Stochastic PDE projection on manifolds: Assumed-Density and Galerkin Filters

GSI 2015, Oct 28, 2015, Paris

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

—

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

—

Full paper to appear in MCSS, see also arXiv.org

Spaces of probability densities

Consider a parametric family of probability densities

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

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

$$p(\cdot,\theta) \mapsto \theta \ni \mathbb{R}^m, \quad (\sqrt{p(\cdot,\theta)} \mapsto \theta \;\; \text{respectively})$$

is a Chart of a m-dim manifold (?) S ($S^{1/2}$).

Spaces of probability densities

Consider a parametric family of probability densities

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

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

$$p(\cdot,\theta) \mapsto \theta \ni \mathbb{R}^m, \quad (\sqrt{p(\cdot,\theta)} \mapsto \theta \ \text{ respectively})$$

is a Chart of a m-dim manifold (?) S ($S^{1/2}$). The topology & differential structure in the chart is the L^2 structure, but two possibilities:

$$S: \quad d_2(p_1, p_2) = \|p_1 - p_2\| \quad \text{(L2 direct distance)}, \quad p_{1,2} \in L^2$$

$$S^{1/2}: \quad d_H(\sqrt{p_1}, \sqrt{p_2}) = \|\sqrt{p_1} - \sqrt{p_2}\| \quad \text{(Hellinger distance)}, \quad p_{1,2} \in L^1$$

where $\|\cdot\|$ is the norm of Hilbert space L^2.

Tangent vectors, metrics and projection

If $\varphi : \theta \mapsto p(\cdot, \theta)$ ($\theta \mapsto \sqrt{p(\cdot, \theta)}$ resp.) is the inverse of a chart then

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

are linearly independent $L^2(\lambda)$ vector that span Tangent Space at θ.

Tangent vectors, metrics and projection

If $\varphi : \theta \mapsto p(\cdot,\theta)$ $(\theta \mapsto \sqrt{p(\cdot,\theta)}$ resp.) is the inverse of a chart then

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

are linearly independent $L^2(\lambda)$ vector that span Tangent Space at θ.
The inner product of 2 basis elements is defined (L^2 structure)

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

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

$\gamma(\theta)$: direct L2 matrix (d_2); $g(\theta)$: famous Fisher-Rao matrix (d_H)

Tangent vectors, metrics and projection

If $\varphi : \theta \mapsto p(\cdot,\theta)$ ($\theta \mapsto \sqrt{p(\cdot,\theta)}$ resp.) is the inverse of a chart then

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

are linearly independent $L^2(\lambda)$ vector that span Tangent Space at θ. The inner product of 2 basis elements is defined (L^2 structure)

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

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

$\gamma(\theta)$: direct L2 matrix (d_2); $g(\theta)$: famous Fisher-Rao matrix (d_H)

d_2 ort. projection: $\Pi_\theta^\gamma[v] = \sum_{i=1}^m [\sum_{j=1}^m \gamma^{ij}(\theta) \langle v, \frac{\partial p(\cdot,\theta)}{\partial \theta_j} \rangle] \frac{\partial p(\cdot,\theta)}{\partial \theta_i}$

(d_H proj. analogous inserting $\sqrt{\cdot}$ and replacing γ with g)

The nonlinear filtering problem for diffusion signals

$$\begin{aligned} dX_t &= f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \quad \text{(signal)} \\ dY_t &= b_t(X_t)\,dt + dV_t, \quad Y_0 = 0 \;\; \text{(noisy observation)} \end{aligned} \tag{1}$$

These are Itô SDE's. We use both Itô and Stratonovich (Str) SDE's. Str SDE's are necessary to deal with manifolds, since second order Itô terms not clear in terms of manifolds [16], *although we are working on a direct projection of Ito equations with good optimality properties (John Armstrong)*

The nonlinear filtering problem for diffusion signals

$$\begin{aligned} dX_t &= f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \quad \text{(signal)} \\ dY_t &= b_t(X_t)\,dt + dV_t, \quad Y_0 = 0 \;\; \text{(noisy observation)} \end{aligned} \tag{1}$$

These are Itô SDE's. We use both Itô and Stratonovich (Str) SDE's. Str SDE's are necessary to deal with manifolds, since second order Itô terms not clear in terms of manifolds [16], *although we are working on a direct projection of Ito equations with good optimality properties (John Armstrong)*

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 $\mathcal{Y}_t := \sigma(Y_s\,,\, 0 \le s \le t)$. Assume π_t has a density p_t: then p_t satisfies the Str SPDE:

The nonlinear filtering problem for diffusion signals

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

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

The nonlinear filtering problem for diffusion signals

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

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

∞-dimensional SPDE. Solutions for even toy systems the like cubic sensor, $f = 0, \sigma = 1, b = x^3$, do not belong in any finite dim $p(\cdot, \theta)$ [19].

The nonlinear filtering problem for diffusion signals

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

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

∞-dimensional SPDE. Solutions for even toy systems the like cubic sensor, $f = 0, \sigma = 1, b = x^3$, do not belong in any finite dim $p(\cdot, \theta)$ [19].

We need finite dimensional approximations. We can project SPDE according to either the L2 direct metric ($\gamma(\theta)$) or, by deriving the analogous equation for $\sqrt{p_t}$, according to the Hellinger metric ($g(\theta)$).

