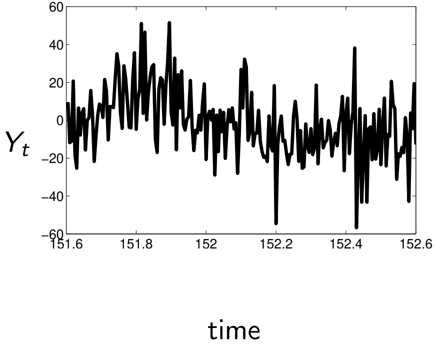
A very brief introduction to Data Assimilation

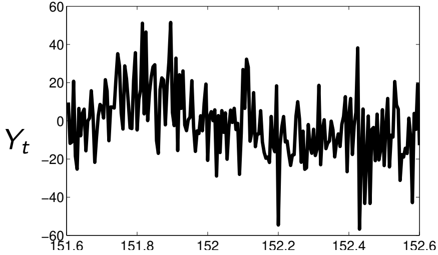
Jochen Bröcker

University of Reading, UK

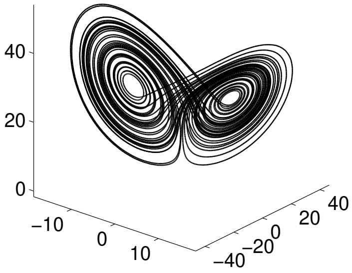
May 7, 2024

Content

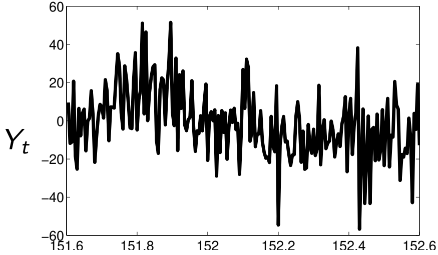

60
40
20
0
−20
−40
−60
151.6
151.8
152
152.2
152.4
152.6
Y_t
time



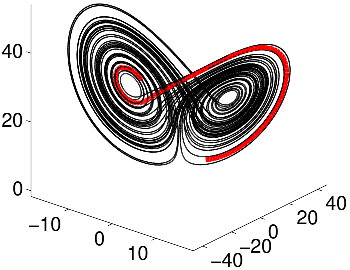
time



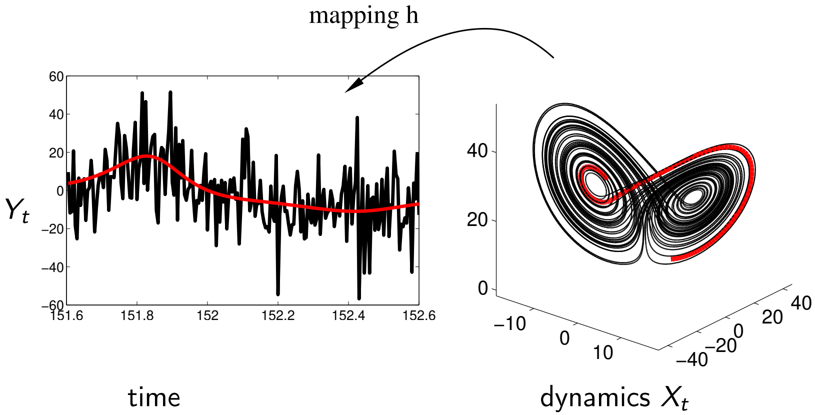
dynamics X_t



time



dynamics X_t


mapping h
60
40
20
0
–20
–40
–60
151.6
151.8
152
152.2
152.4
152.6
Y_t
time
40
20
0
–10
0
10
–40
–20
0
20
40
dynamics X_t

$$Y_t = h(X_t)$$

Problem of Data Assimilation

- The *signal process* $\{X_n, n \in \mathbb{N}_0\}$ is a sequence in some state space E (polish);
- the *observation process* $\{Y_n, n \in \mathbb{N}\}$ is a sequence in $\mathbb{R}^d$ (or finite dim. subspace of E).

