Stochastic Filtering through SPDE projection on Mixtures Manifolds Imperial, AHOI Workshop, 25-27 March 2013

Damiano Brigo
Dept. of Mathematics, Imperial College, London
www.damianobrigo.it

—

John Armstrong
Dept. of Mathematics, King's College, London

—

Full paper available on HAL
`http://hal.archives-ouvertes.fr/hal-00801246` *or damianobrigo.it*

Agenda

Spaces of probability densities I

Consider a parametric family of probability densities

$$S = \{p(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^m\}.$$

let $S^{1/2}$ be the space of square roots

$$S^{1/2} = \{\sqrt{p(\cdot, \theta)}, \theta \in \Theta \subset \mathbb{R}^m\}.$$

If S (or $S^{1/2}$ respectively) is a subset of a function space having an L^2 structure (and hence an inner product, a norm and a metric), then we may ask whether the map

$$p(\cdot, \theta) \mapsto \theta, \quad (\sqrt{p(\cdot, \theta)} \mapsto \theta \ \text{ respectively})$$

taking values in $\mathbb{R}^m$ is a Chart of a m-dimensional manifold (?) S ($S^{1/2}$).

Spaces of probability densities II

The topology and differential structure we should consider in the chart is the topology in L^2, but depending on whether we take S or $S^{1/2}$ the topology and differential structures are different.

$$S : \quad d_2(p_1, p_2) = \|p_1 - p_2, \|$$

where $\| \cdot \|$ denotes the norm of the Hilbert space L^2 (L2 direct distance).

$$S^{1/2} : \quad d_H(\sqrt{p_1}, \sqrt{p_2}) = \|\sqrt{p_1} - \sqrt{p_2}, \|$$

where $\| \cdot \|$ denotes the norm of the Hilbert space L^2 (Hellinger distance).
For the first definition, more restrictive, we need to assume densities in S to be square integrable. This is the case for example if they are integrable and bounded.

Spaces of probability densities III

S and $S^{1/2}$: Not submanifolds of L2 manifolds

Despite being subsets of L^2, neither the set of square roots of all densities nor the set of all densities themselves are locally homeomorphic to L^2, hence they are not infinite dimensional manifolds modeled on L^2. Indeed, any open set of L^2 contains functions which are negative in a set with positive Lebesgue measure. There is no open set of L^2 which contains only positive functions such as the functions of the space of all densities or their square roots.

S and $S^{1/2}$: Not submanifolds of L2 vector subspaces

Furthermore, such spaces are not vector spaces, and hence while they have a metric coming from the L^2 norm, they cannot be equipped with a norm themselves, and are not normed spaces.

Tangent vectors, metrics and projection I

$$\text{If} \quad \varphi : \theta \mapsto p(\cdot,\theta), \quad (\theta \mapsto \sqrt{p(\cdot,\theta)} \ \text{ respectively})$$

is the inverse of the chart (we work only with the single coordinate chart (S, φ^{-1}) and $(S^{1/2}, \varphi^{-1})$ respectively) then

$$\{\frac{\partial \varphi(\cdot,\theta)}{\partial \theta_1}, \cdots, \frac{\partial \varphi(\cdot,\theta)}{\partial \theta_m}\}$$

is a set of linearly independent vectors in $L^2(\lambda)$. Then, according to the chain rule, we compute the following Fréchet derivatives:

$$Dp(,\theta(h))|_{h=0} = \sum_{k=1}^{m} \frac{\partial p(,\theta)}{\partial \theta_k} \ \dot{\theta}_k(0)$$

$$\left(D\sqrt{p(,\theta(h))}|_{h=0} = \sum_{k=1}^{m} \frac{\partial \sqrt{p(,\theta)}}{\partial \theta_k} \ \dot{\theta}_k(0) \quad \text{respectively} \right)$$

Tangent vectors, metrics and projection II

We obtain that a basis for the tangent vector space at $p(\cdot,\theta)$ ($\sqrt{p(\cdot,\theta)}$ respectively) to the space S ($S^{1/2}$) is:

$$T_{p(\cdot,\theta)}S = \text{span}\left\{\frac{\partial p(\cdot,\theta)}{\partial\theta_1},\cdots,\frac{\partial p(\cdot,\theta)}{\partial\theta_m}\right\}\ . \tag{1}$$

$$\left(T_{\sqrt{p(\cdot,\theta)}}S^{1/2} = \text{span}\left\{\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_1},\cdots,\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_m}\right\}\right). \tag{2}$$

If φ is the inverse of a chart, these vectors are actually linearly independent, and they indeed form a basis of the tangent vector space. One has to be careful, because if this were not true, the dimension of the above spanned space could drop.

Tangent vectors, metrics and projection III

The inner product of any two basis elements is defined, according to the L^2 inner product

$$\langle \frac{\partial p(\cdot,\theta)}{\partial \theta_i} \frac{\partial p(\cdot,\theta)}{\partial \theta_j} \rangle = \frac{1}{4} \int \frac{\partial p(x,\theta)}{\partial \theta_i} \frac{\partial p(x,\theta)}{\partial \theta_j} dx = \frac{1}{4} \gamma_{ij}(\theta) \; .$$

$$\left(\langle \frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_i} \frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_j} \rangle = \frac{1}{4} \int \frac{1}{p(x,\theta)} \frac{\partial p(x,\theta)}{\partial \theta_i} \frac{\partial p(x,\theta)}{\partial \theta_j} dx = \frac{1}{4} g_{ij}(\theta) \right) .$$

- $\gamma(\theta)$: direct L2 matrix, associated to the metric d_2;
- $g(\theta)$: the famous Fisher information metric associated with d_H, see for example [1], [2] and [29]

Tangent vectors, metrics and projection IV

The two different matrices define different metrics and differential structures:

$$d_2(p(\cdot,\theta), p(\cdot,\theta+d\theta))^2 = (d\theta)^T\ \gamma(\theta)\ (d\theta),$$

$$d_H(\sqrt{p(\cdot,\theta)}, \sqrt{p(\cdot,\theta+d\theta)})^2 = (d\theta)^T\ g(\theta)\ (d\theta),$$

Tangent vectors, metrics and projection V

Example: The Normal family in canonical param θ and $g(\theta)$

The Gaussian family may be defined as a particular exponential family, represented with canonical parameters θ, given by

$$\{p(x,\theta) = \exp(\theta_1 x + \theta_2 x^2 - \psi(\theta)),\ \theta_2 < 0\}$$

where one has easily $\psi(\theta) = \frac{1}{2}\ln\left(\frac{\pi}{-\theta_2}\right) - \frac{\theta_1^2}{4\theta_2}$ and the Fisher metric is

$$g(\theta) = \begin{bmatrix} -1/(2\theta_2) & \theta_1/(2\theta_2^2) \\ \theta_1/(2\theta_2^2) & 1/(2\theta_2^2) - \theta_1^2/(2\theta_2^3) \end{bmatrix}$$

The familiar representation of Gaussian densities is in terms of mean and variance, given respectively by

$$\mu = -\theta_1/(2\theta_2), \quad v = \sigma^2 = (1/\theta_2 - \theta_1^2/\theta_2^2)/2$$

Tangent vectors, metrics and projection VI

Example: The Normal family in expectation param μ, v and $g(\mu, v)$

We may consider the Fisher metric for the Gaussian family of densities in the parameters μ and v. These are related to the so called expectation parameters μ and $v + \mu^2$. With this coordinate system the Fisher metric is much simpler and the matrix is diagonal, resulting in

$$g(\mu, v) = \frac{1}{v} \left[\begin{array}{cc} 1 & 0 \\ 0 & 1/(2v) \end{array} \right]$$

This can be derived either by applying the change of coordinates formula, or from the metrics Eq. directly, with the parameters θ_1, θ_2 replaced by μ, v.

