Information Geometric Nonlinear Filtering:
a Hilbert Space Approach

Nigel Newton  (University of Essex)

Information Geometry and its Applications IV,
Liblice, June 2016

In honour of Shun-ichi Amari
on the occasion of his 80th birthday
Overview

• Nonlinear Filtering (recursive Bayesian estimation)
  – The need for a proper state space for posterior distributions

• The infinite-dimensional Hilbert manifold of probability measures, $M$, (and Banach variants)

• An $M$-valued Itô stochastic differential equation for the nonlinear filter

• Information geometric properties of the nonlinear filter
Nonlinear Filtering

- Markov “signal” process: \( (X_t \in \mathbb{X}, t \in [0, \infty)) \)
  - \((X, \mu)\) is a metric space, with reference probability measure \(\mu\)
  - Eg. \(X = \mathbb{R}^d, \mu = N(0, I)\)

- Partial “observation” process: \( (Y_t \in \mathbb{R}, t \in [0, \infty)) \)
  \[
  Y_t = \int_0^t h(X_s) ds + W_t
  \]
  Brownian Motion, independent of \(X\)
Nonlinear Filtering

• Markov “signal” process: \((X_t \in \mathbb{X}, t \in [0, \infty))\)
  
  – \((X, \mu)\) is a metric space, with reference probability measure \(\mu\)
  
  – Eg. \(\mathbb{X} = \mathbb{R}^d, \mu = N(0, I)\)

• Partial “observation” process: \((Y_t \in \mathbb{R}, t \in [0, \infty))\)

\[
Y_t = \int_0^t h(X_s)ds + W_t
\]

Brownian Motion, independent of \(X\)

• Estimate \(X_t\) at each time \(t\) from its prior distribution \(P_t\) and the history of the observation:

\[
Y_0^t := (Y_s, s \in [0, t])
\]

• The linear-Gaussian case yields the \textit{Kalman-Bucy filter}
Nonlinear Filtering

- Regular conditional (posterior) distribution: $\Pi_t : \Omega \rightarrow \mathcal{P}(X)$

$$\Pi_t(B) = \mathbf{P}(X_t \in B \mid Y_0^t)$$

- $\Pi_t$ is a random probability measure evolving on $\mathcal{P}(X)$. How should we represent it?
Nonlinear Filtering

- Regular conditional (posterior) distribution: \( \Pi_t : \Omega \rightarrow \mathcal{P}(X) \)

\[
\Pi_t(B) = P \left( X_t \in B \mid Y_0^t \right)
\]

- \( \Pi_t \) is a random probability measure evolving on \( \mathcal{P}(X) \). How should we represent it?

- We could consider the conditional density (w.r.t \( \mu \)), \( \pi_t \)
  - typical differential equation (Shiriyayev, Wonham, Stratonovich, Kushner):

\[
''d\pi_t = \mathcal{A}\pi_t dt + \pi_t (h - \bar{h}_t)(dY_t - \bar{h}_t dt)''
\]

\( \bar{h}_t := \int h(x)\Pi_t(dx) \)

- Spaces of densities are not necessarily optimal
Mean-Square Errors

- Suppose $Ef(X_t)^2 < \infty$ for some $f : X \to \mathbb{R}$
- Then $\bar{f}_t := E_{\Pi_t} f$ minimises the mean-square error

$$E(f(X_t) - \hat{f}_t)^2 = E\left(E_{\Pi_t} (f - \bar{f}_t)^2 + (\bar{f}_t - \hat{f}_t)^2\right)$$

estimation error + approximation error
Mean-Square Errors

• Suppose $Ef(X_t)^2 < \infty$ for some $f : X \to \mathbb{R}$

• Then $\bar{f}_t := E_{\Pi_t} f$ minimises the mean-square error

$$E(f(X_t) - \hat{f}_t)^2 = E \left( E_{\Pi_t} (f - \bar{f}_t)^2 + (\bar{f}_t - \hat{f}_t)^2 \right)$$