The nonlinear filtering problem for diffusion signals

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

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

∞-dimensional SPDE. Solutions for even toy systems the like cubic sensor, $f = 0, \sigma = 1, b = x^3$, do not belong in any finite dim $p(\cdot, \theta)$ [19].

We need finite dimensional approximations. We can project SPDE according to either the L2 direct metric ($\gamma(\theta)$) or, by deriving the analogous equation for $\sqrt{p_t}$, according to the Hellinger metric ($g(\theta)$).

Projection transforms the SPDE to a finite dimensional SDE for θ via the chain rule (hence Str calculus): $dp(\cdot, \theta_t) = \sum_{j=1}^m \frac{\partial p(\cdot,\theta)}{\partial \theta_j} \circ d\theta_j(t)$.

The nonlinear filtering problem for diffusion signals

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

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

∞-dimensional SPDE. Solutions for even toy systems the like cubic sensor, $f = 0, \sigma = 1, b = x^3$, do not belong in any finite dim $p(\cdot, \theta)$ [19].

We need finite dimensional approximations. We can project SPDE according to either the L2 direct metric ($\gamma(\theta)$) or, by deriving the analogous equation for $\sqrt{p_t}$, according to the Hellinger metric ($g(\theta)$).

Projection transforms the SPDE to a finite dimensional SDE for θ via the chain rule (hence Str calculus): $dp(\cdot, \theta_t) = \sum_{j=1}^m \frac{\partial p(\cdot, \theta)}{\partial \theta_j} \circ d\theta_j(t)$.

With Ito calculus we would have terms $\frac{\partial^2 p(\cdot, \theta)}{\partial \theta_i \partial \theta_j} d\langle \theta_i, \theta_j \rangle$ (not tang vec)

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

$$d\theta_t^i = \left[\sum_{j=1}^m \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} dx - \sum_{j=1}^m \gamma^{ij}(\theta_t) \int \frac{1}{2}|b_t(x)|^2 \, \frac{\partial p}{\partial \theta_j} \, dx\right] dt$$

$$+ \sum_{k=1}^d [\sum_{j=1}^m \gamma^{ij}(\theta_t) \int b_t^k(x) \, \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k \ , \quad \theta_0^i \ .$$

The above is the projected equation in d_2 metric and Π^γ.

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

$$d\theta_t^i = \left[\sum_{j=1}^m \gamma^{ij}(\theta_t) \int \mathcal{L}_t^* \, p(x,\theta_t) \frac{\partial p(x,\theta_t)}{\partial \theta_j} dx - \sum_{j=1}^m \gamma^{ij}(\theta_t) \int \frac{1}{2} |b_t(x)|^2 \frac{\partial p}{\partial \theta_j} \, dx \right] dt$$

$$+ \sum_{k=1}^d [\sum_{j=1}^m \gamma^{ij}(\theta_t) \int b_t^k(x) \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k \;, \quad \theta_0^i \; .$$

The above is the projected equation in d_2 metric and Π^γ.
Instead, using the Hellinger distance & the Fisher metric with projection Π^g

$$d\theta_t^i = \left[\sum_{j=1}^m g^{ij}(\theta_t) \int \frac{\mathcal{L}_t^* \, p(x,\theta_t)}{p(x,\theta_t)} \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx - \sum_{j=1}^m g^{ij}(\theta_t) \int \frac{1}{2} \, |b_t(x)|^2 \frac{\partial p}{\partial \theta_j} dx \right] dt$$

$$+ \sum_{k=1}^d [\sum_{j=1}^m g^{ij}(\theta_t) \int b_t^k(x) \frac{\partial p(x,\theta_t)}{\partial \theta_j} \, dx] \circ dY_t^k \;, \quad \theta_0^i \; .$$

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- The tangent space has a simple structure: square roots do not complicate issues thanks to the exponential structure.

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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: $\partial^2_{\theta_i,\theta_j} \psi(\theta) = g_{ij}(\theta)$

Choosing the family/manifold: Exponential

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 exponential families $p(x,\theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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 with Y-function b among $c(x)$ exponents makes filter correction step (projection of dY term) exact

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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 with Y-function b among $c(x)$ exponents makes filter correction step (projection of dY term) exact
- One can define both a local and global filtering error through d_H

Choosing the family/manifold: Exponential

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 exponential families $p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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 with Y-function b among $c(x)$ exponents makes filter correction step (projection of dY term) exact
- One can define both a local and global filtering error through d_H
- Alternative coordinates, expectation param., $\eta = E_\theta[c] = \partial_\theta \psi(\theta)$.

Choosing the family/manifold: Exponential

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 exponential families $p(x,\theta) := \exp[\theta^T c(x) - \psi(\theta)]$. Good combination:

- 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 with Y-function b among $c(x)$ exponents makes filter correction step (projection of dY term) exact
- One can define both a local and global filtering error through d_H
- Alternative coordinates, expectation param., $\eta = E_\theta[c] = \partial_\theta \psi(\theta)$.
- Projection filter in η coincides with classical approx filter: assumed density filter (based on generalized "moment matching")

Mixture families

- However, exponential families do not couple as well with the metric $\gamma(\theta)$. *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?*

Mixture families

- However, exponential families do not couple as well with the metric $\gamma(\theta)$. *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?*
- The answer is affirmative, and this is the *mixture family*.