Tangent vectors, metrics and projection VII

Example: The Normal family in canonical param θ and $\gamma(\theta)$

$$\gamma(\theta) = \frac{1}{8}\frac{\sqrt{2}}{\sqrt{-\theta_2 \pi}}\begin{bmatrix} 1 & \frac{\theta_1}{-\theta_2} \\ \frac{\theta_1}{-\theta_2} & \frac{3}{4}\frac{1}{(-\theta_2)} + \frac{{\theta_1}^2}{{\theta_2}^2} \end{bmatrix}$$

and, as expected, it is different from the Fisher metric seen earlier.

Example: The Normal family in expectation param μ, v and $\gamma(\mu, v)$

We may consider the L^2 metric for the Gaussian family in the coordinates μ, v. The L^2 metric is

$$\gamma(\mu, v) = \frac{1}{8v\sqrt{v\pi}}\begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{4v} \end{bmatrix}$$

and it is different from $g(\mu, v)$, although it is still a diagonal matrix.

Projections

Next, we introduce the orthogonal projection between any linear subspace V of L^2 containing our finite dimensional tangent vector space and the tangent vector space itself.

$$\Pi_\theta^\gamma \ : L^2(\lambda) \supseteq V \longrightarrow \text{span}\{\frac{\partial p(\cdot,\theta)}{\partial \theta_1}, \cdots, \frac{\partial p(\cdot,\theta)}{\partial \theta_m}\}$$

$$\Pi_\theta^\gamma[v] = \sum_{i=1}^m [\sum_{j=1}^m h^{ij}(\theta) \ \langle v, \frac{\partial p(\cdot,\theta)}{\partial \theta_j}\rangle] \ \frac{\partial p(\cdot,\theta)}{\partial \theta_i} \ .$$

$$\left(\begin{array}{l} \Pi_\theta^g \ : L^2(\lambda) \supseteq V \longrightarrow \text{span}\{\frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_1}, \cdots, \frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_m}\} \\ \\ \Pi_\theta^g[v] = \sum_{i=1}^m [\sum_{j=1}^m 4 g^{ij}(\theta) \ \langle v, \frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_j}\rangle] \ \frac{\partial \sqrt{p(\cdot,\theta)}}{\partial \theta_i} \ . \end{array} \right)$$

The nonlinear filtering problem for diffusion signals I

$$\begin{aligned} dX_t &= f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \quad \text{(signal)} \\ dY_t &= b_t(X_t)\,dt + dV_t, \quad Y_0 = 0 \ \ \text{(noisy observation)} \end{aligned} \tag{3}$$

These are Itô SDE's. We shall use both Itô and Stratonovich (Str) SDE's. Str SDE's are necessar to deal with stochastic calculus on manifolds, since second order Itô terms not clear in terms of manifolds [18]).

Nonlinear filtering problem

The nonlinear filtering problem consists in finding the conditional probability distribution π_t of the state X_t given the observations up to time t, i.e. $\pi_t(dx) := P[X_t \in dx \mid \mathcal{Y}_t]$, where $\mathcal{Y}_t := \sigma(Y_s\,, 0 \leq s \leq t)$. We assume that for all $t \geq 0$, the probability distribution π_t has a density p_t w.r.t. the Lebesgue measure.

The nonlinear filtering problem for diffusion signals II

Then $\{p_t\,,\ t \geq 0\}$ satisfies the Str SPDE:

$$dp_t = \mathcal{L}_t^*\, p_t\, dt - \frac{1}{2}\, p_t\, [|b_t|^2 - E_{p_t}\{|b_t|^2\}]\, dt + \sum_{k=1}^{d} p_t\, [b_t^k - E_{p_t}\{b_t^k\}] \circ dY_t^k\,.$$

with the forward diffusion operator $\mathcal{L}_t^*$ defined by

$$\mathcal{L}_t^* \phi = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\, [f_t^i\, \phi] + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\, [a_t^{ij}\, \phi]$$

Projection filter in the metrics h (L2) and g (Fisher) I

This equation can be projected according to either the L2 direct metric (leading to $\gamma(\theta)$) or, by deriving the analogous equation for $\sqrt{p_t}$, according to the Hellinger metric (leading to the Fisher metric $g(\theta)$). First let us project with $\gamma(\theta)$.

$$dp(\cdot,\theta_t) = \Pi^{\gamma}_{\theta_t}\left[\mathcal{L}^*_t\, p(\cdot,\theta_t) \;-\frac{1}{2}\, p(\cdot,\theta_t)\, [|b_t|^2 - E_{p(\cdot,\theta_t)}\{|b_t|^2\}]\right] dt + \\ + \sum_{k=1}^{d} \Pi^{\gamma}_{\theta_t}\left[\sum_{k=1}^{d} p(\cdot,\theta_t)\, [b^k_t - E_{p(\cdot,\theta_t)}\{b^k_t\}]\right] \circ dY^k_t \quad (4)$$

OTOH the lhs can be written with the chain rule

$$dp(\cdot,\theta_t) = \sum_{j=1}^{m} \frac{\partial p(\cdot,\theta)}{\partial \theta_j} \circ d\theta_j(t)$$

Projection filter in the metrics h (L2) and g (Fisher) II

so that we obtain after straightforward calculations (Projection Filter in L2 direct metric $\gamma(\theta)$)

$$
\begin{aligned}
d\theta_t^i \;=\; & [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k \; , \quad \theta_0^i \; .
\end{aligned}
$$

Projection filter in the metrics h (L2) and g (Fisher) III

Instead, using the Hellinger distance and the Fisher metric with projection Π^g we obtain

$$\begin{aligned} d\theta_t^i \; = \; & [\sum_{j=1}^m g^{ij}(\theta_t) \int \frac{\mathcal{L}_t^* \, p(x,\theta_t)}{p(x,\theta_t)} \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\ & - [\sum_{j=1}^m g^{ij}(\theta_t) \int \frac{1}{2} \, |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\ & + \sum_{k=1}^d [\sum_{j=1}^m g^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k \; , \quad \theta_0^i \; . \end{aligned} \tag{5}$$

Notice the differences in the equations stemming from the different metric and the different projection of the diffusion term.