estimation error + approximation error

• If $\hat{f}_t = E_{\hat{\Pi}_t} f$ for some $\hat{\Pi}_t : \Omega \to \mathcal{P}(X)$, and $\Pi_t, \hat{\Pi}_t \ll \mu$ then

$$(\bar{f}_t - \hat{f}_t)^2 \leq E_{\mu} f^2 E_{\mu} (\pi_t - \hat{\pi}_t)^2$$

and so the $L^2(\mu)$ norm on densities may be useful
Mean-Square Errors

• Suppose $\mathbb{E} f(X_t)^2 < \infty$ for some $f : X \to \mathbb{R}$

• Then $\bar{f}_t := \mathbb{E}_{\Pi_t} f$ minimises the mean-square error

\[
\mathbb{E}(f(X_t) - \hat{f}_t)^2 = \mathbb{E}\left(\mathbb{E}_{\Pi_t} (f - \bar{f}_t)^2 + (\bar{f}_t - \hat{f}_t)^2\right)
\]

estimation error + approximation error

• If $\hat{f}_t = \mathbb{E}_{\hat{\Pi}_t} f$ for some $\hat{\Pi}_t : \Omega \to \mathcal{P}(X)$, and $\Pi_t, \hat{\Pi}_t \ll \mu$ then

\[
(\bar{f}_t - \hat{f}_t)^2 \leq \mathbb{E}_{\mu} f^2 \mathbb{E}_{\mu} (\pi_t - \hat{\pi}_t)^2
\]

and so the $L^2(\mu)$ norm on densities may be useful

• Not if $f = 1_B$ and $\Pi_t(B)$ is very small (Eg. fault detection)
Mean-Square Errors

• Suppose $E f(X_t)^2 < \infty$ for some $f : X \to \mathbb{R}$

• Then $\tilde{f}_t := E_{\Pi_t} f$ minimises the mean-square error

  $E(f(X_t) - \tilde{f}_t)^2 = E \left( E_{\Pi_t} (f - \bar{f}_t)^2 + (\bar{f}_t - \hat{f}_t)^2 \right)$

  \[ \text{estimation error} + \text{approximation error} \]

• If $\hat{f}_t = E_{\hat{\Pi}_t} f$ for some $\hat{\Pi}_t : \Omega \to \mathcal{P}(X)$, and $\Pi_t, \hat{\Pi}_t << \mu$ then

  \[ (\hat{f}_t - \bar{f}_t)^2 \leq E_{\mu} f^2 E_{\mu} (\pi_t - \hat{\pi}_t)^2 \]

  and so the $L^2(\mu)$ norm on densities may be useful

• Not if $f = 1_B$ and $\Pi_t(B)$ is very small (Eg. fault detection)

• When topologised in this way, $\mathcal{P}(X)$ has a boundary
Multi-Objective Mean-Square Errors

- Maximising the $L^2$ error over square-integrable functions

\[ M(\hat{\Pi}_t \mid \Pi_t) := \sup_{f \in L^2(\Pi_t)} \frac{(\bar{f}_t - \hat{f}_t)^2}{E_{\Pi_t} (f - \bar{f}_t)^2} \]

\[ = \sup_{f \in F} \left( E_{\Pi_t} f (1 - d\hat{\Pi}_t / d\Pi_t) \right)^2 \]

\[ = E_{\Pi_t} (1 - d\hat{\Pi}_t / d\Pi_t)^2 \]

where \( F := \{ f \in L^2(\Pi_t) : \bar{f}_t = 0, E_{\Pi_t} f^2 = 1 \} \)
Multi-Objective Mean-Square Errors

