Mathematical Foundation of Data Assimilation

Sebastian Reich

Universität Potsdam/ University of Reading

RISDA, January 24th, 2018

Outline

Part 1. Foundation of Bayesian inference

Part 2. Filtering and Smoothing for State Space Models

Part 3. Ensemble Kalman filtering and smoothing

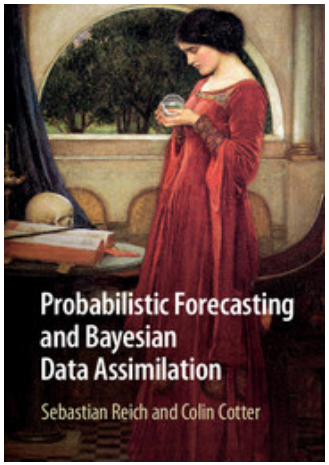
Part 4. Particle filters for high-dimensional systems

My Background

Electrical engineer and applied mathematician by training.

Research interests in:

- numerical analysis
- Hamiltonian and molecular dynamics
- computational fluid dynamics
- data assimilation

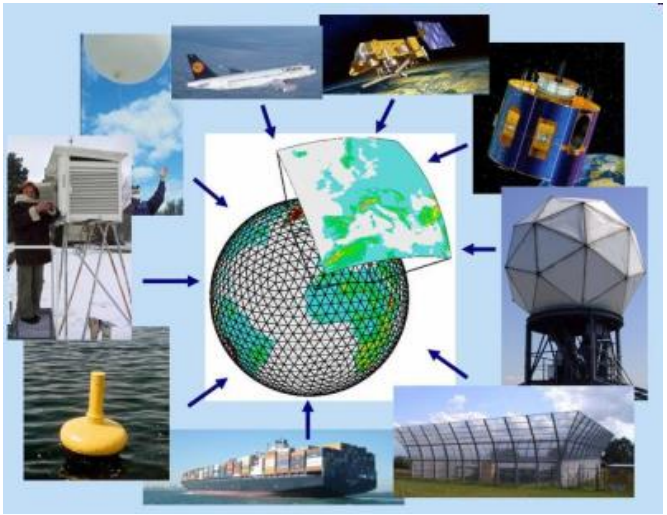


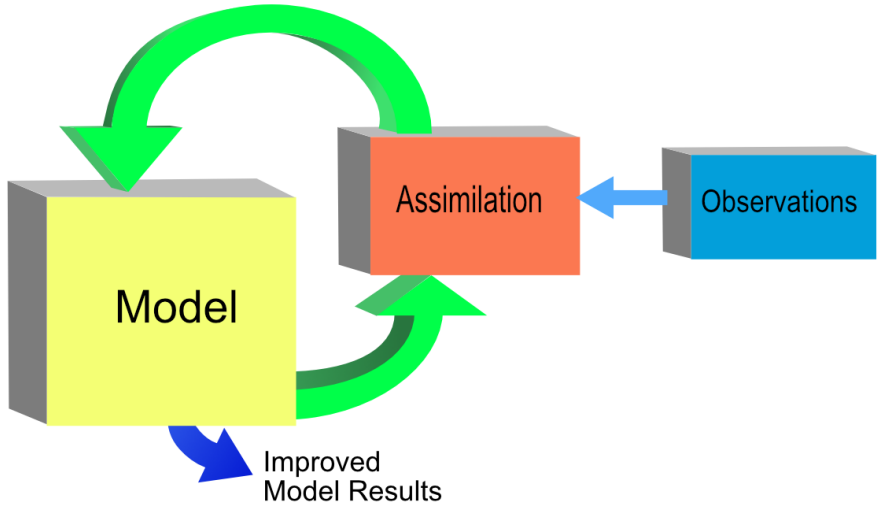
DFG funded Collaborative Research Center on Data Assimilation (www.sfb1294.de)

- maximum funding period: 12 years
- 12 scientific projects
- 24 doctoral and postdoctoral positions
- schools, fellowships, etc.

SFB 1294 Data Assimilation

Numerical Weather Prediction





- **Model**: highly nonlinear discretized partial differential equations
- **Data**: heterogeneous mix of ground-, airborne-, satellite-based and radar data
- 24/7 **data assimilation** service for optimal weather prediction

Bayesian inference

The three key ingredients of Bayesian inference:

- a prior measure over the variable of interest
- the likelihood of an observation given the variable of interest
- the posterior measure over the variable of interest conditioned on the given observation

Note: All variables are treated as random variables contrary to frequentist approach to inference.