Mixture families

- However, exponential families do not couple as well with the metric $\gamma(\theta)$. *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?*
- The answer is affirmative, and this is the *mixture family*.

We define a *simple mixture family* as follows. Given $m+1$ fixed squared integrable probability densities $\underline{q} = [q_1, q_2, \dots, q_{m+1}]^T$, define

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

for all $\theta \in \mathbb{R}^m$. We write $\hat{\theta}$ instead of $\hat{\theta}(\theta)$. Mixture family (simplex):

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

Mixture families

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

$$\frac{\partial p(\cdot,\theta)}{\partial \theta_i} = q_i - q_{m+1} \text{ and } \gamma_{ij}(\theta) = \int (q_i(x) - q_m(x))(q_j(x) - q_m(x))dx$$

(NO inline numeric integr).

Mixture families

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

$$\frac{\partial p(\cdot,\theta)}{\partial \theta_i} = q_i - q_{m+1} \text{ and } \gamma_{ij}(\theta) = \int (q_i(x) - q_m(x))(q_j(x) - q_m(x))dx$$

(NO inline numeric integr). 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

$$\text{span}\{q_1 - q_{m+1}, q_2 - q_{m+1}, \cdots, q_m - q_{m+1}\}$$

Also the L^2 projection becomes particularly simple.

Mixture Projection Filter

- Armstrong and B. (MCSS 2016 [3]) show that the mixture family + metric $\gamma(\theta)$ lead to a Projection filter that is the same as approximate filtering via Galerkin [5] methods.

Mixture Projection Filter

- Armstrong and B. (MCSS 2016 [3]) show that the mixture family + metric $\gamma(\theta)$ lead to a Projection filter that is the same as approximate filtering via Galerkin [5] methods.
- See the full paper for the details. Summing up:

Family $\rightarrow$ Metric $\downarrow$	Exponential	Basic Mixture
Hellinger d_H Fisher $g(\theta)$	Good $\sim$ADF $\approx$ local moment matching	Nothing special
Direct L^2 d_2 matrix $\gamma(\theta)$	Nothing special	Good ($\sim$Galerkin)

Mixture Projection Filter

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

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

Mixture Projection Filter

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

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

- As a consequence, we are going to enrich our family to a mixture where some of the parameters are also in the core densities q.

Mixture Projection Filter

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

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

- As a consequence, 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. For example for a mixture of two Gaussians we have 5 parameters.

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

Mixture Projection Filter

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

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

- As a consequence, 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. 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.

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

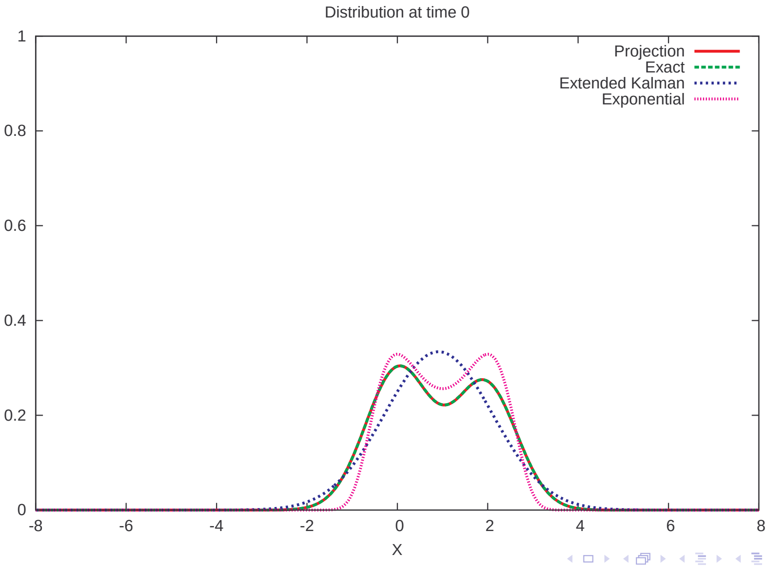
The quadratic sensor

Consider the quadratic sensor

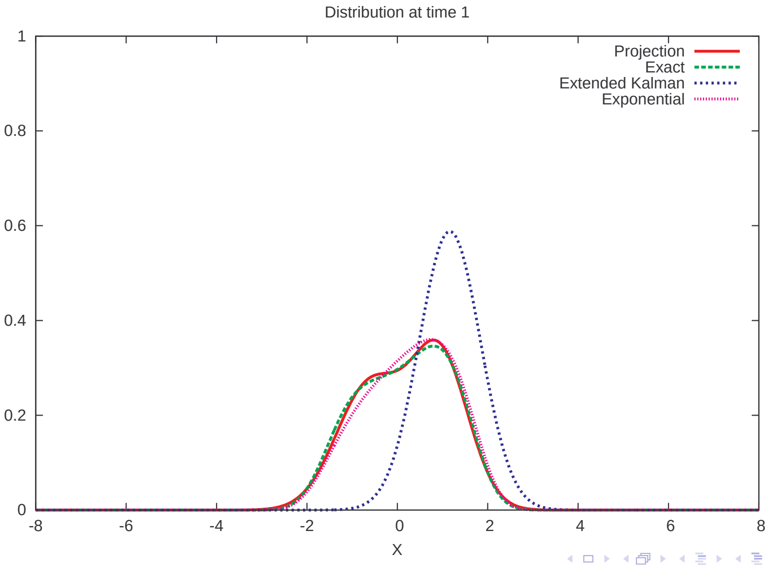
$$\begin{aligned} \mathrm{d}X_t &= \sigma \mathrm{d}W_t \\ \mathrm{d}Y_t &= X^2 \mathrm{d}t + \sigma \mathrm{d}V_t \,. \end{aligned}$$

