University of Reading

Stability of the optimal nonlinear filter

Lea Oljača, Jochen Bröcker, Tobias Kuna

May 1, 2020

University of Reading

Outline

Optimal filter and stability

Main result

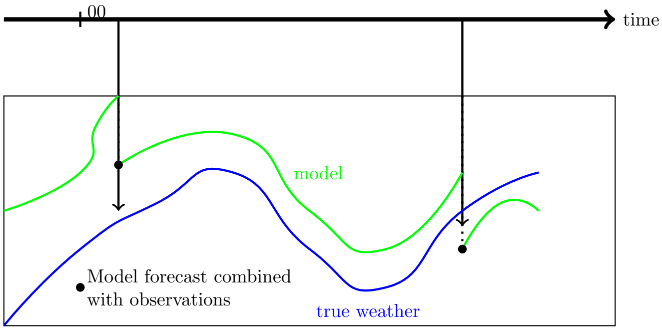
Background and literature

Uniformly hyperbolic dynamics

Idea of proof of main result

Appendix

Data assimilation (DA): initializing the weather forecast from observations



DA is uniquely challenging in weather forecasting

Data assimilation, a Bayesian perspective

Filtering problem/Data assimilation: recursively, at each time t_n

- estimate 'signal' X_n (on a polish space M)
- given all the observations $\{Y_k\}_{k \leq n} \in \mathbb{R}^d$ up to that time.

Optimal filter: optimal estimate of X_n, in the mean square sense.

- in Bayesian framework, it's a seq. of conditional probabilities

$$P_n := \mathbb{P}(X_n | Y_n, \ldots Y_1)$$

- for Gaussian, linear systems the optimal filter is KF
- explicit computation of the posterior is very computationally expensive in nonlinear systems
- in practice, approximation algorithms are used (e.g. the 3DVAR, EnKF, particle filters)

Optimal filter: an iterative two-step process

Let p_{n-1} be the density of the posterior at t_{n-1}, f the dynamics

- **prediction**

$$p^{+}_{n-1} = \mathscr{P} p_{n-1} = \mathbb{P}(X_n | Y_{n-1}, ..., Y_1)$$

  where $\mathscr{P}$ is the transfer operator of f, mapping a probability density of random variable X_{n-1} to the density of $f(X_{n-1})$. This becomes the new prior.

- **update** via Bayes:

$$p_n = \mathcal{B}(Y_n) p^{+}_{n-1}$$

  where $\mathcal{B}(Y_n)$ denotes multiplication with the likelihood and the normalisation and depends on Y_n

- **filtering operator:**

$$\tilde{\mathscr{L}}_{Y_n} := \mathcal{B}(Y_n)\mathscr{P}$$

  so that $p_n = \tilde{\mathscr{L}}_{Y_n} p_{n-1}$.

Stability is an important problem of filtering

Definition[Stability] Given any two initial prior distributions, say P_0 and Q_0, a norm $\|\cdot\|$, and distributions R_n of the observations, the filtering process is said to be stable

$$\lim_{n\to\infty} \|\tilde{\mathscr{L}}_n P_0 - \tilde{\mathscr{L}}_n Q_0\| = 0,$$

where $\tilde{\mathscr{L}}_n$ is shorthand for $\tilde{\mathscr{L}}_n = \tilde{\mathscr{L}}_{Y_n} \circ \ldots \circ \tilde{\mathscr{L}}_{Y_1}$.

Why stability is important:

- initial condition P_0 is required to initialise the filtering
- we don't know the correct initial distribution accurately
- stability with a certain decay rate even allows treatment of approximation errors

Main assumptions

- $X_n = f(X_{n-1})$, for $f : M \to M$ uniformly hyperbolic diffeomorphism, M a smooth manifold
- conditioned on X_n, the observations Y_n are i.i.d.
- Y_n are ergodic (if X_n are drawn from an ergodic distribution)
- assume there is a *likelihood function* g and measure ν s.t.

$$\mathbb{P}(Y_k \in A | X_k = x) = \int_A g(y, x)\nu(dy). \tag{1}$$

- g is Lipschitz continuous $\mathbb{P}$ almost surely and the Lipschitz constant is a tempered random variable

Filtering operator on $L^1(m)$

We can define a new 'filtering' operator that acts on any density $p(x) \in L^1(m)$ by

$$\mathscr{L}_{Y_n} p(x) = g(Y_n, x)\mathscr{P} p(x) \tag{2}$$

and

$$\tilde{\mathscr{L}}_{Y_n} p(x) = \frac{\mathscr{L}_{Y_n} p(x)}{\|\mathscr{L}_{Y_n} p(x)\|_1} \tag{3}$$

where the norm is taken in $L^1(m)$.