- Maximising the $L^2$ error over square-integrable functions

\[
\mathcal{M}(\hat{\Pi}_t | \Pi_t) := \sup_{f \in L^2(\Pi_t)} \frac{(\bar{f}_t - \hat{f}_t)^2}{E_{\Pi_t} (f - \bar{f}_t)^2}
\]

\[
= \sup_{f \in F} \left( E_{\Pi_t} f (1 - d\hat{\Pi}_t / d\Pi_t) \right)^2
\]

\[
= E_{\Pi_t} (1 - d\hat{\Pi}_t / d\Pi_t)^2
\]

where \( F := \{ f \in L^2(\Pi_t) : \bar{f}_t = 0, E_{\Pi_t} f^2 = 1 \} \)

- In time-recursive approximations, the accuracy of $\hat{\Pi}_t$ is affected by that of $\hat{\Pi}_s$ ($s < t$). This naturally induces multi-objective criteria at time $s$ (nonlinear dynamics).
Geometric Sensitivity

- $\mathcal{M}$ is "geometrically sensitive". (It requires small probabilities to be approximated with greater absolute accuracy than large probabilities)

- When topologised by $\mathcal{M}$, $\mathcal{P}(X)$ does not have a boundary
Geometric Sensitivity

- $\mathcal{M}$ is “geometrically sensitive”. (It requires small probabilities to be approximated with greater absolute accuracy than large probabilities)

- When topologised by $\mathcal{M}$, $\mathcal{P}(\mathbf{X})$ does not have a boundary

- This is highly desirable in the context of recursive Bayesian estimation, where conditional probabilities are repeatedly multiplied by the likelihood functions of new observations.
Geometric Sensitivity

• $\mathcal{M}$ is “geometrically sensitive”. (It requires small probabilities to be approximated with greater absolute accuracy than large probabilities.)

• When topologised by $\mathcal{M}$, $P(X)$ does not have a boundary.

• This is highly desirable in the context of recursive Bayesian estimation, where conditional probabilities are repeatedly multiplied by the likelihood functions of new observations.

• $\mathcal{M}$ is Pearson’s $\chi^2$ divergence. It belongs to the one-parameter family of $\alpha$-divergences: $\mathcal{M} = \mathcal{D}_{-3}$
Geometric Sensitivity

• $\mathcal{M}$ is “geometrically sensitive”. (It requires small probabilities to be approximated with greater absolute accuracy than large probabilities.)

• When topologised by $\mathcal{M}$, $\mathcal{P}(X)$ does not have a boundary.

• This is highly desirable in the context of recursive Bayesian estimation, where conditional probabilities are repeatedly multiplied by the likelihood functions of new observations.

• $\mathcal{M}$ is Pearson’s $\chi^2$ divergence. It belongs to the one-parameter family of $\alpha$-divergences: $\mathcal{M} = \mathcal{D}_{-3}$

• It is too restrictive to use in practice
\(\alpha\)-Divergences

- As \(|\alpha|\) becomes larger, \(\mathcal{D}_\alpha\) becomes increasingly "geometrically sensitive"
- The case \(\alpha = 0\) yields the *Hellinger metric*
\(\alpha\)-Divergences

- As \(|\alpha|\) becomes larger \(\mathcal{D}_\alpha\) becomes increasingly “geometrically sensitive”
- The case \(\alpha = 0\) yields the Hellinger metric
- The case \(\alpha = \pm 1\) yields the KL-Divergence:

\[
\mathcal{D}(P \mid Q) := \mathcal{D}_{\pm 1}(P \mid Q) = E_Q \frac{dP}{dQ} \log \frac{dP}{dQ}
\]

- This is widely used in practice.
\(\alpha\)-Divergences

- As \(|\alpha|\) becomes larger \(D_\alpha\) becomes increasingly “geometrically sensitive”
- The case \(\alpha = 0\) yields the **Hellinger metric**
- The case \(\alpha = \pm 1\) yields the **KL-Divergence**:
  \[
  D(P \mid Q) := D_{-1}(P \mid Q) = E_Q \frac{dP}{dQ} \log \frac{dP}{dQ}
  \]
- This is widely used in practice.
- Symmetric error criteria may be appropriate, such as
  \[
  D(\hat{\Pi}_t \mid \Pi_t) + D(\Pi_t \mid \hat{\Pi}_t)
  \]
Connections with Information Theory