- 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
- $\theta p_{\mathcal{N}(\mu_1,v_1)}(x) + (1-\theta) p_{\mathcal{N}(\mu_2,v_2)}(x)$ (red)
- vs $e^{\theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 - \psi(\theta)}$ (pink)
- vs EKF ($\mathcal{N}$) (blue)
- vs exact (green, finite diff. method, grid 1000 state & 5000 time)

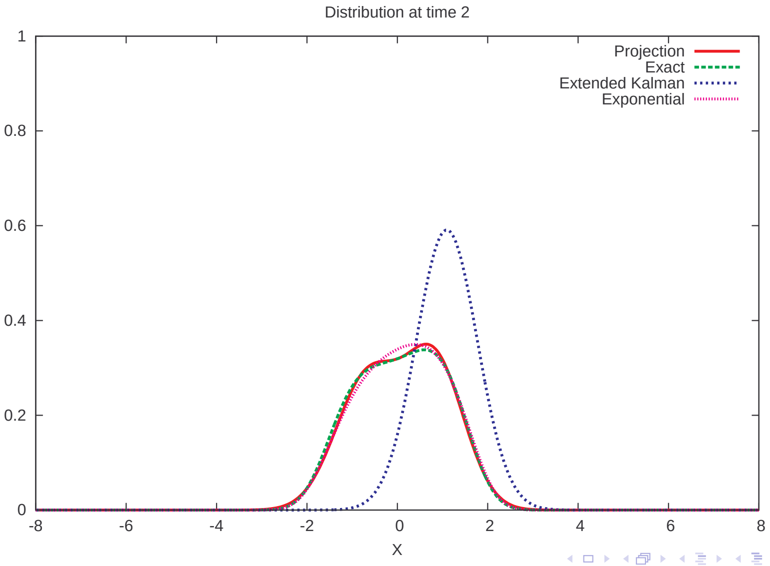
Simulation for the Quadratic Sensor



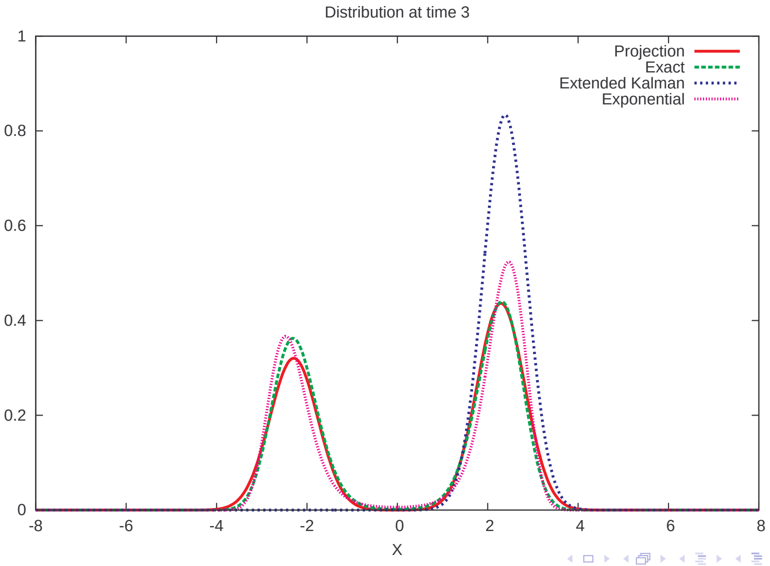
Simulation for the Quadratic Sensor



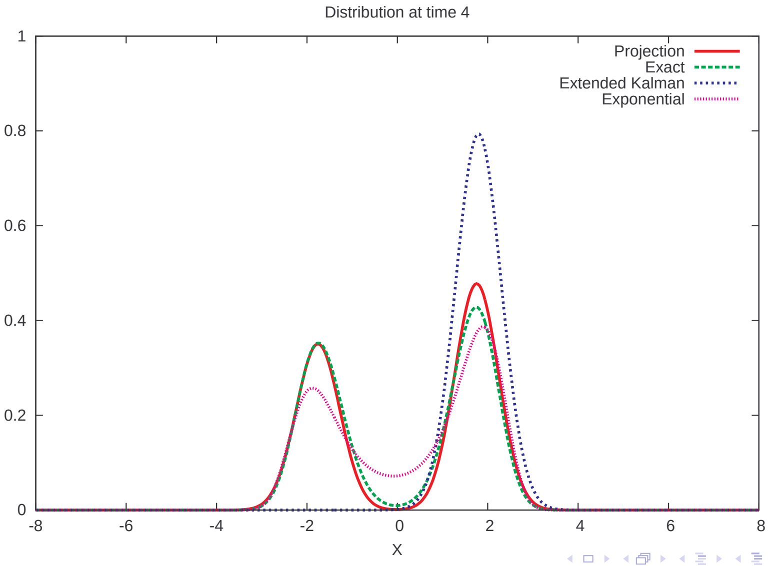
Simulation for the Quadratic Sensor



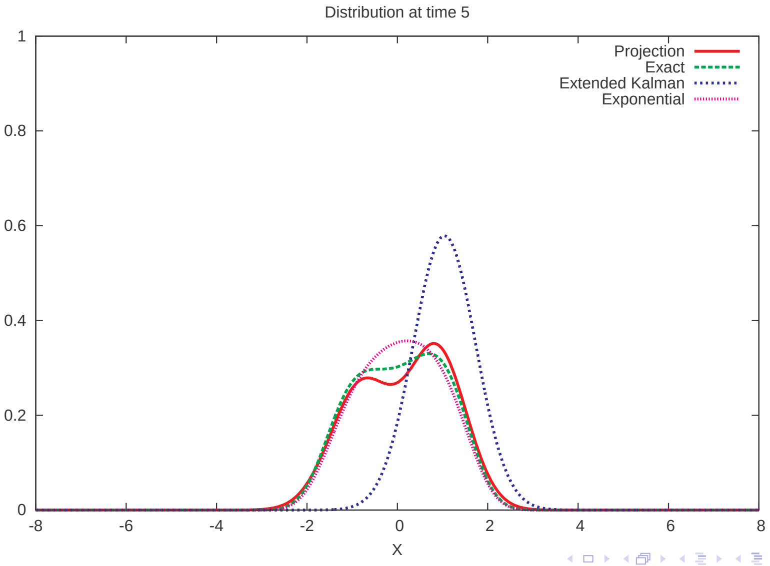