Choosing the family: Exponential I

In past literature and in several papers in Bernoulli, IEEE Automatic Control etc, B. Hanzon and LeGland have developed a theory for the projection filter using the Fisher metric g and

Definition

(Exponential Families) Let $\{c_1, \cdots, c_m\}$ be scalar functions such that $\{1, c_1, \cdots, c_m\}$ are *linearly independent*, and assume the convex set

$$\Theta_0 := \{\theta \in \mathbf{R}^m \ : \ \psi(\theta) = \log \int \exp[\theta^T c(x)]\, dx < \infty\} ,$$

to have *non–empty interior*. Then

$$S = \{p(\cdot, \theta)\, ,\, \theta \in \Theta\}, \qquad p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)] ,$$

where $\Theta \subseteq \Theta_0$ is open, is an exponential family.

Choosing the family: Exponential II

Exponential Families and Hellinger/Fisher $d_H/g(\theta)$ work well together:

- The tangent space has a simple structure: square roots do not complicate issues thanks to the exponential structure.
- The Fisher matrix has a simple structure: $\partial^2_{\theta_i,\theta_j}\psi(\theta) = g_{ij}(\theta)$
- The structure of the projection Π^g is simple for exp families
- Special exp family makes filter correction step exact
- One can define both a local and global fitlering error through d_H
- We have an alternative parameterization in $\eta = E_\theta[c]$, expectation parameters, $\eta = \partial_\theta \psi(\theta)$.
- Projection filter in expectation parameters coincides with classical approximate filter: assumed density filter
- Theory and numerical examples of exponential projection filtering have been developed in [7], [8], [10], [11], [12], [13], [14].

Choosing the family: Exponential III

- However, exponential families do not couple as well with the metric $\gamma(\theta)$. The projection becomes more cumbersome and the filter equations are not as easy.
- *Is there some important family for which the metric $\gamma(\theta)$ is preferable to the classical Fisher metric $g(\theta)$, in that the metric, the tangent space and the filter equations are simpler?*

	Family: Exponential	Family?
Metric:		
Hellinger/Fisher $d_H,\ g(\theta)$	Good	(not good?)
DirectL2 $d_2,\ \gamma(\theta)$	(Not good)	*Good*

- The answer is affirmative, and this is the *mixture family*.

Mixture families I

We define a *simple mixture family* as follows. Suppose we are given $m+1$ fixed squared integrable probability densities, say $\underline{q} = [q_1, q_2, \ldots, q_{m+1}]^T$. Suppose we define the following space of probability densities:

$$S^M(\underline{q}) = \{\theta_1 q_1 + \theta_2 q_2 + \cdots + \theta_m q_m + (1 - \theta_1 - \cdots - \theta_m) q_{m+1}, \theta_i \geq 0 \text{ for all } i,$$

$$\theta_1 + \cdots + \theta_m < 1\}$$

Define the transformation

$$\hat{\theta}(\theta) := [\theta_1, \theta_2, \ldots, \theta_m, 1 - \theta_1 - \theta_2 - \ldots - \theta_m]^T$$

for all θ. We will often write $\hat{\theta}$ instead of $\hat{\theta}(\theta)$.

$$S^M(\underline{q}) = \{\hat{\theta}(\theta)^T \underline{q}, \;\; \theta_i \geq 0 \text{ for all } i, \;\; \theta_1 + \cdots + \theta_m < 1\}$$

Mixture families II

While for exponential families the Hellinger / Fisher is ideal, for mixture families it is not. The calculation of the Fisher information matrix $g(\theta)$ is less immediate, and the related projection is more convoluted.

Instead, if we consider the L^2 / $\gamma(\theta)$ distance, the metric $\gamma(\theta)$ itself and the related projection become very simple. Indeed,

$$\frac{\partial p(\cdot,\theta)}{\partial \theta_i} = q_i - q_{m+1}$$

and

$$\gamma_{ij}(\theta) = \int (q_i(x) - q_m(x))(q_j(x) - q_m(x))dx \quad \text{NO inline numeric integr}$$

Mixture families III

The L^2 metric *does not depend on the specific point* θ *of the manifold*. The same holds for the tangent space at $p(\cdot,\theta)$, which is given by

$$T_{p(\cdot,\theta)}S = \text{span}\{q_1 - q_{m+1}, q_2 - q_{m+1}, \cdots, q_m - q_{m+1}\}$$

Also the L^2 projection becomes particularly simple:

$$\Pi_\theta[v] = \sum_{i=1}^{m}[\sum_{j=1}^{m} h^{ij} \langle v, q_j - q_{m+1}\rangle] (q_i - q_{m+1}) . \qquad (6)$$

Apply the L^2 metric and the related structure to the projection of the infinite dimensional filter onto the mixture family.

Mixture Projection Filter I

- The mixture family + metric $\gamma(\theta)$ lead to a Projection filter that is the same as approximate filtering via Galerkin methods [4].
  `Beard, R. and Gunther, J. (1997). Galerkin Approximations of the Kushner Equation in Nonlinear Estimation. Brigham Young Univ.`
- See the paper in arXiv for the details. Summing up:

	Family:	Exponential	Basic Mixture
Metric:			
Hellinger d_H Fisher $g(\theta)$		Good $\sim$ADF	Not so Good
Direct L2 d_2 $\gamma(\theta)$		Not so Good	Good ($\sim$Galerkin)

Mixture Projection Filter II

- However, despite the simplicity above, the mixture family has an important drawback: for all θ,

$$\min_i \text{mean of } q_i \leq \text{mean of } p(\cdot,\theta) \leq \max_i \text{mean of } q_i$$

- As a consequence of this, we are going to enrich our family to a mixture where some of the parameters are also in the core densities q.
- Specifically, we consider a mixture of GAUSSIAN DENSITIES with MEANS AND VARIANCES in each component not fixed. Means and variances are to be considered as parameters. So for example for a mixture of two Gaussians we have 5 parameters.

$$\theta p_{\mathcal{N}(\mu_1,v_1)}(x) + (1-\theta) p_{\mathcal{N}(\mu_2,v_2)}(x), \quad \text{param.} \quad \theta, \mu_1, v_1, \mu_2, v_2$$

- We are now going to illustrate the Gaussian mixture projection filter (GMPF) in a fundamental example.

Numerical Implementation

The starting point for a numerical implementation is the finite dimensional stochastic SDE given earlier:

$$\begin{aligned} d\theta^i_t \;=\; & [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}^*_t\, p(x,\theta_t)\, \frac{\partial p(x,\theta_t)}{\partial \theta_j}\, dx]\, dt \\ & - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2}|b_t(x)|^2\, \frac{\partial p(x,\theta_t)}{\partial \theta_j}\, dx]\, dt \\ & + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b^k_t(x)\, \frac{\partial p(x,\theta_t)}{\partial \theta_j}\, dx] \circ dY^k_t\ . \end{aligned}$$

Solving a stochastic differential equation

- Writing the SDE more symbollically we have:

$$\gamma \mathrm{d}\theta = H_1(\theta)\mathrm{d}t + H_2(\theta) \circ \mathrm{d}Y$$

Solving a stochastic differential equation

- Writing the SDE more symbollically we have:

$$\gamma \mathrm{d}\theta = H_1(\theta)\mathrm{d}t + H_2(\theta) \circ \mathrm{d}Y$$

- γ is on the right to emphasize that one shouldn't invert γ.