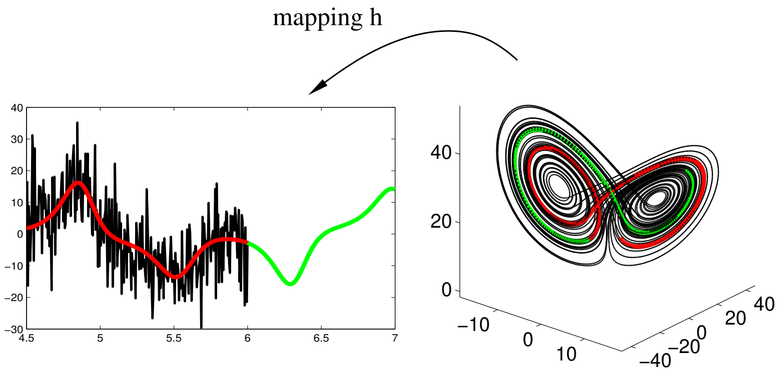
The signal process is *not* observed directly.

Goal of Data Assimilation

"Estimate" X_n from $\{Y_1, \ldots, Y_n\}$ for $n = 1, 2, \ldots$.

Motivation: diagnostics, model evaluation, forecasting

- Estimated trajectories provide a reconstruction of past climate;
- Trajectories that fit well might be indicative of a good model;
- Initial conditions for forecasts.



Different perspectives of data assimilation

Deterministic perspective Signal and observation processes $\{(X_n, Y_n), n \in \mathbb{N}\}$ are solution of some deterministic dynamical system with unknown initial condition. Aim is to estimate X_n as some function $\xi_n = \xi_n(Y_1, \ldots, Y_n)$ $\forall n \in \mathbb{N}$.

Stochastic perspective Signal and observation processes $\{(X_n, Y_n), n \in \mathbb{N}_0\}$ are stochastic processes on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Aim is to compute the **filtering process**, i.e. the conditional Prob'ty

$$\rho_n := \mathbb{P}(X_n \in .|Y_1, \ldots, Y_n) \qquad \forall n \in \mathbb{N}.$$

Basic setup *(Hidden Markov Model)*

- Signal process is homogenous stationary Markov process;
- Observation process is *conditionally independent* given $\{X_n\}$ i.e.

$$\mathbb{P}(Y_1, \dots, Y_n | X_1, \dots, X_n) = \prod_{k=1}^{n} \mathbb{P}(Y_k | X_k).$$

- *Nondegenerate observations:* we have

$$\mathbb{P}(Y_k \in B | X_k) = \int_B g(y, X_k) \, \mathrm{d}\nu(y).$$

for some sigma–finite measure ν on $\mathbb{R}^d$ and a *likelihood* $g : \mathbb{R}^d \times E \to \mathbb{R}_{\geq 0}$.

Iterative representation of filtering process

The filtering process $\{\rho_n, n \in \mathbb{N}\}$ is a random process in $\mathcal{P}_E$, the set of Borel distributions over E. For HMM's the filtering process satisfies

$$\begin{aligned} \rho_0 &= \mathbb{P}(X_0 \in .), \\ \rho_n &= \mathcal{L}_{Y_n}\rho_{n-1}, \end{aligned} \tag{1}$$

with operators $\mathcal{L}_y : \mathcal{P}_E \to \mathcal{P}_E$, for $y \in \mathbb{R}^d$. (Precise form depends on specific problem.)

For $m \leq n \in \mathbb{N}$, write $\mathcal{L}_{Y_{m:n}} := \mathcal{L}^{Y_n} \circ \ldots \circ \mathcal{L}^{Y_m}$.

Mathematical questions

Accuracy Does ρ_n concentrate strongly on X_n, the true state of the system?

Stability Does ρ_n depend on ρ_0 for large n, i.e. does

$$d(\mathcal{L}_{Y_{1:n}}\pi^{(1)}, \mathcal{L}_{Y_{1:n}}\pi^{(2)}) \to 0$$

hold in a suitable sense for $\pi^{(1,2)} \in \mathcal{P}_E$?