Simulation for the Quadratic Sensor



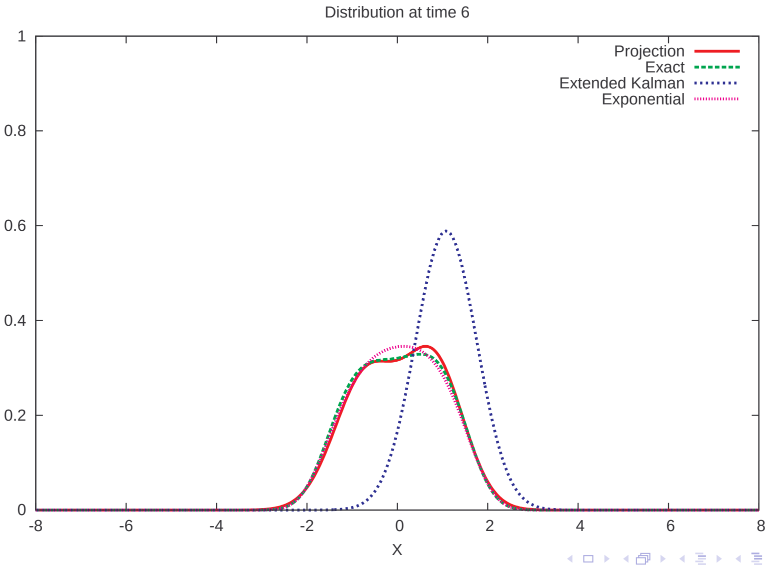
Simulation for the Quadratic Sensor



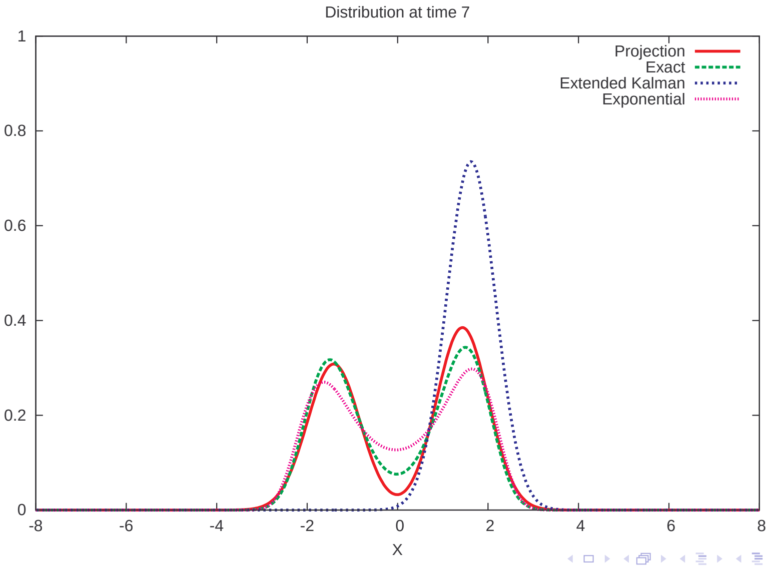
Simulation for the Quadratic Sensor



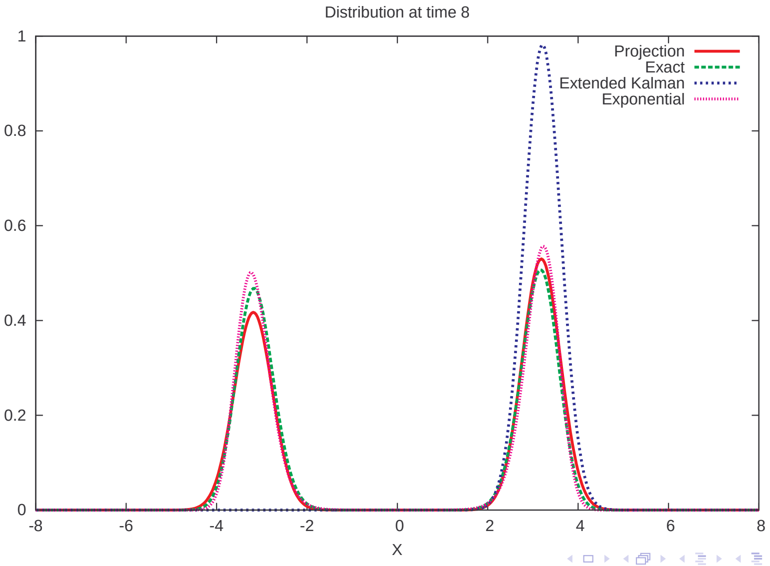
Simulation for the Quadratic Sensor



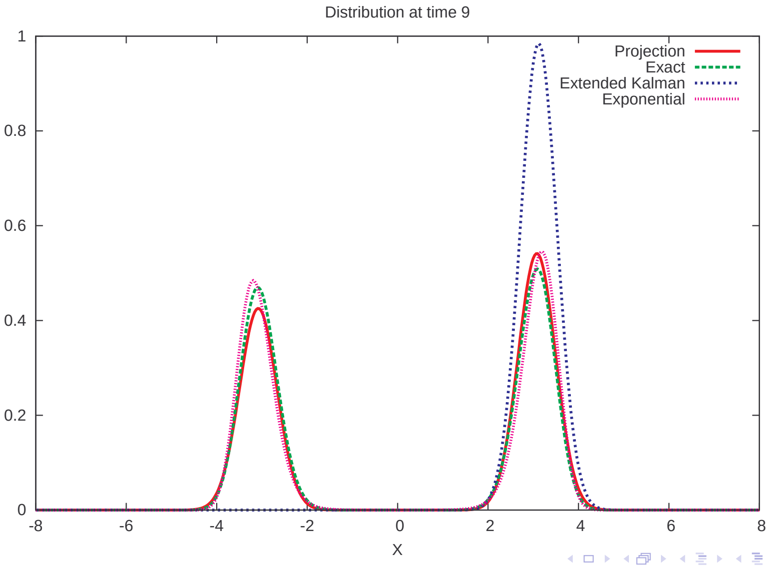
Simulation for the Quadratic Sensor



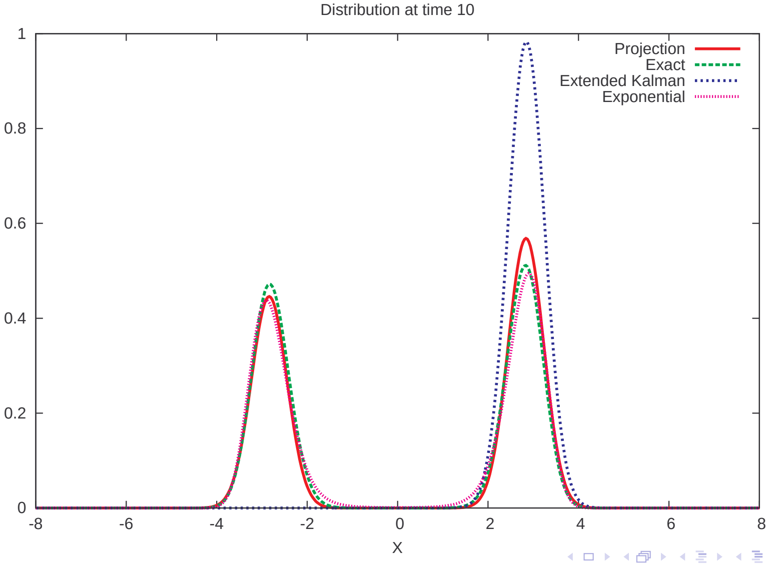
Simulation for the Quadratic Sensor



Simulation for the Quadratic Sensor



Simulation for the Quadratic Sensor

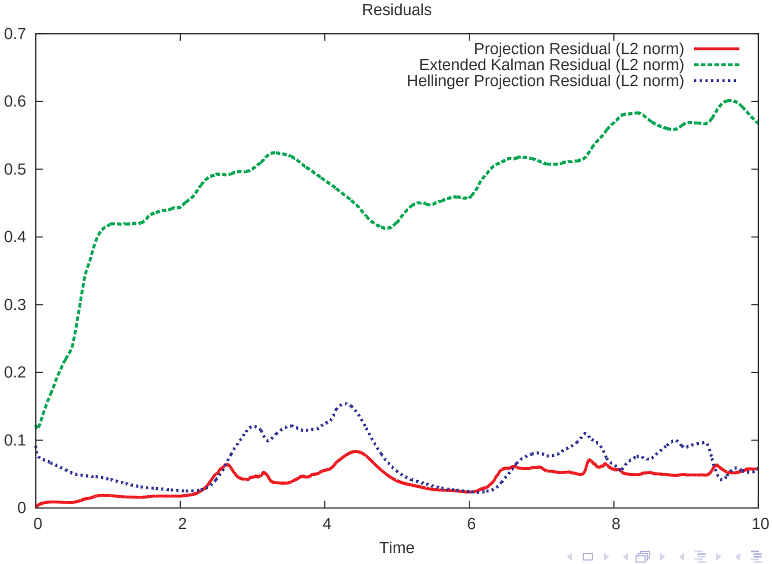


Comparing local approximation errors (L^2 residuals) ε_t