Solving a stochastic differential equation

- Writing the SDE more symbollically we have:

$$\gamma \mathrm{d}\theta = H_1(\theta)\mathrm{d}t + H_2(\theta) \circ \mathrm{d}Y$$

- γ is on the right to emphasize that one shouldn't invert γ.
- So long as we can compute $H_1(\theta)$ and $H_2(\theta)$ and the matrix γ we can solve this using an appropriate numerical scheme.

Solving a stochastic differential equation

- Writing the SDE more symbollically we have:

$$\gamma \mathrm{d}\theta = H_1(\theta)\mathrm{d}t + H_2(\theta) \circ \mathrm{d}Y$$

- γ is on the right to emphasize that one shouldn't invert γ.
- So long as we can compute $H_1(\theta)$ and $H_2(\theta)$ and the matrix γ we can solve this using an appropriate numerical scheme.
- Since this is a Stratonovich equation we use the Euler–Heun scheme.

Computing the coefficients

$$
\begin{aligned}
d\theta_t^i \quad = \quad & [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k
\end{aligned}
$$

Computing the coefficients

$$
\begin{aligned}
d\theta_t^i \;=\;& [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k
\end{aligned}
$$

- A point on our manifold is represented by a density function p.

Computing the coefficients

$$
\begin{aligned}
d\theta_t^i \quad = \quad & [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k
\end{aligned}
$$

- A point on our manifold is represented by a density function p.
- Tangent vectors are represented by the functions $\frac{\partial p}{\partial \theta_i}$

Computing the coefficients

$$
\begin{aligned}
d\theta_t^i \;=\;& [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& - [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \, dt \\
& + \sum_{k=1}^{d} [\sum_{j=1}^{m} \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k
\end{aligned}
$$

- A point on our manifold is represented by a density function p.
- Tangent vectors are represented by the functions $\frac{\partial p}{\partial \theta_i}$
- Once we know these functions, the coefficients can be computed using the following operations on functions: *addition, differentiation, multiplication, integration*.

The FunctionRing abstraction

- Choose an in memory representation of a function.

The FunctionRing abstraction

- Choose an in memory representation of a function.
- Implement code to manipulate these functions.

FunctionRing
Function add(Function x, Function y) Function multiply(Function x, Function y) Function differentiate(Function x) Real integrate(Function x)

Representing the Manifold

- The Manifold maps a set of parameter values θ to a Function in our FunctionRing
- The Manifold can tell us the tangent vectors $\frac{\partial p}{\partial \theta_i}$ at a given point

Representing the Manifold

- The Manifold maps a set of parameter values θ to a Function in our FunctionRing
- The Manifold can tell us the tangent vectors $\frac{\partial p}{\partial \theta_i}$ at a given point

Manifold
Function probabilityDensity(θ) Function[] tangentVectors(θ)

Representing the Manifold

- The Manifold maps a set of parameter values θ to a Function in our FunctionRing
- The Manifold can tell us the tangent vectors $\frac{\partial p}{\partial \theta_i}$ at a given point

Manifold
Function probabilityDensity(θ) Function[] tangentVectors(θ)

- Mathematically the Manifold abstraction corresponds to a parameterization of the manifold
- The results are sensitive to the choice of parameterization
- Our abstraction allows us to change the manifold or parameterization at different time steps.

L2Projection engine

The end result is a general program which given choices of

- FunctionRing
- Manifold
- Coefficients for the filtering equation in the FunctionRing
- Initial state θ
- Measurements Y

is able to compute the evolution of θ.

Example Manifold and FunctionRing

- Data:
  - FunctionRing $\mathcal{R}$ given by linear combinations of polynomials and Gaussians times polynomials.
  - Manifold given by mixtures of Gaussians
  - Coefficients for the filtering equation in the FunctionRing

Example Manifold and FunctionRing

- Data:
  - FunctionRing $\mathcal{R}$ given by linear combinations of polynomials and Gaussians times polynomials.
  - Manifold given by mixtures of Gaussians
  - Coefficients for the filtering equation in the FunctionRing
- In this case all integrals can be performed analytically.

Example Manifold and FunctionRing

- Data:
  - FunctionRing $\mathcal{R}$ given by linear combinations of polynomials and Gaussians times polynomials.
  - Manifold given by mixtures of Gaussians
  - Coefficients for the filtering equation in the FunctionRing
- In this case all integrals can be performed analytically.
- Approximate coefficients outside the function ring using Taylor series.

The quadratic sensor

Consider the quadratic sensor

$$\begin{aligned} \mathrm{d}X_t &= \sigma \mathrm{d}W_t \\ \mathrm{d}Y_t &= X^2 \mathrm{d}t + \sigma \mathrm{d}V_t \ . \end{aligned}$$

The quadratic sensor

Consider the quadratic sensor

$$\begin{aligned} \mathrm{d}X_t &= \sigma \mathrm{d}W_t \\ \mathrm{d}Y_t &= X^2 \mathrm{d}t + \sigma \mathrm{d}V_t \ . \end{aligned}$$

- The measurements tell us nothing about the sign of X

The quadratic sensor

Consider the quadratic sensor

$$\begin{aligned} \mathrm{d}X_t &= \sigma \mathrm{d}W_t \\ \mathrm{d}Y_t &= X^2 \mathrm{d}t + \sigma \mathrm{d}V_t \,. \end{aligned}$$

- The measurements tell us nothing about the sign of X
- Once it seems likely that the state has moved past the origin, the distribution will become nearly symmetrical

