Why are nonlinear filters stable?

Ramon van Handel

Department of Operations Research & Financial Engineering

PRINCETON UNIVERSITY

5th Oxford-Princeton Conference, March 27, 2009
Filtering models

Markov additive process \((X_t, Y_t)_{t \geq 0}\):
- \((X_t, Y_t)_{t \geq 0}\) is a Markov process with càdlàg paths.
- **Signal** \((X_t)_{t \geq 0}\) is itself a Markov process.
- **Observations** \((Y_t)_{t \geq 0}\) conditionally independent increments.

Standard examples:
1. White noise observations: \(dY_t = h(X_t) \, dt + \sigma \, dW_t\).
2. Counting observations: \(Y_t\) Poisson with rate \(\lambda(X_t)\).
3. Marked point process observations, stochastic volatility, etc.

Counterpart in discrete time: **Hidden Markov Models**.
Nonlinear filtering and stability

Definition

The **nonlinear filter** is the measure-valued process \((\pi_t)_{t \geq 0}\) such that \(\pi_t(f)\) is the optional projection of \((f(X_t))_{t \geq 0}\) on \((\mathcal{F}_t^Y)_{t \geq 0}\) for every \(f\).

**Notation:**

- \(\mathcal{F}_t^Y = \sigma\{Y_s : s \leq t\}\), etc. (suitably augmented).
- Under \(\mathbb{P}^\mu\), the signal has initial measure \(X_0 \sim \mu\). The corresponding filter is denoted \((\pi_t^\mu)_{t \geq 0}\), i.e., \(\pi_t^\mu(f) = \mathbb{E}^{\mu}(f(X_t)|\mathcal{F}_t^Y)\).

**Question**

When is the filter **stable**, i.e., \(\mathbb{E}^{\mu}(||\pi_t^\mu - \pi_t^\nu||) \xrightarrow{t \to \infty} 0\)?

- Problem lies at the heart of the asymptotic theory of nonlinear filters: key to ergodic theory and other uniform properties of the filter.
Example (discrete time)

\[ X_n = 0.9X_{n-1} + \beta_n, \quad Y_n = X_n + \gamma_n \]
Example (discrete time)

Kalman/SIS/SIS-R

\[ X_n = 0.9X_{n-1} + \beta_n, \quad Y_n = X_n + \gamma_n \]
Intuition

Filter stability is caused by two mechanisms:

1. When the signal is \textit{ergodic}, the filter should be also.
2. When the observations are sufficiently \textit{informative}, the resulting information gain should obsolete the prior measure.

In the special \textit{linear-Gaussian} case (Kalman filter), intuition can be made explicit: \textit{ergodic, observable, detectable} models.

\textbf{Goal:} develop a general theory.

- Proof in linear-Gaussian case is useless!
- Most results need very strong assumptions (uniform contraction).
- Ergodic case: all known \textit{general} results are based on a paper by Kunita (1971). However, the key step in his proof is incorrect.
- Results beyond the ergodic case very limited.
Ergodic signal: a general result

**Ergodicity Assumption**

The signal possesses an invariant probability measure \( \lambda \) such that
\[
\| P^z(X_t \in \cdot) - \lambda \|_{TV} \to 0 \quad \text{as} \quad t \to \infty \quad \text{for} \quad \lambda\text{-a.e.} \ z.
\]

**Nondegeneracy Assumption**

\[
P^\mu |_{\mathcal{F}_t^X \vee \mathcal{F}_t^Y} \sim P^\mu |_{\mathcal{F}_t^X \otimes \Phi|_{\mathcal{F}_t^Y}} \quad \text{for all} \quad t < \infty, \mu.
\]

**Theorem**

Suppose that the above assumptions hold. Then
\[
E^\mu (\| \pi^\mu_t - \pi^\lambda_t \|_{TV}) \to 0 \quad \text{iff} \quad \| P^\mu |_{\sigma(X_t)} - \lambda \|_{TV} \to 0.
\]
Idea of proof