Bayes' formula I

Random variable of interest $\mathbf{Z}$ with prior distribution/measure/density

$$\mathbf{Z} \sim \mathbb{P} \qquad \text{or} \qquad \mathbf{Z} \sim \pi$$

The expectation (expected value) of a function $g(\mathbf{z})$ under $\mathbb{P}$ is defined as

$$\mathbb{E}[g] = \int g(\mathbf{z})\, \mathbb{P}(\mathrm{d}\mathbf{z})$$

or

$$\mathbb{E}[g] = \int g(\mathbf{z})\, \pi(\mathbf{z})\, \mathrm{d}\mathbf{z}\,.$$

We also use the shorthand

$$\bar{g}, \qquad \pi[g], \qquad \mathbb{P}[g].$$

Bayes' formula II

The likelihood characterizes the probability of observing y given $\mathbf{z}$:

$$l(y|\mathbf{z})$$

Note: We assume for simplicity that l is normalized, i.e.

$$\int l(y|\mathbf{z})\,\mathrm{d}y = 1.$$

Evidence of y under the prior $\mathbb{P}$:

$$\begin{aligned}\bar{l}(y) &= \int l(y|\mathbf{z})\,\mathbb{P}(\mathrm{d}\mathbf{z}) \\ &= \mathbb{P}[l(y,\cdot)] = \pi[l(y,\cdot)]\,.\end{aligned}$$

Bayes' formula III

Rules of conditional probabilities

$$\pi(\mathbf{z}, y) = l(y|\mathbf{z})\,\pi(\mathbf{z}) = \pi(\mathbf{z}|y)\,\bar{l}(y)$$

yield the posterior density

$$\pi(\mathbf{z}|y) = \frac{\pi(\mathbf{z}, y)}{\bar{l}(y)} = \frac{l(y|\mathbf{z})\,\pi(\mathbf{z})}{\bar{l}(y)}$$
$$\propto l(y|\mathbf{z})\,\pi(\mathbf{z})\,.$$

Notation. We use the shorthand $\nu^*(\mathbf{z})$ for $\pi(\mathbf{z}|y)$ or, more generally, in case of a posterior measure, $\mathbb{Q}^*(\mathrm{d}\mathbf{z})$.

Remark. Normalizing constant, i.e. evidence $\bar{l}(y)$, is important when comparing models, e.g. different prior distributions $\mathbb{P}_\theta$.

Bayes' formula IV

Bayes' formula needs to be generalized when the prior is a measure.

Radon-Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{P}} = \frac{l(y|\cdot)}{\bar{l}(y)} \propto l(y|\cdot)$$

of posterior wrt prior measure.

In words: $\mathbb{Q}^*$ is absolute continuous with respect to $\mathbb{P}$ with density $l(y, \mathbf{z})/\bar{l}(y)$.

In equation form: $\mathbb{Q}^* \ll \mathbb{P}$ and

$$\begin{aligned}\mathbb{Q}^*[g] &= \int g(\mathbf{z})\, \mathbb{Q}^*(\mathrm{d}\mathbf{z}) \\ &= \int g(\mathbf{z})\, \frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{P}}\, \mathbb{P}(\mathrm{d}\mathbf{z}) = \int g(\mathbf{z})\, \frac{l(y|\mathbf{z})}{\bar{l}(y)}\, \mathbb{P}(\mathrm{d}\mathbf{z}) \\ &= \frac{1}{\bar{l}(y)} \mathbb{P}[g\, l(y|\cdot)].\end{aligned}$$

Kullbeck-Leibler divergence:

$$D(\mathbb{Q}|\mathbb{Q}^*) = \int \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Q}^*} \mathbb{Q}(\mathrm{d}\mathbf{z}) = \mathbb{Q}\left[\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Q}^*}\right].$$

It holds that

$$D(\mathbb{Q}|\mathbb{Q}^*) > 0 \quad \text{for all} \quad \mathbb{Q} \neq \mathbb{Q}^*.$$

Donsker-Varadhan principle:

$$-\log \bar{l}(y) = \inf_{\mathbb{Q} \ll \mathbb{P}} \left\{-\mathbb{Q}[\log l(y|\cdot)] + D(\mathbb{Q}|\mathbb{P})\right\}$$

with the infimum taken over all measures $\mathbb{Q}$ which are absolutely continuous wrt $\mathbb{P}$. The infimum is achieved for $\mathbb{Q} = \mathbb{Q}^*$.

$\mathcal{F} = -\log \bar{l}(y)$ is called the free energy.