• Conditional mutual information (un-averaged):

\[ I(X; Y | Z) := \mathcal{D}(P_{XY|Z} \mid P_{X|Z} \otimes P_{Y|Z}) \]

• Additivity property:

\[ I(X; (Y, Z)) = I(X; Z) + \mathbf{E}I(X; Y | Z) \]
Connections with Information Theory

- Conditional mutual information (unaveraged):
  \[ I(X;Y | Z) := \mathcal{D}(P_{XY|Z} | P_X|Z \otimes P_Y|Z) \]

- Additivity property:
  \[ I(X;(Y,Z)) = I(X;Z) + \mathbf{E} I(X;Y | Z) \]

- Information Supply to the nonlinear filter:
  \[ S(t) := I(X;Y_0^t) \]

- The filter continuously fuses new observation information
  \[ S(t) = S(s) + \mathbf{E} I(X;Y_s^t | Y_0^s) \]
Appropriate Metrics on $\mathcal{P}(X)$

- The KL divergence is **bilinear** in the density and its log (regarded as elements of dual spaces of functions).

- For $P, Q \in \mathcal{P}(X)$ with $P, Q \ll \mu$

$$D(P \mid Q) = \langle p, \log p \rangle - \langle p, \log q \rangle$$

where $p$ and $q$ are the densities
Appropriate Metrics on $\mathcal{P}(\mathbf{X})$

- The KL divergence is **bilinear** in the density and its log (regarded as elements of dual spaces of functions).

- For $P, Q \in \mathcal{P}(\mathbf{X})$ with $P, Q \ll \mu$

  \[
  \mathcal{D}(P | Q) = \langle p, \log p \rangle - \langle p, \log q \rangle
  \]

  where $p$ and $q$ are the densities

- So we would like the metric to “control” both $p$ and $\log p$
Maximal Exponential Model
(G. Pistone et al.)

• \( \mathcal{E}(\mu) = \{ P \in \mathcal{P}(X) : p = \exp(a - K_\mu(a)) \mid a \in S_\mu \} \)

• **Model space** (exponential Orlicz):

\[ B_\mu = \{ a : X \to \mathbb{R} : E_\mu a = 0, E_\mu \cosh(\alpha a) < \infty \text{ for some } \alpha > 0 \} \]
Maximal Exponential Model
(G. Pistone et al.)

• $\mathcal{E}(\mu) = \{ P \in \mathcal{P}(X) : p = \exp(a - K_\mu(a)) \mid a \in S_\mu \}$

• **Model space** (exponential Orlicz):

  $B_\mu = \{ a : X \rightarrow \mathbb{R} : E_\mu a = 0, E_\mu \cosh(\alpha a) < \infty \text{ for some } \alpha > 0 \}$

• **Global Chart**: $s_\mu : \mathcal{E}(\mu) \rightarrow B_\mu$

  $s_\mu(P) := \log(p) - E_\mu \log(p)$
Maximal Exponential Model
(G. Pistone et al.)

• $\mathcal{E}(\mu) = \{ P \in \mathcal{P}(\mathbf{X}) : p = \exp(a - K_\mu(a)) | a \in S_\mu \}$

• Model space (exponential Orlicz):

$$B_\mu = \{ a : \mathbf{X} \to \mathbb{R} : \mathbb{E}_\mu a = 0, \mathbb{E}_\mu \cosh(\alpha a) < \infty \text{ for some } \alpha > 0 \}$$

• Global Chart: $s_\mu : \mathcal{E}(\mu) \to B_\mu$

$$s_\mu(P) := \log(p) - \mathbb{E}_\mu \log(p)$$

• Mixture Map: $\eta_\mu : \mathcal{E}(\mu) \to \star B_\mu$

$$\eta_\mu(P) := p - 1$$

Injective and of class $C^\infty$, but not homeomorphic
The Hilbert Manifold $\mathcal{M}$