Problem can be reduced to the case $\mu \ll \lambda$. We can prove:

$$E^\mu(\|\pi^\mu_t - \pi^\lambda_t\|_{TV}) =$$

$$E^\lambda\left(\left\| E^\lambda\left(\frac{d\mu}{d\lambda}(X_0) \bigg| \mathcal{F}^Y_\infty \vee \mathcal{F}^X_{[t,\infty[} \right) - E^\lambda\left(\frac{d\mu}{d\lambda}(X_0) \bigg| \mathcal{F}^Y_t \right) \left\| \right.\right).$$

By martingale convergence,

$$\bigcap_{t \geq 0} \mathcal{F}^Y_\infty \vee \mathcal{F}^X_{[t,\infty[} = \mathcal{F}^Y_\infty \implies E^\mu(\|\pi^\mu_t - \pi^\lambda_t\|_{TV}) \xrightarrow{t \to \infty} 0.$$ 

Wrong proof

$$(X_t)_{t \geq 0} \text{ ergodic} \implies \bigcap_{t \geq 0} \mathcal{F}^X_{[t,\infty[} \text{ is trivial} \implies \bigcap_{t \geq 0} \mathcal{F}^Y_\infty \vee \mathcal{F}^X_{[t,\infty[} = \mathcal{F}^Y_\infty.$$ 

This fundamental mistake is made in Kunita (1971)!
Idea of proof

Correct statement (von Weizsäcker 1983):

$$\bigcap_{t \geq 0} \mathcal{F}_Y \vee \mathcal{F}_X^{t, \infty} = \mathcal{F}_Y^{\infty} \quad \mathbb{P}^\lambda \text{-a.s.} \quad \iff \quad \bigcap_{t \geq 0} \mathcal{F}_X^{t, \infty} \quad \mathbb{P}^\lambda(\cdot | \mathcal{F}_Y^{\infty}) \text{-trivial} \quad \mathbb{P}^\lambda \text{-a.s.}$$

So, must prove that \((X_t)_{t \geq 0}\) is ergodic under \(\mathbb{P}^\lambda(\cdot | \mathcal{F}_\infty^Y)\).

Key ideas:

- \((X_t)_{t \geq 0}\) is a Markov pr. in a random environment under \(\mathbb{P}^\lambda(\cdot | \mathcal{F}_\infty^Y)\).
- Prove a general ergodic theorem for such processes.
- Use coupling, disintegration and time reversal methods to relate the ergodic properties under \(\mathbb{P}^\lambda(\cdot | \mathcal{F}_\infty^Y)\) to those under \(\mathbb{P}^\lambda\).
- Nondegeneracy enters in the last step.
Informative observations: a general result

**Definition**

Model is called **uniformly observable** if \( \forall \varepsilon > 0, \exists \delta > 0 \) such that

\[
\| P^\mu |_{\mathcal{F}_Y} - P^\nu |_{\mathcal{F}_Y} \|_{TV} < \delta \quad \text{implies} \quad \| \mu - \nu \|_{BL} < \varepsilon.
\]

Model is called **observable** if \( P^\mu |_{\mathcal{F}_Y} = P^\nu |_{\mathcal{F}_Y} \) implies \( \mu = \nu \).

**Theorem**

If the model is uniformly observable, then

\[
E^\mu (\| \pi^\mu_t - \pi^\nu_t \|_{BL}) \xrightarrow{t \to \infty} 0 \quad \text{whenever} \quad P^\mu |_{\mathcal{F}_Y} \ll P^\nu |_{\mathcal{F}_Y}.
\]

Moreover, if \( (X_t)_{t \geq 0} \) is Feller and takes values in a compact state space, then the conclusion already holds if the model is observable.

**Proof:** Martingale convergence arguments.
Verifying observability

How to prove (uniform) observability?