Remark. $-\mathbb{Q}[\log l(y|\cdot)]$ is called the expected loss under $\mathbb{Q}$.

Machine learning vs data assimilation

Key element of both machine learning (ML) and DA:

$$\text{joint probability}: \qquad \pi(\mathbf{z}, y) = l(y|\mathbf{z})\,\pi(\mathbf{z})$$

ML: the (effective) dimension of the data y is much larger than the (effective) dimension of the parameters $\mathbf{z}$ (big data)

DA: the (effective) dimension of $\mathbf{z}$ is much larger than the (effective) dimension of the data y (complex models)

In addition:

- ML addresses mostly static inference problems
- DA has an element of forgetting (not just learning)
- Both ML and DA lead to complex minimization and quantification of uncertainty (UQ) problems

Computational approaches

Overview:

- distributional approximations (deterministic)
  - point estimators such as MAP estimator:

    $$\mathbf{z}^* := \operatorname{argmin} V(\mathbf{z}), \qquad V(\mathbf{z}) := -\log \nu^*(\mathbf{z}))$$

    leading to 3DVar, 4DVar from meteorology
  - variational Bayes (VB)
- Monte Carlo approximations (random)
  - Markov chain Monte Carlo (MCMC)
  - importance sampling (IS)

Approximate posterior ν^* by a Gaussian distribution

$$\nu(\mathbf{z}) = (2\pi)^{-N_z/2}|P|^{-N_z/2}e^{-\frac{1}{2}(\mathbf{z}-\mu)^{\mathsf{T}}P^{-1}(\mathbf{z}-\mu)}$$

with mean μ and covariance P chosen such that the variational free energy

$$\mathcal{F}(\nu) = -\nu[\log l(y|\cdot)] + D(\nu|\pi)$$

is minimised.

Critical points (μ^*, P^*) satisfy:

$$0 = \nu\left[\nabla_{\mathbf{z}} \log \nu^*\right], \qquad (P^*)^{-1} = -\nu\left[\nabla_{\mathbf{z}}\nabla_{\mathbf{z}} \log \nu^*\right]$$

Remark. Compare to Laplace approximation:

$$0 = \nabla_{\mathbf{z}} \log \nu^*, \qquad (P^*)^{-1} = -\nabla_{\mathbf{z}}\nabla_{\mathbf{z}} \log \nu^*_{|\mathbf{z}=\mu^*}.$$

Monte Carlo methods: Random algorithms for producing (weighted) samples $\mathbf{z}^i = \mathbf{Z}^i(\omega)$, $i = 1, \ldots, M$, from $\mathbb{Q}^*$.

The target measure $\mathbb{Q}^*$ is approximated by the associated random measure:

$$\mathbb{Q}^* \approx \frac{1}{M} \sum_{i=1}^{M} w_i \, \delta(\mathbf{z} - \mathbf{z}^i),$$

$\delta(\cdot)$ the standard Dirac delta measure.

Two examples:

- Markov chain Monte Carlo (MCMC): $w_i = 1$
- importance sampling (IS): nonuniform weights w_i

General idea of MCMC:

Find a transition kernel $q(\mathrm{d}\mathbf{z}'|\mathbf{z})$ such that

$$\text{Invariance:} \qquad \mathbb{Q}^*(\mathrm{d}\mathbf{z}') = \int q(\mathrm{d}\mathbf{z}'|\mathbf{z})\,\mathbb{Q}^*(\mathrm{d}\mathbf{z})$$

holds.

Produce correlated samples $\mathbf{z}^i$, $i = 1, \ldots, M$, sequentially

$$\mathbf{z}^i = \mathbf{Z}^i(\omega) \sim q(\cdot|\mathbf{z}^{i-1}), \qquad i = 1, \ldots, M\,.$$

Efficiency: equivalent number, M_{eff}, of independent samples required to produce the same accuracy; typically

$$M_{\text{eff}} \ll M\,.$$

Example. Consider gradient flow SDE (Brownian dynamics)

$$\mathrm{d}\mathbf{Z}_t = \nabla_{\mathbf{z}} \log \nu^*(\mathbf{Z}_t)\,\mathrm{d}t + \sqrt{2}\mathrm{d}W_t\,,$$

W_t N_z-dimensional standard Brownian motion.

This SDE has ν^* as a stationary distribution.

Discretize in time by Euler-Maryama method

$$\mathbf{z}^i = \mathbf{z}^{i-1} + \nabla_{\mathbf{z}} \log \nu^*(\mathbf{z}^{i-1})\,\Delta t + \sqrt{2\Delta t}\Xi^i$$

with $\Xi^i \sim N(0, I)$ and step-size $\Delta t > 0$.