We define the filtering operator acting on measures as

$$\bar{\mathcal{L}}_\omega \mu(\psi) := \frac{\int g(\omega, x) \circ f \psi \circ f d\mu}{\int g(\omega, x) \circ f d\mu} \tag{4}$$

- $\mathscr{L}_{Y_n}$ **is a linear operator** related to $\mathscr{P}$ via the likelihood function g

Main result: Stability

Theorem[L.Oljaca, J. Bröcker, T.Kuna]: There exists a regular conditional probability measure $\mu : \Omega \times \mathscr{M} \to [0,1]$ and a set of full measure $\Omega_1 \subseteq \Omega$ such that for all $\omega \in \Omega_1$ and continuous ψ, it holds that

$$\bar{\mathcal{L}}_\omega \mu_\omega(\psi) = \mu_{T\omega}(\psi). \tag{5}$$

Furthermore, there exists a constant $\tilde{\beta} > 0$ such that for all strictly positive functions ϕ s.t $\log \phi$ is ν-Hölder continuous, all $\omega \in \Omega_1$ and $\hat{\mu}$-Hölder continuous $\psi : Q \to \mathbb{R}$, it holds that

$$\lim_{n\to\infty} n^{-1} \log \left| \int \psi \tilde{\mathscr{L}}_\omega^n \phi dm - \int \psi d\mu_{T^n\omega} \right| \leq -\tilde{\beta}.$$

Literature I: Stability has been investigated for various situations

The problem of stability can be studied in various contexts

- linear/nonlinear dynamics
- Gaussian/non-Gaussian priors/errors
- random/non-random signals
- optimal and non-optimal filters (e.g. 3DVAR and EnKF, particle filters).

For linear, random dynamics

- stability of KF holds under some broad conditions on the signal and observations [7]

For linear, deterministic systems (no model error)

- in context of DA, by Bouquet, Gurumoorthy, Apte, Carassi, Grudzien and Jones, 2017 [2].

University of Reading

Literature II: Stability for nonlinear, random dynamical systems

Most work has focused on stochastic dynamics and relies on mixing properties of the signal due to randomness

- based on work by Kunita[8], Ocone and Pardoux [12] show L^p type convergence. Exponential convergence shown for e.g. KF
- Atar and Zeituni [1] extend result by showing a.s. exponential stability in TV norm
- while Le Gland and Oudjane are able to further relax the ergodicity assumptions on the signal process [9]
- Tong and Van Handel [13], show that the Stochastic 2D N-S equations satisfy conditions in [5]. However does not provide a rate of convergence.

Literature III: Non-optimal filter stability

Question of whether they converge to the optimal filter or at least, of estimating the error.

- It has been shown that stability with a certain decay rate implies uniform convergence of the asymptotic approximation error, see e.g. Crisan and Haine [6] or [10], [11], [9] and [4].
- Stable filters can be approximated numerically, with errors that are bounded uniformly in time
- Furthermore, **Crisan et al** [4], show that a class of approximations is stable and converge uniformly to the optimal filter, <u>whether the filter itself is stable or not</u>.

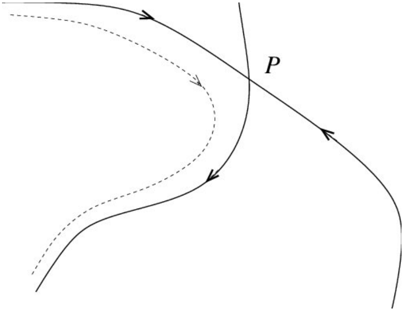
Literature IV: Deterministic nonlinear dynamics more challenging

As there is no mixing due to stochastic effects, any “forgetting” of the initial condition has to come from the dynamics

- Bröcker and Del Magno, 2017, show exponential stability for expanding maps for sufficiently smooth initial condition [3]

Our aim: extend analysis to deterministic signals of hyperbolic dynamical systems, using their chaotic nature

Uniformly hyperbolic dynamics



- $f : M \to f(M)$ to be a diffeomorphism on a manifold M
- dynamics at every point in $\Lambda \subset M$ (maximal invariant set) has a contracting and expanding direction, which are transversal
- Anosov diffeomorphism, Axiom A systems

University of Reading

Y_n is stationary and ergodic

Theorem: f uniformly hyperbolic as described above, then f admits a unique SRB measure μ_0.

Proof: See Viana [14], "Stochastic dynamics of deterministic systems."