Robustness What is the influence of model misspecification on the filtering process?

Numerics How to efficiently implement and approximate filtering?

Existing results

Stability:

Strongly mixing signal processes
: Ocone and Pardoux [1996], Kunita [1971], Atar and Zeitouni [1997], LeGland and Oudjane [2004]; see also van Handel [2009] (very general result but no rate).

Linear state space models with Gaussian errors
: Classical results, see e.g. Anderson and Moore [2012]. For deterministic signal process: Bocquet et al. [2017] .

Very informative observations
: Crisan and Heine [2008].

Implementation:

Kalman filter
: is an exact solution for linear state space models with Gaussian errors. *Vast* number of "extensions" to nonlinear systems.

Sequential Monte Carlo
: are particle approximations [van Leeuwen et al., 2019, Crisan and Rozovskii, 2011]

Very few results on accuracy and robustness.

Deterministic signal process

Signal process $\{X_n, n \in \mathbb{Z}\}$ is given through

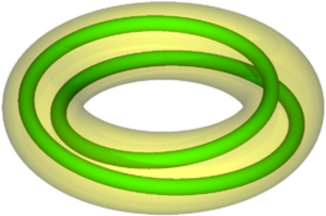
$$X_{n+1} = f(X_n), \qquad \mathbb{P}(X_0 \in .) = \rho_0,$$

where

- $f : M \to M$ diffeo on compact Riemannian M'fold M;
- f is solenoid–type and transitive.
- ρ_0 is the (uniquely defined) SRB–measure.

The likelihood $g(y, .)$ is assumed log–Hölder continuous with integrable coefficient.

Filter dynamics on densities



Smale–Williams Attractor or *Solenoid* (Source: Wikipedia).

Representation of $\mathcal{L}_y$ on densities (wrt. m)

$$\mathscr{L}^y \phi(z) = \frac{g(z, y)\, f_* \phi(z)}{\int_M \cdots \,\mathrm{d}m(z)},$$

Main result

Theorem [J.B. and Del Magno, 2017, Oljača, Kuna, and J.B., 2021]

There exist $\Omega_1 \subset \Omega$ with $\mathbb{P}(\Omega_1) = 1$ and positive α, β, so that

- for all $\omega \in \Omega_1$,
- for all densities $\phi : U \to \mathbb{R}_{>0}$ α-log Hölder cont',
- for all functions $\psi : U \to \mathbb{R}$ α-Hölder cont',

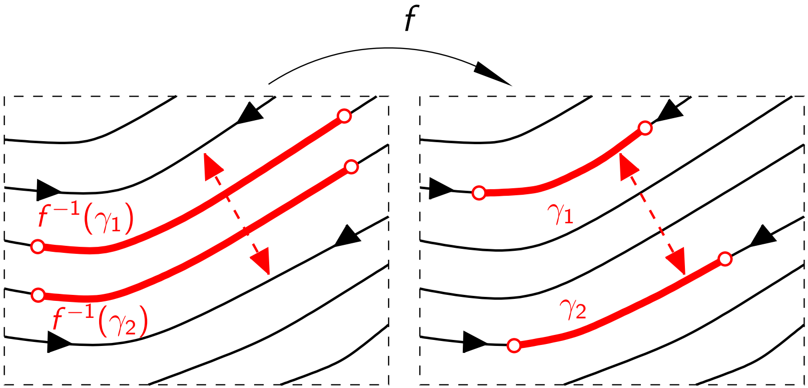
we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \left| \int \psi \cdot \mathscr{L}_{Y_{1:n}} \phi \, \mathrm{d}m - \mathbb{E}(\psi(X_n) | Y_{1:n}) \right| \leq -\beta.$$

If ψ is merely continuous, the convergence still takes place albeit not necessarily with exponential rate. (Convergence in L^1 holds in the case of uniformly expanding dynamics.)

Regularity of densities for hyperbolic dynamics

Expanding and contracting directions!



stable foliation under f with local stable leaves γ_1, γ_2.