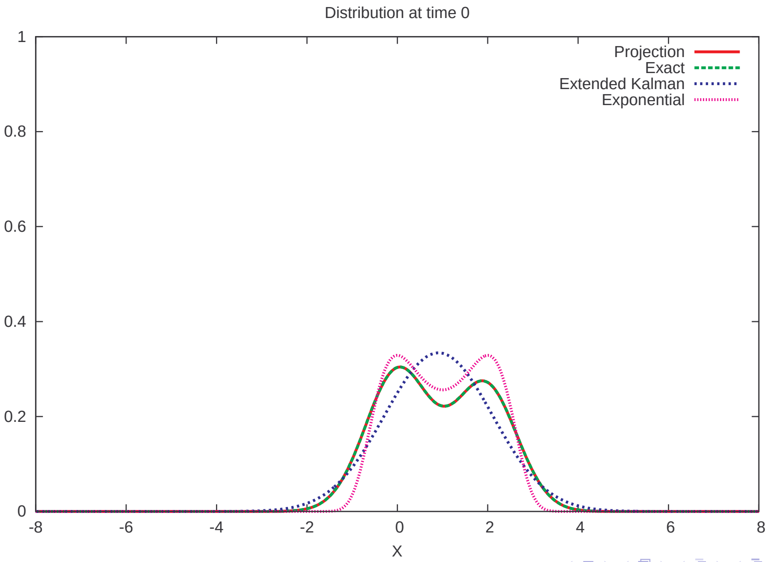
The quadratic sensor

Consider the quadratic sensor

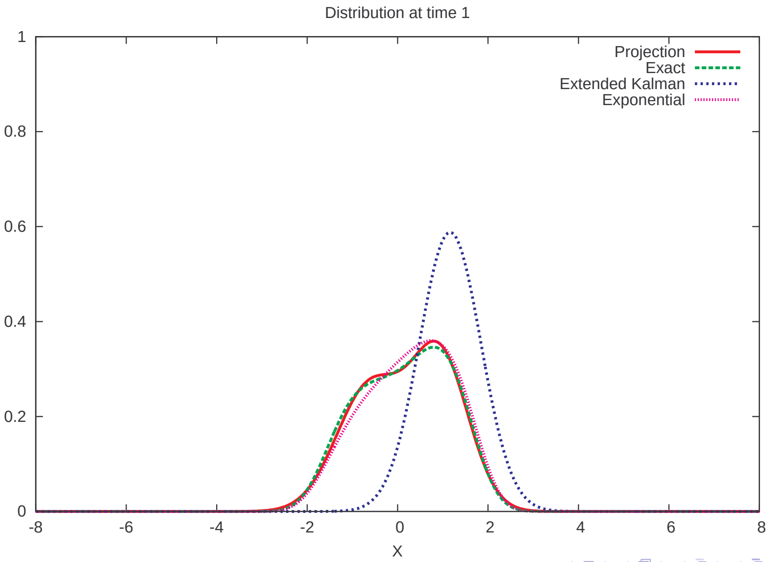
$$\begin{aligned} \mathrm{d}X_t &= \sigma \mathrm{d}W_t \\ \mathrm{d}Y_t &= X^2 \mathrm{d}t + \sigma \mathrm{d}V_t \ . \end{aligned}$$

- The measurements tell us nothing about the sign of X
- Once it seems likely that the state has moved past the origin, the distribution will become nearly symmetrical
- We expect a bimodal distribution