From the above, it can be shown that:

- if $X_0 \sim \mu_0$, signal process generated by $X_n = f(X_{n-1})$ is ergodic,
- if Y_n is i.i.d. conditioned on X_n, it is also an ergodic process.

Can assume that Y_n is a stationary and ergodic process

- there is an ergodic $T : \Omega \to \Omega$ s.t.

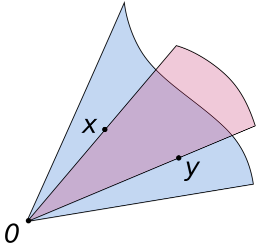
$$Y_n(\omega) = Y_1(T^{n-1}\omega)$$

University of Reading

Idea of proof

- Find a space on which the filtering operator is a contraction
- Use 'cones' of functions with Hilbert projective metric

- A *cone* in vector space E, is a subset $C \subset E \backslash \{0\}$ satisfying $v \in C, t > 0 \Rightarrow tv \in C$
- Convex if $t_1 v_1 + t_2 v_2 \in C$ for any $t_1, t_2 > 0$ and $v_1, v_2 \in C$
- A cone is proper if $\overline{C} \cap -\overline{C} = \emptyset$



Source: Wikipedia

Hilbert projective metric

Given $v_1, v_2 \in C$ we define

$$\alpha(v_1, v_2) = \sup\{t > 0; v_2 - tv_1 \in C\},$$

$$\beta(v_1, v_2) = \inf\{s > 0; sv_1 - v_2 \in C\}.$$

Hilbert Projective Metric

Let C be a proper convex cone. Given $v_1, v_2 \in C$, define the projective metric

$$\theta(v_1, v_2) = \log \frac{\beta(v_1, v_2)}{\alpha(v_1, v_2)}$$

with $\theta(v_1, v_2) = +\infty$ if $\alpha(v_1, v_2) = 0$ or $\alpha(v_1, v_2) = +\infty$.

θ induces a distance in the projective quotient of C:
$\theta(v_1, v_2) = 0 \iff v_1 = tv_2$ for some $t > 0$

Hilbert projective metric II

- E_1 and E_2 vector spaces
- $C_i \subset E_i$ proper convex cones
- $L : E_1 \to E_2$ be a linear operator such that $L(C_1) \subset C_2$

$$\begin{aligned}\alpha_1(v_1, v_2) &= \sup\{t > 0; v_2 - tv_1 \in C_1\} \\ &\leq \sup\{t > 0; L(v_2 - tv_1) \in C_2\} = \alpha_2(Lv_1, Lv_2)\end{aligned}$$

Also, $\beta_1(v_1, v_2) \geq \beta_2(Lv_1, Lv_2) \Rightarrow \theta_1(v_1, v_2) \geq \theta_2(L(v_1), L(v_2))$.

L is a strict contraction if diameter is finite

Let $D = \sup\{\theta_2(Lv_1, Lv_2); v_1, v_2 \in C_1\}$. If $D < +\infty$ then

$$\theta_2(Lv_1, Lv_2) \leq (1 - e^{-D})\theta_1(v_1, v_2).$$

Hilbert metric allows us to work with the linear part of the filtering operator as the normalization can be ignored

Idea of proof

Key idea (adapted from Viana [14]): Average densities along the stable direction against test functions

- along the stable manifolds we have contraction, which amplifies oscillations (or errors in initial density)
- must allow the density to become 'singular' on stable leaves
- expansion along unstable direction has effect of making densities smoother

Idea of proof: back to Solenoid example

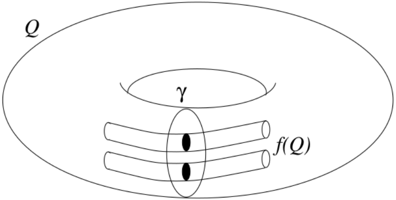


Figure 4.1: The solenoid on the solid torus Q

Source: M. Viana, Stochastic dynamics of deterministic systems

- contraction along the B^2 disks (foliation of stable leaves); for the solenoid uniquely defined by the $\theta \in S^1$
- expansion along the $S^1 x\{z\}$, for any $z \in B^2$
- every leaf γ has two pre-images, γ_j, also leaves, such that $f(\gamma_j) \subset \gamma$, for $j = 1, 2$

Idea of proof: back to Solenoid example II

- γ and δ two nearby local stable leaves
- $\pi : \gamma \to \delta$ be the projection along the unstable direction, that is
$$x = (\theta_1, z) \in \gamma \to \pi x = (\theta_2, z) \in \delta$$
- let $d(x, \pi x)$ is distance measured along the unstable leaves
- define a distance on the space of stable leaves
$$d(\gamma, \delta) = \sup_{x \in \gamma} d(x, \pi(x))$$
- **map induced by f on the stable leaves is expanding for this distance**
$$d(\gamma_j, \delta_j) \leq \lambda_u d(\gamma, \delta),$$
where $\lambda_u < 1$.