$\mathcal{M}$ is the subset of $\mathcal{P}(\mathbf{X})$ whose members have the following properties:

$$P \sim \mu, \quad E_\mu p^2 < \infty \quad \text{and} \quad E_\mu \log^2 p < \infty$$
The Hilbert Manifold $\mathcal{M}$

- $\mathcal{M}$ is the subset of $\mathcal{P}(\mathbf{X})$ whose members have the following properties:

$$P \sim \mu, \quad E_\mu p^2 < \infty \quad \text{and} \quad E_\mu \log^2 p < \infty$$

- **Model space:**

$$H = L^2_0(\mu) = \{ a : \mathbf{X} \to \mathbb{R} : E_\mu a = 0, \quad E_\mu a^2 < \infty \}$$

- **Global Chart:** $\phi : \mathcal{M} \to H$

$$\phi(P) := p - 1 + \log p - E_\mu \log p$$
The Hilbert Manifold $M$

- $M$ is the subset of $\mathcal{P}(X)$ whose members have the following properties:

\[ P \sim \mu, \quad \mathbb{E}_\mu p^2 < \infty \quad \text{and} \quad \mathbb{E}_\mu \log^2 p < \infty \]

- **Model space:**

\[ H = L_0^2(\mu) = \{ a : X \to \mathbb{R} : \mathbb{E}_\mu a = 0, \mathbb{E}_\mu a^2 < \infty \} \]

- **Global Chart:** $\phi : M \to H$

\[ \phi(p) := p - 1 + \log p - \mathbb{E}_\mu \log p \]

- **Proposition 1:** $\phi$ is a bijection onto $H$
$M$ as a Generalised Exponential Family

- The exponential function is replaced by the inverse of the function $(0, \infty) \ni y \mapsto y - 1 + \log y \in \mathbb{R}$:

\[
p(x) = \psi(a(x) + Z(a)) \quad \text{where} \quad a = \phi(P)
\]
\( M \) as a Generalised Exponential Family

- The exponential function is replaced by the inverse of the function \((0, \infty) \ni y \rightarrow y - 1 + \log y \in \mathbb{R}\):

\[
p(x) = \psi(a(x) + Z(a)) \quad \text{where} \quad a = \phi(P)
\]

- Convex, linear growth, bounded derivatives of all orders.
Mixture and Exponential Maps

• The maps \( m, e : M \rightarrow H \), defined by

\[
m(P) = p - 1 \quad \text{and} \quad e(P) = \log p - \mathbb{E}_\mu \log p
\]

are injective, but not homeomorphic (like \( \eta_\mu \) of \( \mathcal{E}(\mu) \))
Mixture and Exponential Maps

• The maps \( m, e : M \to H \), defined by
  \[
m(P) = p - 1 \quad \text{and} \quad e(P) = \log p - E\mu \log p
  \]
  are injective, but not homeomorphic (like \( \eta_\mu \) of \( \mathcal{E}(\mu) \))

• They satisfy:
  \[
  \mathcal{D}(P | Q) + \mathcal{D}(Q | P) = \langle m(P) - m(Q), e(P) - e(Q) \rangle_H
  \]

• So that
  \[
  \| m(P) - m(Q) \|^2_H + \| e(P) - e(Q) \|^2_H \leq \| \phi(P) - \phi(Q) \|^2_H
  \]

  and \( \mathcal{D}(P | Q) + \mathcal{D}(Q | P) \leq \frac{1}{2} \| \phi(P) - \phi(Q) \|^2_H \)
The Tangent Bundle

- Global Chart: $\Phi : TM \to H \times H$

$$\Phi(P, U) := (\phi(P), U\phi_p)$$
The Tangent Bundle

• Global Chart: $\Phi : TM \to H \times H$

$$\Phi(P,U) := (\phi(P), U\phi_p)$$

• $m$ and $e$ representations:

$$\Phi_m(P,U) := (\phi(P), Um_p) \in H \times H, \quad \Phi_e(P,U) := (\phi(P), Ue_p) \in H \times H$$