$$\varepsilon_t^2 = \int (p_{\text{exact},t}(x) - p_{\text{approx},t}(x))^2 dx$$

- $p_{\text{approx},t}(x)$: three possible choices.
- $\theta p_{\mathcal{N}(\mu_1, v_1)}(x) + (1-\theta) p_{\mathcal{N}(\mu_2, v_2)}(x)$ (red)
- vs $e^{\theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 - \psi(\theta)}$ (blue)
- vs EKF ($\mathcal{N}$) (green)

L^2 residuals for the quadratic sensor



Comparing local approx errors (Prokhorov residuals) ε_t

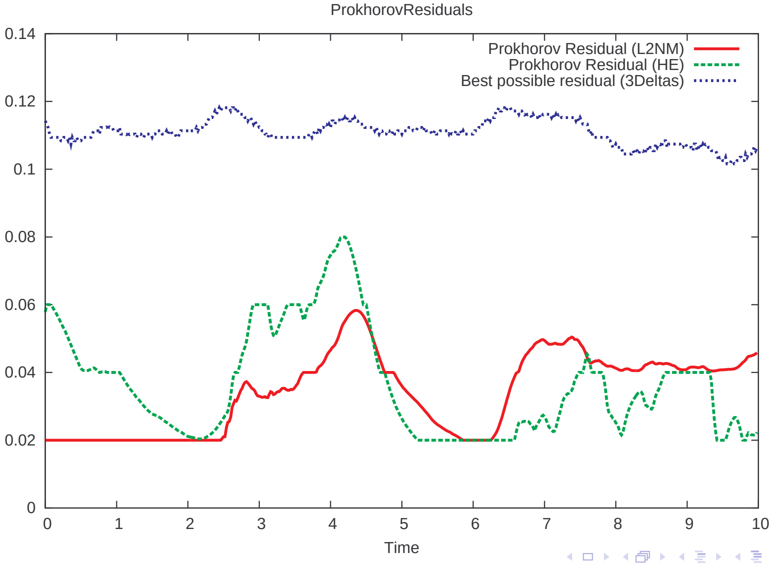
$$\varepsilon_t = \inf\{\epsilon : F_{\text{exact},t}(x-\epsilon) - \epsilon \le F_{\text{approx},t}(x) \le F_{\text{exact},t}(x+\epsilon) + \epsilon \ \ \forall x\}$$