Conclusions for part II

- Stochastic interpretation of Data Assimilation leads to concept of *filtering*;
- Several interesting mathematical questions regarding properties of the filtering process (stability, robustness etc);
- Example: stability result for expansive and solenoid–type dynamics (extension to Anosov systems is work in progress);
- Approximation and implementation of filters is *huge* field not touched upon.

State space model

Signal process $\{X_n \in E, n \in \mathbb{N}\}$ is given through

$$X_{n+1} = f(X_n) \qquad X_0 \text{ unknown.}$$

Here f might be time–τ flow of differential equation.

Observations are given by

$$Y_n = h(X_n), \qquad \text{with} \qquad h : E \mapsto \mathbb{R}^d.$$

Observer problem

For every $n \in \mathbb{N}$, approximate X_n as a function $\xi_n = \xi_n(Y_1, \ldots, Y_n)$.

Mathematical questions

The observer problem is well studied in discrete and continuous time (especially for finite dimensional dynamics).

Mathematical questions:

- Asymptotic behaviour for large time;
- Robustness to model misspecification;
- Stability with respect to initial condition;
- Sensitivity to observational noise.

The last point overlaps with the stochastic perspective.

Linear error feedback

We attempt to reconstruct $\{X_n, n \in \mathbb{N}\}$ by computing estimates $\{\xi_n \in H, n \in \mathbb{N}\}$ recursively as follows:
At time n, let

$$\begin{aligned} \xi^-_{n+1} &= f(\xi_n), && \text{(prediction step)} \\ \xi_{n+1} &= \xi^-_{n+1} + K_n \underbrace{\left(Y_{n+1} - h(\xi^-_{n+1})\right)}_{\textit{error}}, && \text{(update step)}, \end{aligned}$$

with some *feedback gain* K_n. In Data Assimilation parlance

- ξ^-_n is called *Background*
- ξ_n is called *Analysis*

How to find "optimal" K_n

In ***3DVar***, the analysis ξ_n (for given background) is found by minimising

$$J(z) = (\xi_n^- - z)^T B^{-1}(\xi_n^- - z) + (Y_n - h(z))^T S^{-1}(Y_n - h(z)),$$

which has some probabilistic motivation. For linear $h(z) = Hz$, this gives

$$\xi_n = (1 - KH)\xi_n^- + KY_n,$$

with "optimal" Kalman Gain K depending on B, S, H.

- Brett et al. [2013] consider 3DVar and other systems (Lorenz-type) with bounded noise.
- Blömker et al. [2013] consider 3DVar in continuous time.
- *Vast* literature on 3DVar in applications.

Examle: 2D Navier Stokes

On torus $\mathbb{T} = [0,1] \times [0,1]$ with periodic BC, consider Navier–Stokes

$$\partial_t u - \nu \Delta u + (u \cdot \nabla) u + \nabla p = \phi,$$
$$\nabla \cdot u = 0,$$

Here $E = L^2(\mathbb{T})$, and we let f, the time–τ flow map of Navier–Stokes.
Observations: Let $H : E \to E$ be orthogonal projection onto subspace E_N spanned by first N Fourier modes. We consider observations of the form

$$Y_n = HX_n \qquad n \in \mathbb{N},$$

and $K = 1$.

Movies (without observational noise)

Movie 1: 3 by 3 modes $nu = 2 \cdot 10^{-5}$ dt = 0.1	Movie 4: 3 by 3 modes $nu = 2 \cdot 10^{-8}$ dt = 0.2
Movie 2: 7 by 7 stations $nu = 2 \cdot 10^{-5}$ dt = 0.1	Movie 5: 3 by 3 modes $nu = 2 \cdot 10^{-6}$ dt = 0.2
Movie 3: 5 by 5 stations $nu = 2 \cdot 10^{-7}$ dt = 0.1	Movie 6: 5 by 5 modes $nu = 2 \cdot 10^{-6}$ dt = 0.2

Previous work on this behaviour

- Foiaș and Prodi [1967] show that the solution of 2D-NS is fully determined by its projection onto finitely many degrees of freedom.
- Jones and Titi [1992] carry this further (quantitative analysis).
- Hayden et al. [2011] apply this in the context of data assimilation.