Idea of proof

Define a metric space of densities

- (Cones with Hilbert projective metrics) on which the operator $\mathscr{L}_\omega$ is a contraction

$$\mathscr{C}(c, a, \mu, \nu) := \{\phi; \frac{\int_\gamma \phi\rho}{\int_\delta \phi\pi^*\rho} \leq e^{cd(\gamma,\delta)^\nu}, \forall\gamma,\delta \in \Gamma, \rho \in \mathcal{D}(a,\mu,\gamma)\}, \tag{6}$$

where we define the density $\pi^*\rho(y) := \rho(\pi(y))|\det D\pi(y)|$.
We define the (random) cones

$$\mathscr{L}_\omega \mathscr{C}(c_\omega, a_\omega, \mu, \nu) \subset \mathscr{C}(c_{T\omega}, a_{T\omega}, \mu, \nu).$$

Idea of proof and concluding remarks

- we would like to use a fixed point theorem to deduce asymptotic convergence to a density
- however, clearly we cannot have completeness in the cones, as the densities become singular on stable leaves
- use a 'weak' fixed point theorem to deduce that there is convergence to a distribution, regardless of initial density (up to some Hölder regularity)

In conclusion, dynamics can be sufficiently mixing so that filter forgets initial condition

- initial condition needs to be a density with some smoothness

University of Reading

Appendix

Assumptions and notation

- conditioned on X_n, the observations Y_n are i.i.d.

$$\mathbb{P}(Y_n, ..., Y_1|\mathcal{X}_n) = \Pi_{k=1}^{n}\mathbb{P}(Y_k|X_k) \tag{7}$$

- assume there is a *likelihood function* g and measure ν s.t.

$$\mathbb{P}(Y_k \in A|X_k = x) = \int_A g(y, x)\nu(dy). \tag{8}$$

- denote the transition kernel of X_n

$$K(z, B) = \mathbb{P}(X_n \in B|X_{n-1} = z) \tag{9}$$

- let $P_0(B) = \mathbb{P}(X_0 \in B)$ be an the initial prior distribution.

Filtering equations I

Proposition Under assumptions above, the filtering process $P_n := \mathbb{P}(X_n | Y_n, \dots Y_1)$ satisfies the following recursion

$$P_n(\psi) = \frac{\int_M \psi(x) g(Y_n, x) dP^+_{n-1}(x)}{\int_M g(Y_n, x) dP^+_{n-1}(x)}, \tag{10}$$

where we define P^+_{n-1} as

$$P^+_{n-1}(\psi) = \int_M \int_M \psi(x) K(z, dx) dP_{n-1}(dz) \tag{11}$$

for all continuous $\psi : M \to \mathbb{R}$.

Filtering equations II: deterministic dynamics

Assume that X_n is deterministic, $X_n = f(X_{n-1})$, for $f : M \to M$, M a smooth manifold

- X_n is completely determined by X_0; all uncertainty comes from the uncertainty in the initial condition

Proposition Suppose P_n has a density $p_n(x)$ w.r.t. to Riemannian volume m. Then also P_{n+1} has a density $p_{n+1}(x)$ given by

$$p_{n+1}(x) = \frac{g(Y_n, x)\mathscr{P} p_n(x)}{\int_M g(Y_n, x)\mathscr{P} p_n(x) dm(x)} \tag{12}$$

where $\mathscr{P}$ is the transfer operator mapping a probability density of random variable X to the density of $f(X)$.

Uniformly hyperbolic dynamics

- $f : M \to f(M)$ to be a diffeomorphism on a manifold M
- Λ the maximal invariant set
- Λ is a uniformly hyperbolic set if there exists a splitting of the tangent bundle to M on Λ into stable and unstable directions;

$$T_\Lambda M = E^s_\Lambda \oplus E^u_\Lambda,$$

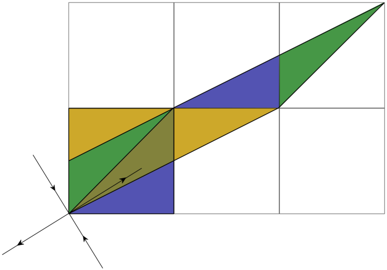
and a constant $\lambda_0 < 1$ such that for some Riemannian metric $\|\cdot\|$ on M it holds that

1. $Df(x) \cdot E^s_x = E^s_{f(x)}$ and $Df^{-1}(x) \cdot E^u_x = E^u_{f^{-1}(x)}$
2. $\|Df(x)|E^s_x\| \leq \lambda_0$ and $\|Df^{-1}(x)|E^u_x\| \leq \lambda_0$ for every $x \in \Lambda$.

