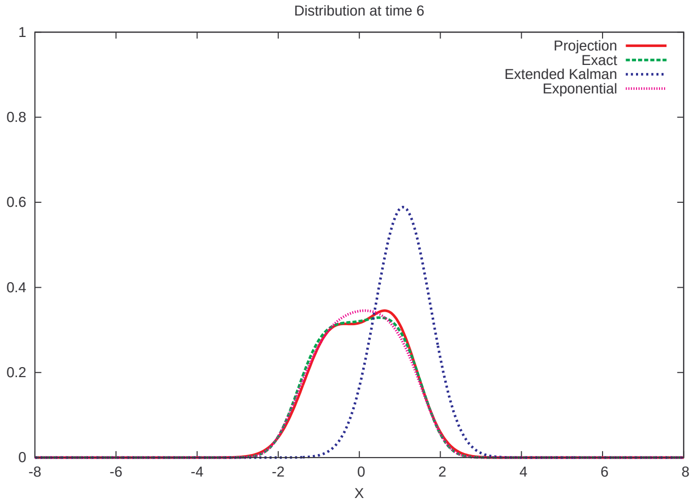
Simulation for the Quadratic Sensor



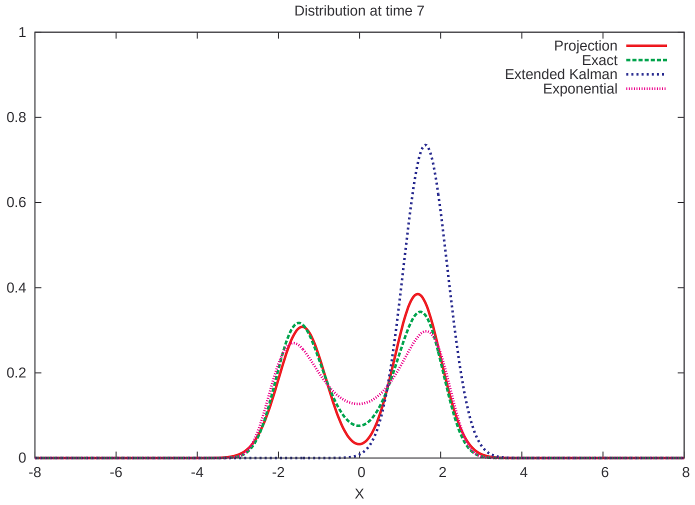
Simulation for the Quadratic Sensor



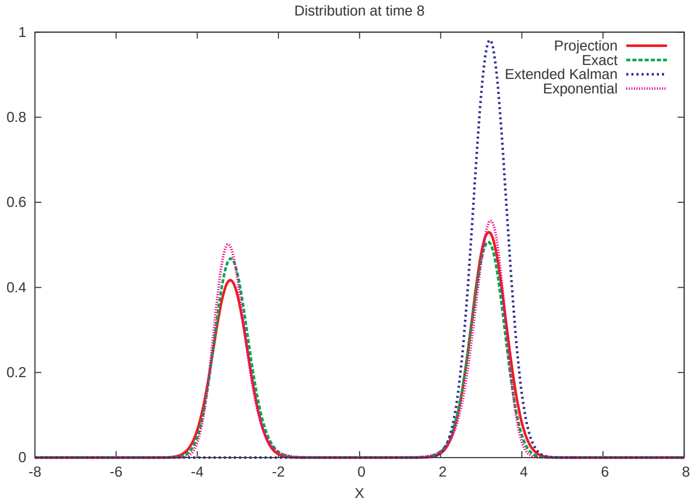
Simulation for the Quadratic Sensor



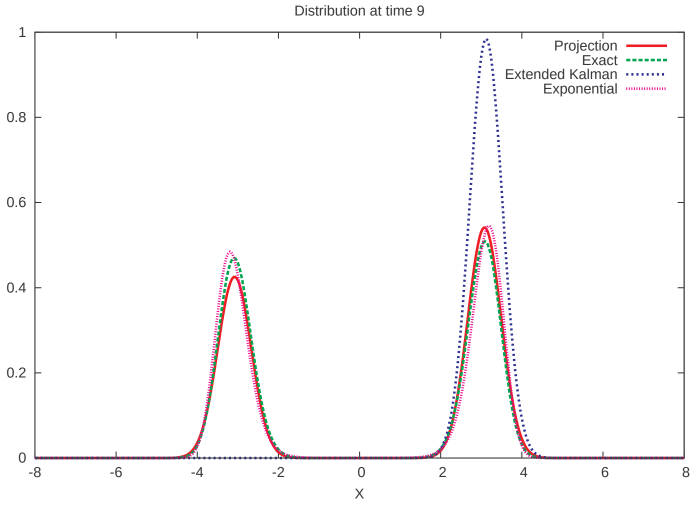
Simulation for the Quadratic Sensor



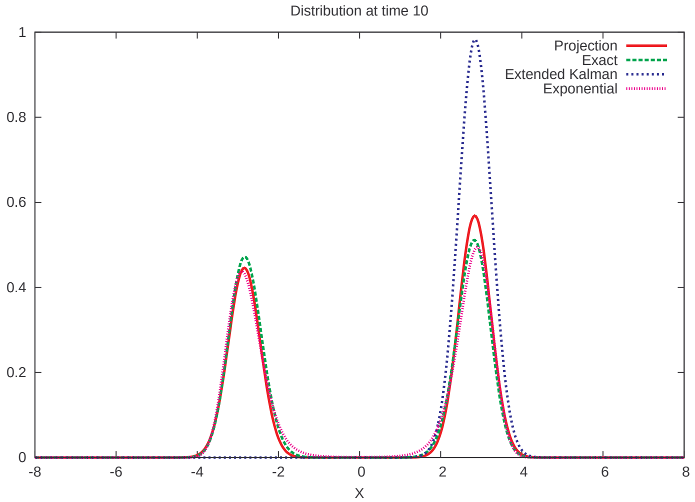
Simulation for the Quadratic Sensor



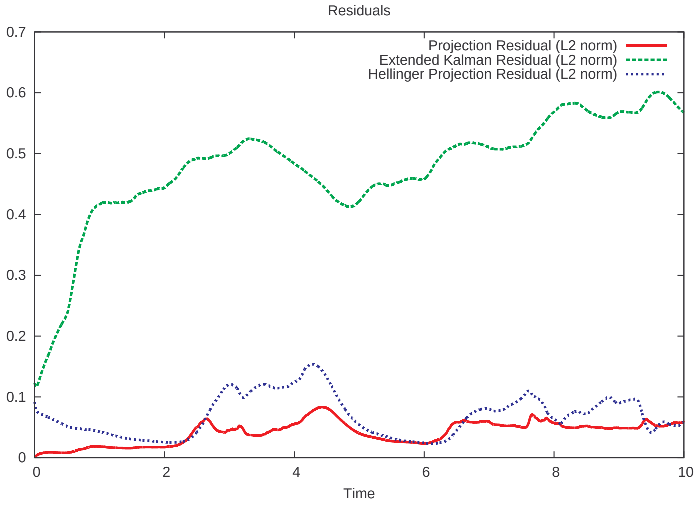
Simulation for the Quadratic Sensor