Injective but not homeomorphic
The Tangent Bundle

• Global Chart: \( \Phi : TM \rightarrow H \times H \)
  \[ \Phi(P,U) := (\phi(P), U\phi_P) \]

• \( m \) and \( e \) representations:
  \[ \Phi_m(P,U) := (\phi(P), Um_P) \in H \times H, \quad \Phi_e(P,U) := (\phi(P), Ue_p) \in H \times H \]
  \textbf{Injective but not homeomorphic}

• The Fisher metric: for \( U, V \in T_pM \)
  \[ \langle U, V \rangle_p := -UVD_p = \langle Um_P, Ve_P \rangle_H \quad \text{(Eguchi)} \]
The Tangent Bundle

- Global Chart: \( \Phi : TM \to H \times H \)
  \[
  \Phi(P, U) := (\phi(P), U\phi_P)
  \]
- \( m \) and \( e \) representations:
  \[
  \Phi_m(P, U) := (\phi(P), Um_p) \in H \times H, \quad \Phi_e(P, U) := (\phi(P), Ue_p) \in H \times H
  \]
  In**jective** but not homeomorph**omorphic**
- The Fisher metric: for \( U, V \in T_pM \)
  \[
  \langle U, V \rangle_p := -UV\mathcal{D}_p = \langle Um_p, Ve_p \rangle_H \quad \text{(Eguchi)}
  \]
- \( (T_pM, < \cdot, \cdot >) \) is an inner product space with
  \[
  \| U \|_p = \langle Um_p, Ue_p \rangle_H \leq \| U\phi \|_H
  \]
$e$ and $m$ Parallel Transport

- These are obtained by considering the inclusions:

$$\Phi_m(TM) \subset H \times H \quad \text{and} \quad \Phi_e(TM) \subset H \times H$$

together with the parallel transport on $H \times H$ defined by:

$$T_{a,b}(a,u) = (b,u)$$
\textbf{e and m Parallel Transport}

- These are obtained by considering the inclusions:

\[ \Phi_m(TM) \subset H \times H \quad \text{and} \quad \Phi_e(TM) \subset H \times H \]

- Together with the parallel transport on \( H \times H \) defined by:

\[ T_{a,b}(a,u) = (b,u) \]

- Like the \( m \) parallel transport on the maximal exponential model, they coincide with \( m \) parallel transport on the tangent bundle only in special cases.

- \( \alpha \)-parallel transports can be defined in the same way on statistical Hilbert bundles.
Submanifolds

Like the maximal exponential model, $M$ admits many useful submanifolds. For example…

• **Proposition 2**: If $N \subset M$ is a finite-dimensional exponential family, then it is a $C^\infty$-embedded submanifold of $M$, on which $m$, $e$ and $D$ are of class $C^\infty$
Submanifolds

Like the maximal exponential model, $M$ admits many useful submanifolds. For example...

• **Proposition 2**: If $N \subset M$ is a finite-dimensional exponential family, then it is a $C^\infty$-embedded submanifold of $M$, on which $m$, $e$ and $D$ are of class $C^\infty$

• **Example**: the non-singular Gaussian measures on $\mathbb{R}^m$ form a $C^\infty$-embedded submanifold of $M(\mathbb{R}^m, \mu)$, where

$$
\mu(dx) := 2^{-m} \exp(-|x|) \, dx
$$
Submanifolds

Like the maximal exponential model, $M$ admits many useful submanifolds. For example...

- **Proposition 2:** If $N \subset M$ is a finite-dimensional exponential family, then it is a $C^\infty$-embedded submanifold of $M$, on which $m$, $e$ and $\mathcal{D}$ are of class $C^\infty$.

- **Example:** the non-singular Gaussian measures on $\mathbb{R}^m$ form a $C^\infty$-embedded submanifold of $M(\mathbb{R}^m, \mu)$, where

  \[ \mu(dx) := 2^{-m} \exp(-|x|) \, dx \]

- Similar results hold for mixture models and $\alpha$-models.