- **Finite state space**: observability reduces to linear algebra.
- **Kalman filter**: observability $\iff$ uniform observability.
- **Additive noise**: the model

$$dX_t = b(X_t) \, dt + g(X_t) \, dW_t, \quad dY_t = h(X_t) \, dt + \sigma \, dB_t,$$

is uniformly observable if $h$ is strongly invertible.

**Proposition**

Let $\mu, \nu, \xi \in \mathcal{P}(\mathbb{R}^d)$ and let $| \int e^{i \cdot k \cdot x} \, \xi(dx) | > 0$. Then

$$\forall \varepsilon > 0, \ \exists \delta > 0 \ \text{s.t.} \ \| \mu * \xi - \nu * \xi \|_{BL} < \delta \implies \| \mu - \nu \|_{BL} < \varepsilon.$$

**Proof**: basic ideas from Banach space theory and harmonic analysis.
A necessary and sufficient condition

Detectability Assumption

For every pair $\mu, \nu$ of initial measures, either
1. $P^\mu|_{\mathcal{F}_Y^\infty} \neq P^\nu|_{\mathcal{F}_Y^\infty}$; or
2. $\|P^\mu|_{\sigma(X_t)} - P^\nu|_{\sigma(X_t)}\|_{TV} \to 0$ as $t \to \infty$.

Theorem

Suppose that $(X_t)_{t \geq 0}$ is a finite state Markov process and that the observations are nondegenerate. Then the following are equivalent:
1. The detectability condition is satisfied.
2. $E^\mu(\|\pi^\mu_t - \pi^\nu_t\|_{TV}) \to 0$ whenever $P^\mu|_{\mathcal{F}_Y^\infty} \ll P^\nu|_{\mathcal{F}_Y^\infty}$.

- Detectability is necessary and sufficient!
- Very satisfying, but proof does not generalize (so far...)
Filter approximation: a general result

**Theorem**

Let \((\pi^N_k)_{k \geq 0}\), \(N \geq 1\) be a sequence of recursive approximations of the nonlinear filter \((\pi_k)_{k \geq 0}\). Suppose that the following assumptions hold:

1. The signal is ergodic and the observations are nondegenerate.
2. The one step transition probability \(\Pi^N\) of \((X_k, \pi^N_k)_{k \geq 0}\) converges to the transition probability \(\Pi\) of \((X_k, \pi_k)_{k \geq 0}\) uniformly on compacts.
3. The family \(\{\pi^N_k : k \geq 0, N \geq 1\}\) is tight.

Then \((\pi^N_k)_{k \geq 0}\) approximates \((\pi_k)_{k \geq 0}\) uniformly in time average:

\[
\lim_{N \to \infty} \sup_{T \geq 0} \mathbb{E} \left[ \frac{1}{T} \sum_{k=1}^{T} \|\pi^N_k - \pi_k\|_{BL} \right] = 0.
\]

Inspired by an argument of Budhiraja and Kushner (2001), but the new stability results are key to developing the technique in its generality.
Particle filters

- SIS-R algorithm satisfies condition 2, SIS violates it.
- To prove the approximation property, need “only” prove that the particle system is *tight*. This is surprisingly difficult!
- Tightness proofs for geometrically ergodic signals with either (1) bounded observations, or (2) radially unbounded observations.
- Significant improvement over previous results (Del Moral 2004), and at present the only approach that can feasibly be extended.
- Continuous time should be no problem; nonergodic case is a mystery.

\[ X_n = 0.9X_{n-1} + \beta_n, \quad Y_n = X_n + \gamma_n \]
Conclusion

- A surprisingly general asymptotic theory answers the basic question: *why are nonlinear filters stable?*

- Application: new insight into the performance of particle filters.

- Various open problems remain both in the fundamental theory and in applications (particle filters, stochastic control, statistical inference).

References at [http://www.princeton.edu/~rvan/](http://www.princeton.edu/~rvan/)