Arnold's cat map

Hyperbolic toral automorphism $F : \mathbb{T}^2 \to \mathbb{T}^2$ defined by

$$F(x) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} x \mod 1 \tag{13}$$



Source: Wikipedia

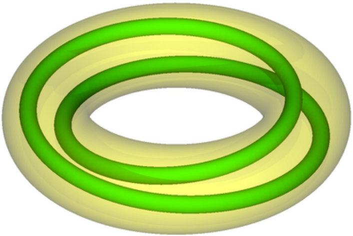
University of Reading

Solenoid or Smale -Williams attractor

In this case the manifold is a solid torus given by $\mathbf{T} = S^1 \times B^2$ and the dynamics is produced by the map $f : \mathbf{T} \to \mathbf{T}$ given by

$$S^1 \times B^2 \ni (\theta, z) \to (2\theta \mod \mathbb{Z}, \rho e^{2\pi i \theta} + \lambda z),$$

with suitable constants $\lambda < \rho$ and $\lambda + \rho < 1$ producing a contraction in the B^2 direction.



Source: Wikipedia

R. Atar and O. Zeitouni. "Exponential stability for nonlinear filtering". In: Ann. Inst. H. Poincaré Probab. Statist. 33.6 (1997), pp. 697–725. URL: https://doi.org/10.1016/S0246-0203(97)80110-0.

M. Bocquet et al. "Degenerate Kalman Filter Error Covariances and Their Convergence onto the Unstable Subspace". In: SIAM/ASA Journal on Uncertainty Quantification 5.1 (2017), pp. 304–333. URL: https://doi.org/10.1137/16M1068712.

J. Bröcker and G. Del Magno. "Asymptotic stability of the optimal filter for random chaotic maps". In: Nonlinearity 30.5 (2017), pp. 1809–1833.

D. Crisan, A. Lopez-Yela, and J. Miguez. Stable approximation schemes for optimal filters. 2018.

R. van Handel. "The stability of conditional Markov processes and Markov chains in random environments". In:

University of Reading

Ann. Probab. 37.5 (2009), pp. 1876–1925. URL: https://doi.org/10.1214/08-AOP448.

K. Heine and D. Crisan. “Uniform Approximations of Discrete-Time Filters”. In: Advances in Applied Probability 40.4 (2008), pp. 979–1001. URL: http://www.jstor.org/stable/20443621.

A. Jazwinski. Stochastic processes and filtering theory. Mathematics in science and engineering 64. New York, NY [u.a.]: Acad. Press, 1970. XIV, 376. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+021832242&sourceid=fbw_bibsonomy.

H. Kunita. “Asymptotic behavior of the nonlinear filtering errors of Markov processes”. In: Journal of Multivariate Analysis 1.4 (1971), pp. 365 –393. URL: http://www.sciencedirect.com/science/article/pii/0047259X71900157.

F. Le Gland and N. Oudjane. "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters". In: Ann. Appl. Probab. 14.1 (2004), pp. 144–187. URL: https://doi.org/10.1214/aoap/1075828050.

P. D. Moral. "A Uniform Convergence Theorem for the Numerical Solving of the Nonlinear Filtering Problem". In: Journal of Applied Probability 35.4 (1998), pp. 873–884. URL: http://www.jstor.org/stable/3215523.

P. D. Moral and A. Guionnet. "On the stability of interacting processes with applications to filtering and geneticalgorithms". In: Annales de l'Institut Henri Poincare (B) Probability and Statistics 37.2 (2001), pp. 155 –194. URL: http://www.sciencedirect.com/science/article/pii/S0246020300010645.

D. Ocone and E. Pardoux. "Asymptotic Stability of the Optimal Filter with Respect to Its Initial Condition". In: SIAM Journal on Control and Optimization 34.1 (1996), pp. 226–243. URL: https://doi.org/10.1137/S0363012993256617.

X. T. Tong and R. van Handel. "Conditional ergodicity in infinite dimension". In: Ann. Probab. 42.6 (Nov. 2014), pp. 2243–2313. URL: https://doi.org/10.1214/13-AOP879.

M Viana. Stochastic Dynamics of Deterministic Systems. Instituto de Matemtica Pura e Aplicada, 1997.