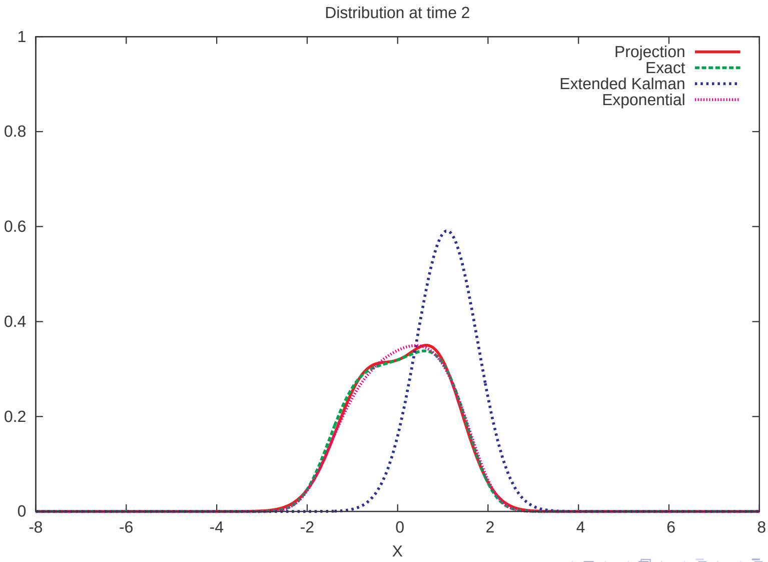
Simulation for the Quadratic Sensor



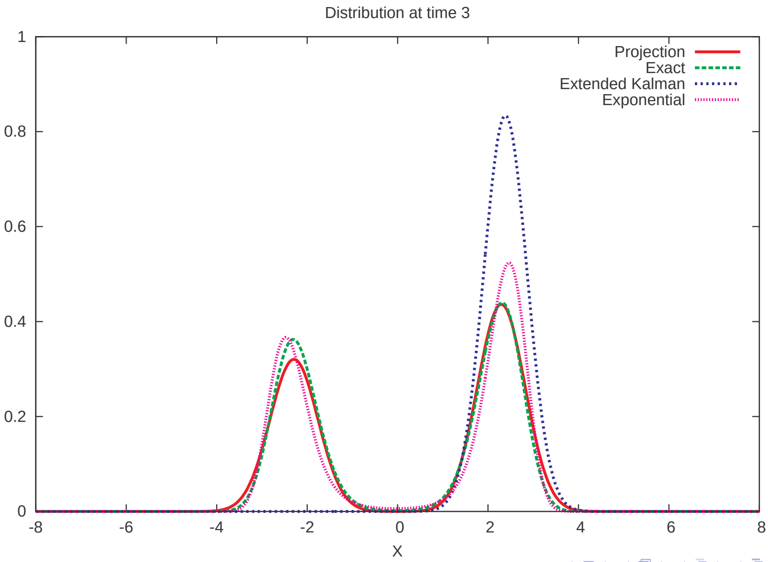
Simulation for the Quadratic Sensor



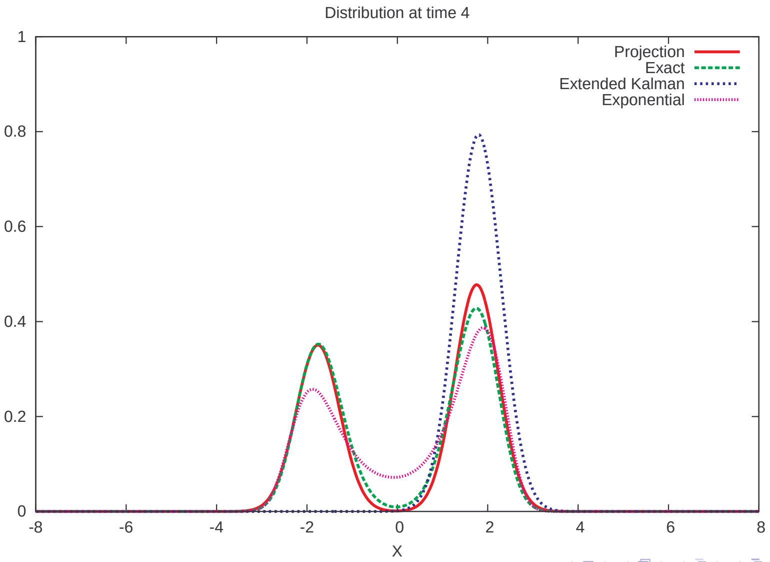
Simulation for the Quadratic Sensor



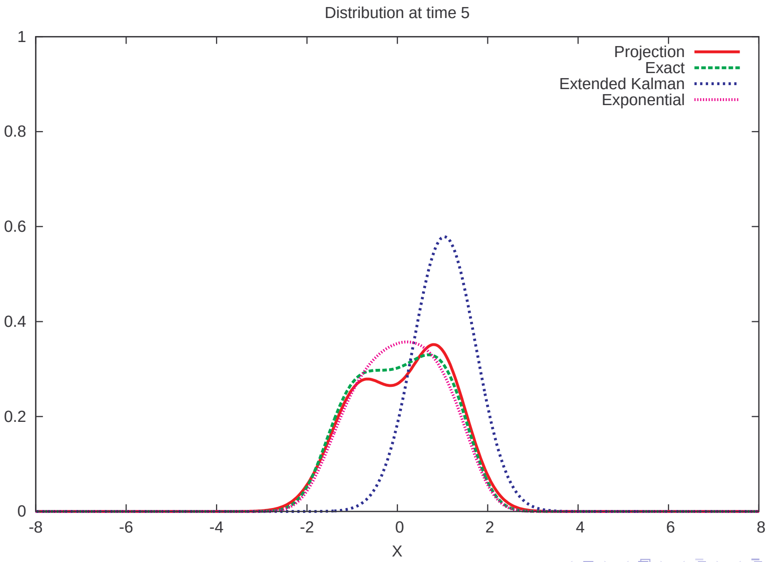
Simulation for the Quadratic Sensor



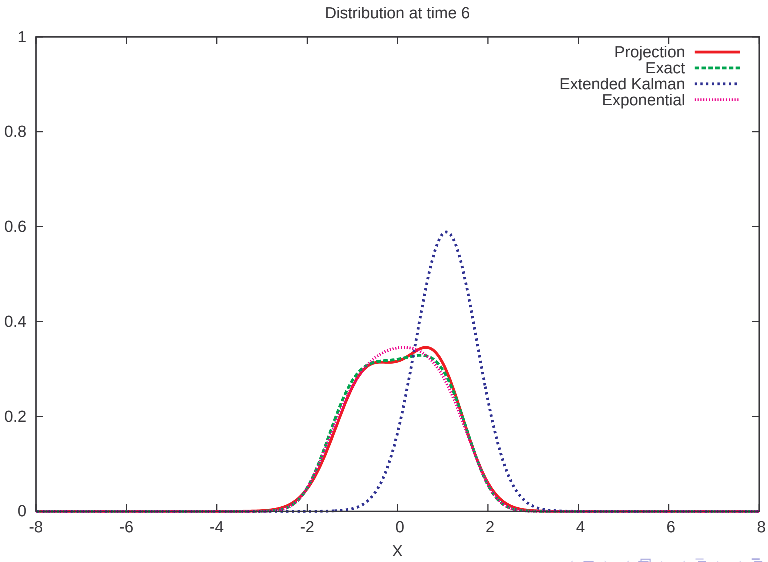
Simulation for the Quadratic Sensor



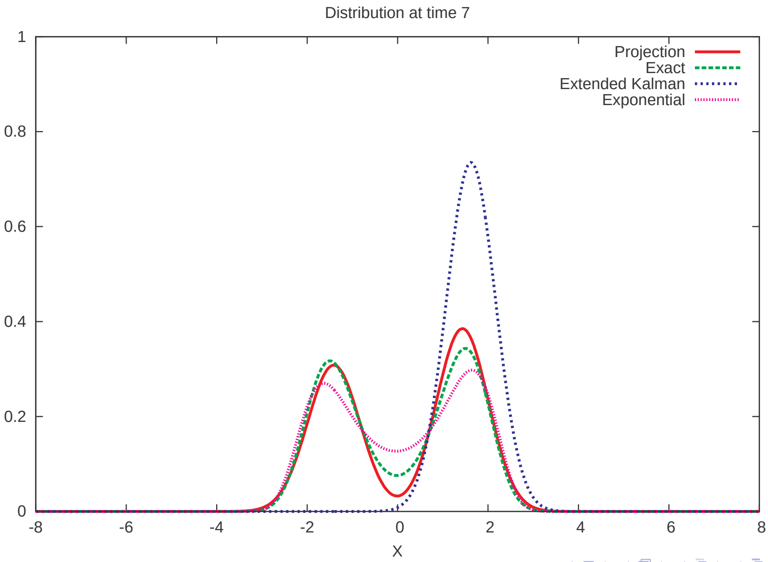
Simulation for the Quadratic Sensor



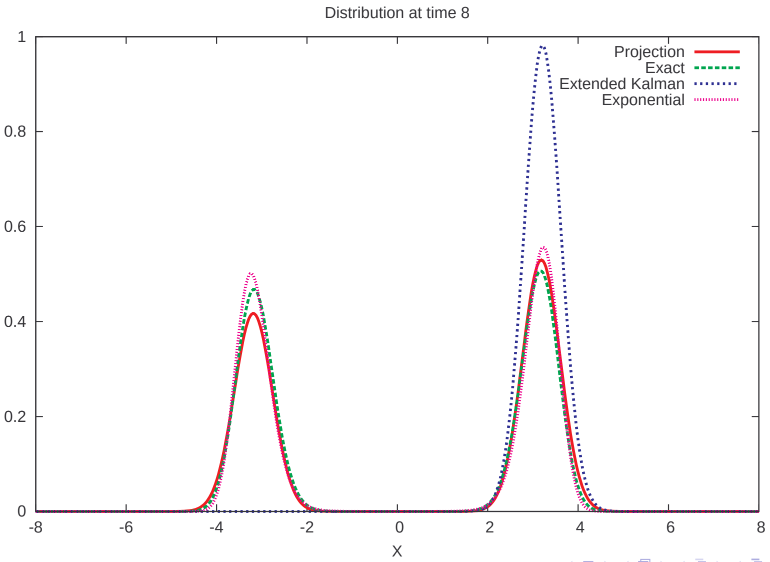
Simulation for the Quadratic Sensor