Error dynamics with noise

We consider observations *with noise*

$$Y_n = HX_n + \sigma w_n, \qquad n \in \mathbb{N},$$

where $\{w_n, n \in \mathbb{N}\}$ i.i.d. random variables (with values in $E_N = \text{range}(H)$). We get error dynamics

$$\delta_{n+1} = X_{n+1} - \xi_{n+1} = (1 - H)\,(f(X_n) - f(\xi_n)) - \sigma w_{n+1}.$$

Here we look for **Individual bounds**

$$\limsup_{n \to \infty} (\|\delta_n\| - c_n \sigma) \leq 0$$

where c_n is a stationary process.

A local squeezing property

Write $Q = 1 - H$.

Local squeezing property [Hayden et al., 2011]

There are α, β (dep. on τ, N, R) so that

1. whenever $\|u\|^2 \leq R$ and $\|v\|^2 \leq R$ then

$$\|Q\left(f(u) - f(v)\right)\|^2 \leq \alpha \|Q(u - v)\|^2 + \beta \|H(u - v)\|^2,$$

2. for each R there is N, τ so that $\alpha < 1$.

Without noise, this would imply $\|\delta_n\|^2 \leq \alpha^n \|\delta_0\|^2 \to 0$, **if** we have independent proof of the apriori bounds.

Synchronisation with noise

Simple induction on the local squeezing property yields

$$\|Q\delta_n\|^2 \leq \alpha_m^n \|Q\delta_m\|^2 + \sigma^2 \sum_{k=m}^{n-1} \alpha_k^{n-1} \beta_k \|w_k\|^2$$

with $\alpha_m^n := \prod_{l=m}^{n-1} \alpha_l$ and $\alpha_n := \alpha(\tau, P, \|X_n\| \vee \|\xi_n\|)$.

Problem: $\|X_n\|$ and $\|\xi_n\|$ are unbounded! A priori bounds replaced by asymptotic stationarity of $\{X_n\}$ and $\{\xi_n\}$.

Accuracy of data assimilation in 2D–NS with unbounded observation noise

Theorem [Oljača, J.B., and Kuna, 2019]

Fix $\sigma_0 > 0$. Then

1. The stochastic processes $\{\alpha_n\}, \{\beta_n\}$ are stationary and ergodic with $\mathbb{E}(\log_+ \beta_0) < \infty$.
2. Suppose that $\mathbb{E}(\|w_n\|^3) < \infty$. Then for τ small enough and N large enough (i.e. sufficiently many modes observed) we have $\mathbb{E}(\alpha_n) < 1$.

This implies that there is a stationary process c_n depending only on σ_0 so that for all $\sigma \leq \sigma_0$ we have

$$\|\delta_n\|^2 \leq c_n \sigma^2$$

whenever n is sufficiently large.

Accuracy of data assimilation in 2D–NS with unbounded observation noise (short version)

Theorem [Oljača, J.B., and Kuna, 2019]

Fix $\sigma_0 > 0$ and suppose that $\mathbb{E}(\|w_n\|^3) < \infty$. Then for τ small enough and N large enough (i.e. sufficiently many modes observed), there is a stationary process c_n depending only on σ_0 so that for all $\sigma \leq \sigma_0$ we have

$$\|\delta_n\|^2 \leq c_n \sigma^2$$

whenever n is sufficiently large.

Conclusions for part III

- Data Assimilation from a deterministic perspective strongly linked to Observer Problem;
- In operational practice, most algorigthms employ linear error feedback, with some probabilistic motivation for the feedback gain;
- In analysis of 2D Navier Stokes, local squeezing property is fundamental and describes amount of "information" contained in observations (depending on the viscosity and attractor size).
- A number of ideas presented apply to other dissipative dynamical systems (such as Lorenz63 and Lorenz96).