- Subspaces of $H$ also provide natural submanifolds of $M$. 
Banach Variants

• The $\alpha$-divergences are twice differentiable on $M$.

• Greater regularity can be obtained by the use of stronger topologies on the model space: $L^\lambda(\mu)$, for $\lambda > 2$

• This enables the definition of $\alpha$-covariant derivatives on the statistical bundles mentioned above.

• Details in:

Nonlinear Filtering

• Markov “signal” process: \((X_t \in \mathbf{X}, t \in [0, \infty))\)
  - \((\mathbf{X}, \mu)\) is a metric space, with reference probability measure \(\mu\)
  - Eg. \(\mathbf{X} = \mathbb{R}^d\), \(\mu = \mathcal{N}(0, I)\)

• Partial “observation” process: \((Y_t \in \mathbb{R}, t \in [0, \infty))\)
  \[Y_t = \int_0^t h(X_s)ds + W_t\]

• Estimate \(X_t\) at each time \(t\) from its prior distribution \(P_t\) and the history of the observation:
  \[Y_0^t := (Y_s, s \in [0, t])\]

• Typical equation for the density:
  \[d\pi_t = A\pi_t dt + \pi_t (h - \bar{h}_t) d\bar{W}_t \quad \text{where } d\bar{W}_t := dY_t - \bar{h}_t dt\]
$M$-Valued Nonlinear Filters

Proposition 3: Under some technical conditions:

1. $P(\Pi_t \in M \text{ for all } t \geq 0) = 1$
\textbf{M-Valued Nonlinear Filters}

**Proposition 3:** Under some technical conditions:

1. \( \mathbf{P}(\Pi_t \in M \text{ for all } t \geq 0) = 1 \)

2. The coordinate representation \( \phi(\Pi) \) satisfies the following (infinite-dimensional) Itô equation

\[
\frac{d\phi(\Pi_t)}{d\Pi_t} = (u_t - \zeta_t)dt + v_t \, d\bar{W}_t
\]

where

\[
\begin{align*}
    u_t &:= \Lambda(1 + \pi_t^{-1})A\pi_t \\
    \zeta_t &:= \Lambda(h - \bar{h}_t)^2 / 2 \\
    v_t &:= \Lambda(\pi_t + 1)(h - \bar{h}_t)
\end{align*}
\]

\[
\Lambda f = \begin{cases} 
    f - E_{\mu}f & \text{if } f \in L^2(X, \mu) \\
    0 & \text{otherwise}
\end{cases}
\]
Components

• Since $H$ is of countable dimension, it admits a complete orthonormal basis $(\eta_i, i = 1, 2, 3, \ldots)$

• So the filter equations can be written in terms of the components:

$$\phi(\Pi_t)^i := \langle \phi(\Pi_t), \eta_i \rangle_H \text{ for } i = 1, 2, 3, \ldots$$
Components

- Since $H$ is of countable dimension, it admits a complete orthonormal basis $(\eta_i, i = 1, 2, 3, \ldots)$

- So the filter equations can be written in terms of the components:

$$\phi(\Pi_t)^i := \langle \phi(\Pi_t), \eta_i \rangle_H \quad \text{for} \quad i = 1, 2, 3, \ldots$$

- The Fisher metric can be expressed in terms of the $(\eta_i)$

$$\langle U, V \rangle_P = G(P)_{i,j} u^i v^j$$

where $G(P)_{i,j} = \langle D_i, D_j \rangle_P$, $(P, D_i) = \Phi^{-1}(\phi(P), \eta_i)$ and $U = u^i D_i$
Components