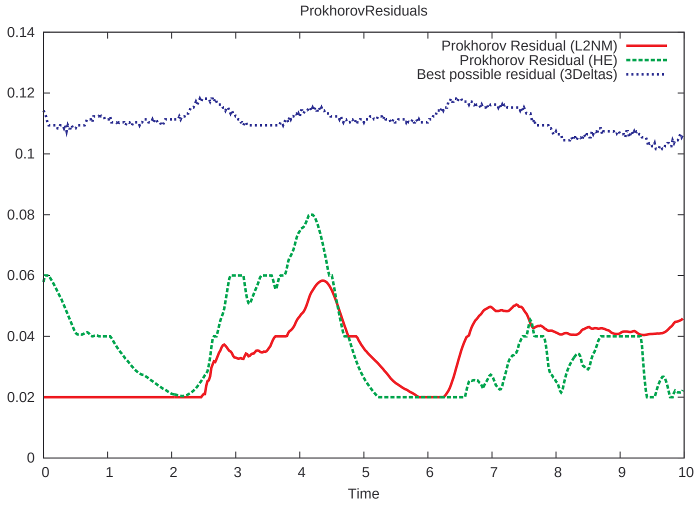
Simulation for the Quadratic Sensor



L^2 residuals for the quadratic sensor



Lévy residuals for the quadratic sensor



Conclusions

- Projection methods allow us to approximate the solution to nonlinear problems with surprising accuracy using only low dimensional manifolds.
- This conclusion holds for a variety of projection metrics and manifolds.
- L^2 projection of normal mixtures is particularly promising since all integrals can be computed analytically.