with F the CDF of p's.

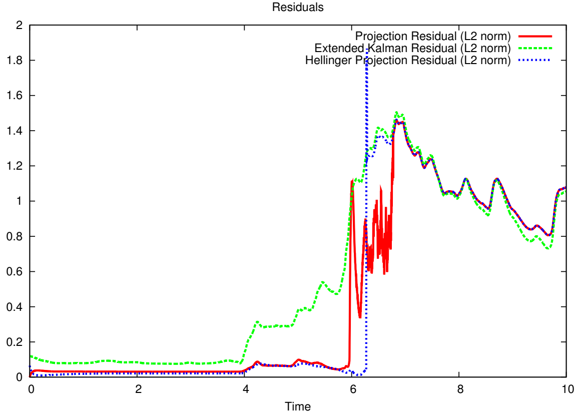
Levy-Prokhorov metric works well with singular densities like particles where L2 metric not ideal.

- $\theta p_{\mathcal{N}(\mu_1,v_1)}(x) + (1-\theta) p_{\mathcal{N}(\mu_2,v_2)}(x)$ (red)
- vs $e^{\theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 - \psi(\theta)}$ (green)
- vs best three particles (blue)

Lévy residuals for the quadratic sensor

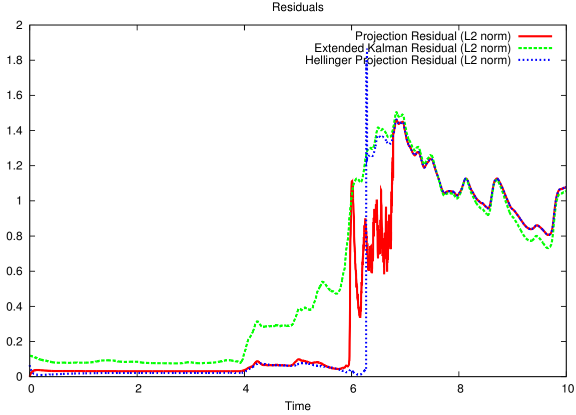


Cubic sensors



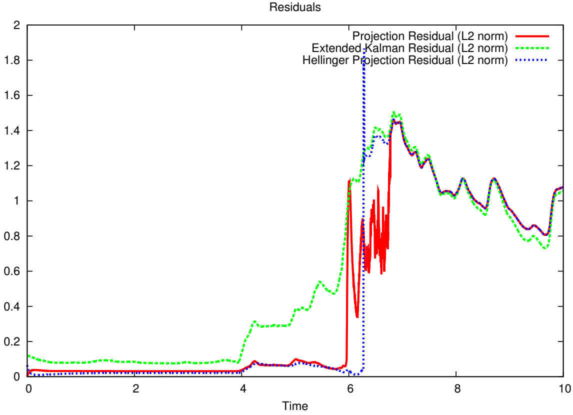
- Qualitatively similar results up to a stopping time

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.