- Since $H$ is of countable dimension, it admits a complete orthonormal basis $(\eta_i, i = 1, 2, 3, \ldots)$
- So the filter equations can be written in terms of the components:
  \[ \phi(\Pi_t)^i := \langle \phi(\Pi_t), \eta_i \rangle_H \quad \text{for} \quad i = 1, 2, 3, \ldots \]
- The Fisher metric can be expressed in terms of the $(\eta_i)$
  \[ \langle U, V \rangle_p = G(P)_{i,j} u^i v^j \]
  where \( G(P)_{i,j} = \langle D_i, D_j \rangle_p \), \( (P, D_i) = \Phi^{-1}(\phi(P), \eta_i) \) and \( U = u^i D_i \)
- The basis can be chosen to suit the problem (wavelets)
- Truncated series could be used in approximations
Quadratic Variation

- Semimartingales on $M$ have well-defined quadratic variation in the Fisher metric; in particular

$$[\Pi]_t := \int_0^t G(\Pi_s)_{i,j} d[\phi(\Pi)^i, \phi(\Pi)^j]_s$$
Quadratic Variation

• Semimartingales on $M$ have well-defined quadratic variation in the Fisher metric; in particular

$$[\Pi]_t := \int_0^t G(\Pi_s)_{i,j} \, d[\phi(\Pi)^i, \phi(\Pi)^j]_s$$

• **Proposition 4**: Under the conditions of Proposition 3:

$$I(X; Y'_s | Y^s_0) = \frac{1}{2} \mathbb{E}([\Pi]_t - [\Pi]_s | Y^s_0)$$
Quadratic Variation

- Semimartingales on $M$ have well-defined quadratic variation in the Fisher metric; in particular

$$[\Pi]_t := \int_0^t G(\Pi_{s,i,j}) d[\phi(\Pi)^i, \phi(\Pi)^j]_s$$

- **Proposition 4:** Under the conditions of Proposition 3:

$$I(X; Y^t_s | Y^s_0) = \frac{1}{2} \mathbb{E}(\left[\Pi\right]_t - [\Pi]_s | Y^s_0)$$

- Results of this type are of interest in *Non-equilibrium Statistical Mechanics*, where interactions between systems set up “flows of entropy”.

NJN U of E 2016
Finite Dimensional Filters

• A number of filters are known to evolve on finite-dimensional exponential manifolds (Kalman-Bucy, Benes…)

NJN U of E 2016
Finite Dimensional Filters

• A number of filters are known to evolve on finite-dimensional exponential manifolds (Kalman-Bucy, Benes…)

• Proposition 5: Under some technical conditions, $\Pi$ is the unique strong solution of the following intrinsic Stratonovich equation on such a manifold:

$$\circ d\Pi_t = \left( U_t(\Pi_t) - \frac{1}{2} \nabla_{V_t}^{(-1)} V_t(\Pi_t) \right) dt + V_t(\Pi_t) \circ d\overline{W}_t$$

where $\nabla^{(-1)}$ is Amari’s $(-1)$-covariant derivative, and $U$ and $V$ are suitably regular, time-dependent vector fields.
Projections onto Submanifolds
(Brigo, Pistone, Hanzon, Le Gland, Armstrong…)

1. Choose a suitable $C^2$-embedded finite-dimensional submanifold $N \subset M$.
2. The tangent space $T_pN$ is complete w.r.t. the Fisher metric.
3. Evaluate $u_t - z_t$ and $v_t$ at points of $N$. (These are tangent vectors of $M$.)
4. Project onto $T_pN$ in the Fisher metric to obtain an evolution equation on $N$. 
Projections onto Submanifolds
(Brigo, Pistone, Hanzon, Le Gland, Armstrong…)

1. Choose a suitable $C^2$-embedded finite-dimensional submanifold $N \subset M$.
2. The tangent space $T_pN$ is complete w.r.t. the Fisher metric.
3. Evaluate $u_t - z_t$ and $v_t$ at points of $N$. (These are tangent vectors of $M$.)
4. Project onto $T_pN$ in the Fisher metric to obtain an evolution equation on $N$.

- The Hilbert manifold is very suited to this purpose
- One could also project in the model space metric
Details in:


Related Work


Related Work (cont.)