References I

Thank you!

B.D.O. Anderson and J.B. Moore. *Optimal filtering*. Courier Corporation, 2012.

Rami Atar and Ofer Zeitouni. Exponential stability for nonlinear filters. *Ann. Inst. H. Poincaré Prob. Statist.*, 36:691–725, 1997.

D. Blömker, K. Law, A. M. Stuart, and K. C. Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the Navier-Stokes equation. *Nonlinearity*, 26(8):2193–2219, 2013. ISSN 0951-7715. doi: https://doi.org/10.1088/0951-7715/26/8/2193.

Marc Bocquet, Karthik S. Gurumoorthy, Amit Apte, Alberto Carrassi, Colin Grudzien, and Christopher K. R. T. Jones. Degenerate Kalman filter error covariances and their convergence onto the unstable subspace. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):304–333, 2017. doi: 10.1137/16M1068712. URL https://doi.org/10.1137/16M1068712.

References II

CEA Brett, Kei Fong Lam, KJH Law, DS McCormick, MR Scott, and AM Stuart. Accuracy and stability of filters for dissipative pdes. *Physica D: Nonlinear Phenomena*, 245(1):34–45, 2013.

J.B. and Gianluigi Del Magno. Asymptotic stability of the optimal filter for random chaotic maps. *Nonlinearity*, 2017.

D. Crisan and K. Heine. Stability of the discrete time filter in terms of the tails of noise distributions. *J. Lond. Math. Soc. (2)*, 78(2):441–458, 2008. ISSN 0024-6107. doi: 10.1112/jlms/jdn032. URL https://doi.org/10.1112/jlms/jdn032.

Dan Crisan and Boris Rozovskii, editors. *The Oxford handbook of nonlinear filtering*. Oxford University Press, 2011.

C. Foiaș and G. Prodi. Sur le comportement global des solutions non-stationnaires des équations de Navier-Stokes en dimension 2. *Rendiconti del Seminario Matematico della Universitá di Padova*, 39:1–34, 1967. ISSN 0041-8994.

References III

Kevin Hayden, Eric Olson, and Edriss S. Titi. Discrete data assimilation in the lorenz and 2d navier–stokes equations. *Physica D: Nonlinear Phenomena*, 240(18):1416 – 1425, 2011. doi: http://dx.doi.org/10.1016/j.physd.2011.04.021.

Don A. Jones and Edriss S. Titi. On the number of determining nodes for the 2D Navier-Stokes equations. *J. Math. Anal. Appl.*, 168(1):72–88, 1992. ISSN 0022-247X. doi: 10.1016/0022-247X(92)90190-O. URL http://dx.doi.org/10.1016/0022-247X(92)90190-O.

Hiroshi Kunita. Asymptotic behavior of the nonlinear filtering errors of markov processes. *Journal of Multivariate Analysis*, 1:365–393, 1971.

F. LeGland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert matric and application to particle filters. *The Annals of Applied Probability*, 14(1):144–187, 2004.

Daniel Ocone and Etienne Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim.*, 34(1):226–243, 1996. ISSN 0363-0129. doi: 10.1137/S0363012993256617. URL https://doi.org/10.1137/S0363012993256617.

References IV

Lea Oljača, Tobias Kuna, and J.B. Stability and asymptotic properties of the optimal filter for signals with deterministic hyperbolic dynamics. *arXiv:2103.01190*, 2021.

Lea Oljača, J.B., and Tobias Kuna. Almost sure error bounds for data assimilation in dissipative systems with unbounded observation noise. *SIAM Journal on Applied Dynamical Systems*, 17(4), August 2019.

Ramon van Handel. The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.*, 37(5):1876–1925, 2009. ISSN 0091-1798. doi: 10.1214/08-AOP448. URL http://dx.doi.org/10.1214/08-AOP448.

Peter Jan van Leeuwen, Hans R. Künsch, Lars Nerger, Roland Potthast, and Sebastian Reich. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 145(723): 2335–2365, 2019. doi: 10.1002/qj.3551.