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure
- Two different metrics: direct L^2 and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure
- Two different metrics: direct L^2 and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)
- Fisher works well with exponential families:
  - multimodality,
  - correction step exact,
  - simplicity of implementation
  - equivalence with Assumed Density Filters “moment matching”

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure
- Two different metrics: direct L^2 and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)
- Fisher works well with exponential families:
  - multimodality,
  - correction step exact,
  - simplicity of implementation
  - equivalence with Assumed Density Filters “moment matching”
- Direct L2 works well with mixture families
  - even simpler filter equations, no inline numerical integration
  - basic version equivalent to Galerkin methods
  - suited also for multimodality (quadratic sensor tests, L2 global error)
  - comparable with particle methods

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure
- Two different metrics: direct L^2 and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)
- Fisher works well with exponential families:
  - multimodality,
  - correction step exact,
  - simplicity of implementation
  - equivalence with Assumed Density Filters “moment matching”
- Direct L2 works well with mixture families
  - even simpler filter equations, no inline numerical integration
  - basic version equivalent to Galerkin methods
  - suited also for multimodality (quadratic sensor tests, L2 global error)
  - comparable with particle methods
- Further investigation: convergence, more on *optimality*?

Conclusions

- Approximate finite-dimensional filtering by rigorous projection on a chosen manifold of densities
- Projection uses overarching L^2 structure
- Two different metrics: direct L^2 and Hellinger/Fisher (L^2 on $\sqrt{\cdot}$)
- Fisher works well with exponential families:
  - multimodality,
  - correction step exact,
  - simplicity of implementation
  - equivalence with Assumed Density Filters "moment matching"
- Direct L2 works well with mixture families
  - even simpler filter equations, no inline numerical integration
  - basic version equivalent to Galerkin methods
  - suited also for multimodality (quadratic sensor tests, L2 global error)
  - comparable with particle methods
- Further investigation: convergence, more on *optimality*?
- Optimality: introducing new projections (forthcoming J. Armstrong)

Thanks

With thanks to the organizing committee.

Thank you for your attention.

Questions and comments welcome

References I

[1] J. Aggrawal: Sur l'information de Fisher. In: Theories de l'Information (J. Kampe de Feriet, ed.), Springer-Verlag, Berlin–New York 1974, pp. 111-117.

[2] Amari, S. Differential-geometrical methods in statistics, Lecture notes in statistics, Springer-Verlag, Berlin, 1985

[3] Armstrong, J., and Brigo, D. (2016). Nonlinear filtering via stochastic PDE projection on mixture manifolds in L^2 direct metric, *Mathematics of Control, Signals and Systems*, 2016, accepted.

[4] 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.

References II

[5] Beard, R. and Gunther, J. (1997). Galerkin Approximations of the Kushner Equation in Nonlinear Estimation. Working Paper, Brigham Young University.

[6] Barndorff-Nielsen, O.E. (1978). Information and Exponential Families. John Wiley and Sons, New York.

[7] 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

[8] Brigo, D, On SDEs with marginal laws evolving in finite-dimensional exponential families, STAT PROBABIL LETT, 2000, Vol: 49, Pages: 127 – 134

References III

[9] 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

[10] 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

[11] Brigo, D, Hanzon, B, Le Gland, F, Approximate nonlinear filtering by projection on exponential manifolds of densities, BERNOULLI, 1999, Vol: 5, Pages: 495 – 534

[12] D. Brigo, *Filtering by Projection on the Manifold of Exponential Densities*, PhD Thesis, Free University of Amsterdam, 1996.

References IV

[13] Brigo, D., and Pistone, G. (1996). Projecting the Fokker-Planck Equation onto a finite dimensional exponential family. Available at arXiv.org

[14] Crisan, D., and Rozovskii, B. (Eds) (2011). The Oxford Handbook of Nonlinear Filtering, Oxford University Press.

[15] 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.

[16] Elworthy, D. (1982). Stochastic Differential Equations on Manifolds. LMS Lecture Notes.

References V

[17] 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.

[18] B. Hanzon, Identifiability, recursive identification and spaces of linear dynamical systems, CWI Tracts 63 and 64, CWI, Amsterdam, 1989

[19] 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.

[20] J. Jacod, A. N. Shiryaev, Limit theorems for stochastic processes. Grundlehren der Mathematischen Wissenschaften, vol. 288 (1987), Springer-Verlag, Berlin,

References VI

[21] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[22] 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.

[23] 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

[24] R. Z. Khasminskii (1980). Stochastic Stability of Differential Equations. Alphen aan den Reijn

[25] R.S. Liptser, A.N. Shiryayev, *Statistics of Random Processes I, General Theory* (Springer Verlag, Berlin, 1978).

References VII

[26] M. Murray and J. Rice - Differential geometry and statistics, Monographs on Statistics and Applied Probability 48, Chapman and Hall, 1993.

[27] 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)

[28] 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