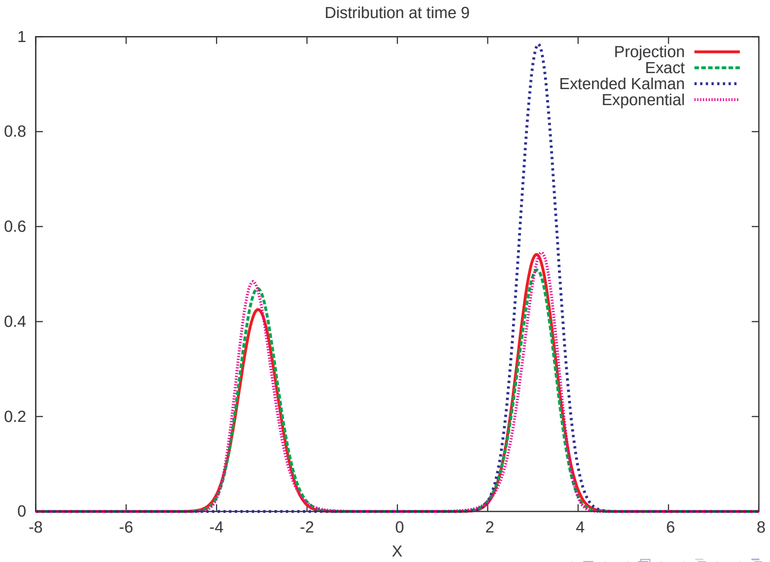
Simulation for the Quadratic Sensor



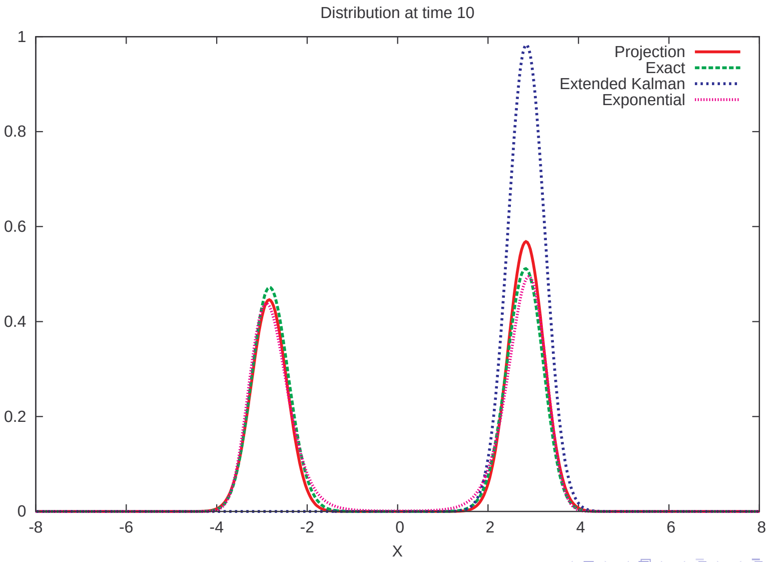
Simulation for the Quadratic Sensor



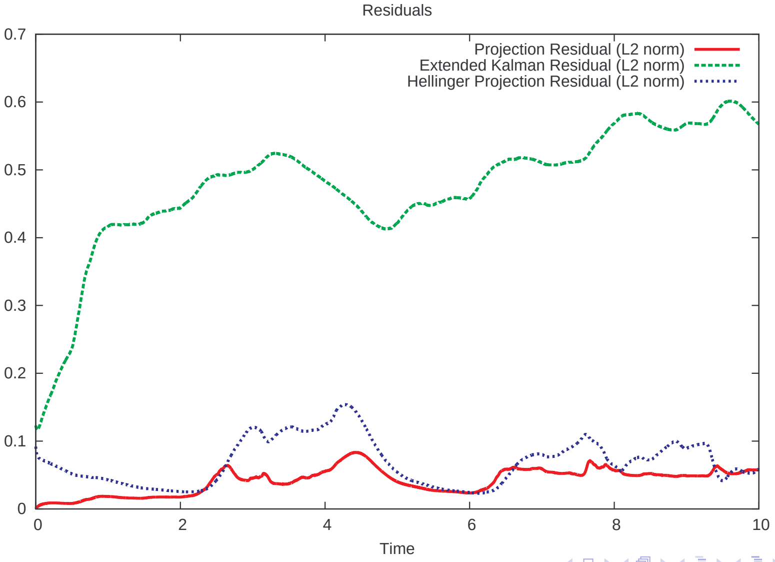
Simulation for the Quadratic Sensor



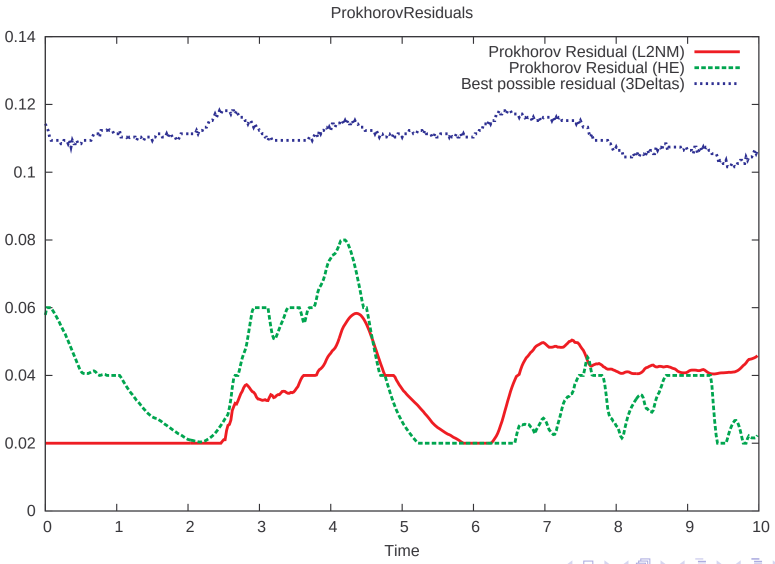
Simulation for the Quadratic Sensor



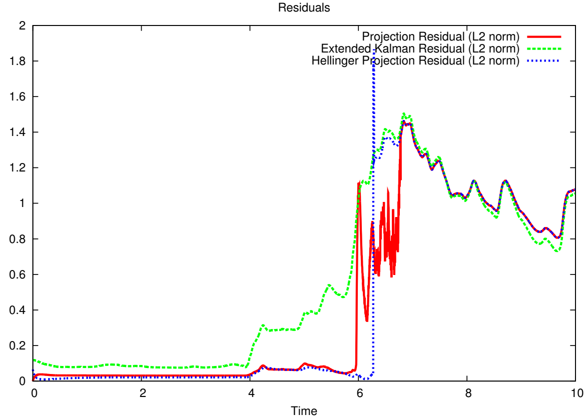
L^2 residuals for the quadratic sensor



Lévy residuals for the quadratic sensor

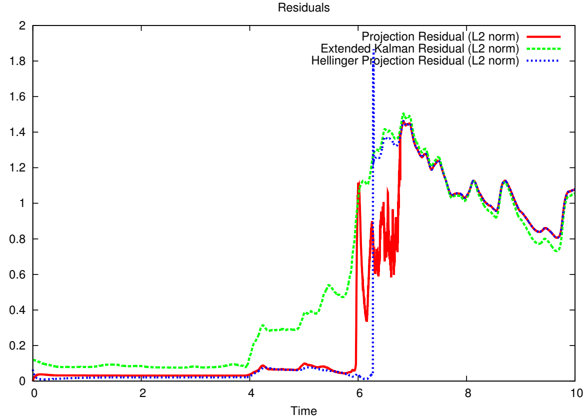


Cubic sensors



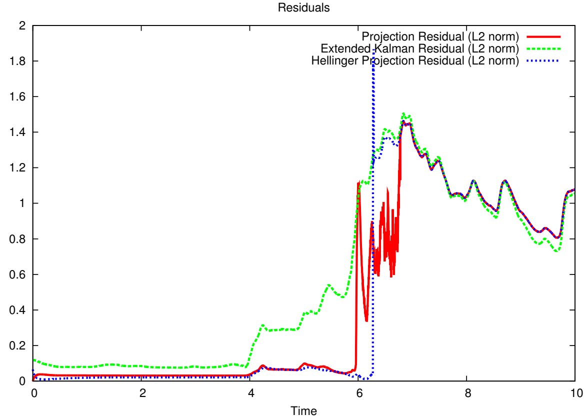
- Qualitatively similar results up to a stopping time

Cubic sensors



- Qualitatively similar results up to a stopping time
- As one approaches the boundary γ_{ij} becomes singular

Cubic sensors



- Qualitatively similar results up to a stopping time
- As one approaches the boundary γ_{ij} becomes singular
- The solution is to dynamically change the parameterization and even the dimension of the manifold.

Conclusion I

- Approximate filtering can be achieved by mathematically rigorous projection methods on the space of densities
- Manifold structure depending on an overarching L^2 structure
- Two different metrics: direct $L2$ and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)
- Fisher works well with exponential families:
  - good for conditional signal with thin tails
  - multimodality,
  - correction step exact,
  - simplicity of implementation
  - equivalence with Assumed Density Filters
- Direct L2 works well with mixture families
  - even simpler filter equations, no inline numerical integration
  - equivalence with Galerkin methods
  - suited also for fat tailed conditional distributions and extreme multimodality (quadratic sensor tests, L2 global filter error)

Thanks

With Thanks to the organizing committee and Almut Veraart in particular for your kind invitation.

Thank you for your kind attention.

Questions and comments very welcome

References I

- J. Aggrawal: Sur l'information de Fisher. In: Theories de l'Information (J. Kamp de Friet, ed.), Springer-Verlag, Berlin–New York 1974, pp. 111-117.
- Amari, S. Differential-geometrical methods in statistics, Lecture notes in statistics, Springer-Verlag, Berlin, 1985
- Beard, R., Kenney, J., Gunther, J., Lawton, J., and Stirling, W. (1999). Nonlinear Projection Filter based on Galerkin approximation. AIAA Journal of Guidance Control and Dynamics, 22 (2): 258-266.
- Beard, R. and Gunther, J. (1997). Galerkin Approximations of the Kushner Equation in Nonlinear Estimation. Working Paper, Brigham Young University.

References II

- Bain, A., and Crisan, D. (2010). Fundamentals of Stochastic Filtering. Springer-Verlag, Heidelberg.
- Barndorff-Nielsen, O.E. (1978). Information and Exponential Families. John Wiley and Sons, New York.
- Brigo, D. Diffusion Processes, Manifolds of Exponential Densities, and Nonlinear Filtering, In: Ole E. Barndorff-Nielsen and Eva B. Vedel Jensen, editor, Geometry in Present Day Science, World Scientific, 1999
- Brigo, D, On SDEs with marginal laws evolving in finite-dimensional exponential families, STAT PROBABIL LETT, 2000, Vol: 49, Pages: 127 – 134

References III

- Brigo, D. (2011). The direct L2 geometric structure on a manifold of probability densities with applications to Filtering. Available on arXiv.org and damianobrigo.it
- Brigo, D, Hanzon, B, LeGland, F, A differential geometric approach to nonlinear filtering: The projection filter, IEEE T AUTOMAT CONTR, 1998, Vol: 43, Pages: 247 – 252
- Brigo, D, Hanzon, B, Le Gland, F, Approximate nonlinear filtering by projection on exponential manifolds of densities, BERNOULLI, 1999, Vol: 5, Pages: 495 – 534
- D. Brigo, On the nice behaviour of the Gaussian Projection Filter with small observation noise, *Systems and Control Letters* **26** (1995) 363–370

References IV

- D. Brigo, New results on the Gaussian projection filter with small observation noise, to appear in *Systems and Control Letters*.
- D. Brigo, *Filtering by Projection on the Manifold of Exponential Densities*, PhD Thesis, Free University of Amsterdam, 1996.
- Brigo, D., and Pistone, G. (1996). Projecting the Fokker-Planck Equation onto a finite dimensional exponential family. Available at arXiv.org
- Crisan, D., and Rozovskii, B. (Eds) (2011). The Oxford Handbook of Nonlinear Filtering, Oxford University Press.
- M. H. A. Davis, S. I. Marcus, An introduction to nonlinear filtering, in: M. Hazewinkel, J. C. Willems, Eds., *Stochastic Systems: The Mathematics of Filtering and Identification and Applications* (Reidel, Dordrecht, 1981) 53–75.

References V

- Elworthy, D. (1982). Stochastic Differential Equations on Manifolds. LMS Lecture Notes.
- Hanzon, B. A differential-geometric approach to approximate nonlinear filtering. In C.T.J. Dodson, Geometrization of Statistical Theory, pages 219 - 223,ULMD Publications, University of Lancaster, 1987.
- B. Hanzon, Identifiability, recursive identification and spaces of linear dynamical systems, CWI Tracts 63 and 64, CWI, Amsterdam, 1989
- M. Hazewinkel, S.I.Marcus, and H.J. Sussmann, Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem, *Systems and Control Letters* **3** (1983) 331–340.

References VI

- J. Jacod, A. N. Shiryaev, Limit theorems for stochastic processes. Grundlehren der Mathematischen Wissenschaften, vol. 288 (1987), Springer-Verlag, Berlin,
- A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- M. Fujisaki, G. Kallianpur, and H. Kunita (1972). Stochastic differential equations for the non linear filtering problem. Osaka J. Math. Volume 9, Number 1 (1972), 19-40.
- Kenney, J., Stirling, W. Nonlinear Filtering of Convex Sets of Probability Distributions. Presented at the 1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July 1999

References VII

- R. Z. Khasminskii (1980). Stochastic Stability of Differential Equations. Alphen aan den Reijn
- J. Lévine, Finite dimensional realizations of stochastic PDE's and application to filtering, *Stochastics and Stochastic Reports* **43** (1991) 75–103.
- R.S. Liptser, A.N. Shiryayev, *Statistics of Random Processes I, General Theory* (Springer Verlag, Berlin, 1978).
- M. Murray and J. Rice - Differential geometry and statistics, Monographs on Statistics and Applied Probability 48, Chapman and Hall, 1993.
- D. Ocone, E. Pardoux, *A Lie algebraic criterion for non-existence of finite dimensionally computable filters*, Lecture notes in mathematics 1390, 197–204 (Springer Verlag, 1989)

References VIII

Pistone, G., and Sempi, C. (1995). An Infinite Dimensional Geometric Structure On the space of All the Probability Measures Equivalent to a Given one. The Annals of Statistics 23(5), 1995