If exact sampling is desired, apply a Metropolis-Hastings accept-reject criterion to correct for numerical errors.

Importance Sampling I

Find proposal density $\mathbb{Q}$ such that

- $\mathbb{Q}^* \ll \mathbb{Q}$, i.e.

$$\bar{g} := \int g(\mathbf{z})\, \mathbb{Q}^*(\mathrm{d}\mathbf{z}) = \int g(\mathbf{z})\, \frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{Q}}\, \mathbb{Q}(\mathrm{d}\mathbf{z})$$

- $\mathbb{Q}$ can be easily sampled from.

Example. $\mathbb{Q} = \mathbb{P}$.

Approximate $\mathbb{Q}^*$ by

$$\mathbb{Q}^* \approx \frac{1}{M} \sum_{i=1}^{M} w_i\, \delta(\mathbf{z} - \mathbf{z}^i)$$

with

$$\mathbf{z}^i = \mathbf{Z}^i(\omega) \sim \mathbb{Q}, \qquad w_i = \frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{Q}}(\mathbf{z}^i)\,.$$

Example. $\mathbb{Q} = \mathbb{P}$, $w_i \propto l(y|\mathbf{z}^i)$, $\sum_i w_i = M$.

Importance Sampling II

Notation: Radon-Nikodym derivative (e.g. likelihood):

$$L(\mathbf{z}) := \frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{Q}}(\mathbf{z})$$

Effective sample size:

$$M_{\text{eff}} := \frac{1}{\frac{1}{M}\sum_i w_i^2} M \leq M, \qquad w_i := L(\mathbf{z}^i)\,.$$

Law of large numbers:

$$M_{\text{eff}} \approx \rho M$$

with

$$\rho := \frac{1}{\mathbb{Q}[L^2]} = \frac{\mathbb{Q}[L]^2}{\mathbb{Q}[L^2]} \leq 1\,.$$

Upper bound:

$$\rho \leq e^{-2D(\mathbb{Q}^*|\mathbb{Q})} \leq 1\,.$$

References

Christian Robert, The Bayesian Choice, Springer, 2007

Christian Robert, George Casella, Monte Carlo Statistical Methods, Springer, 2010

Sebastian Reich and Colin Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015

Andrew Stuart, Inverse problems: A Bayesian perspective, Acta Numerica, 2010, 451–559

Manfred Opper and Cedric Archambeau, The Variational Gaussian Approximation Revisited, Journal Neural Computation, 21, 2009, 786–792

Signal process I

In case of state space models, the prior measure $\mathbb{P}$ is defined recursively.

Discrete-time models:

$$X_n = \mathcal{M}(X_{n-1}) + Q^{1/2}\Xi_n\,,$$

$\Xi_n \sim N(0, I)$, $X_0 \sim \pi_0$, $n = 1, \ldots, N$, $x_n = X_n(\omega) \in \mathbb{R}^{N_x}$.

Continuous-time models (SDEs):

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t\,,$$

W_t N_x-dimensional standard Brownian motion, $X_0 \sim \pi_0$, $t \in [0, T]$.

Variable of interest (discrete time):

$$\mathbf{z} = x_{0:N} = (x_0, x_1, \ldots, x_N)$$

Prior density/measure:

$$\pi(\mathbf{z}) = \pi(x_{0:N}) = \pi_0(x_0)\,\pi(x_1|x_0)\cdots\pi(x_N|x_{N-1})$$

Transition kernel:

$$X_n \sim \pi(\cdot\,|x_{n-1})$$

with

$$\pi(x|x') \propto \exp\left(-\frac{1}{2}(x - \mathcal{M}(x'))^{\mathsf{T}} Q^{-1} (x - \mathcal{M}(x'))\right).$$

Remark. It is easy to generate samples from prior density π.

Formal derivation of prior measure for time-continuous models.

Euler-Maruyama method for SDEs:

$$X_n = X_{n-1} + f(X_{n-1})\,\Delta t + Q^{1/2}\Xi_n\,,$$

with $T = \Delta t\, N$, $\Xi_n \sim N(0, \Delta t I)$.

Transition kernel of Euler-Maruyama method for finite Δt:

$$\pi_{\Delta t}(x|x') \propto \exp\left(-\frac{1}{2\Delta t}(x - x' - f(x')\Delta t)^{\mathsf{T}} Q^{-1}(x - x' - f(x')\Delta t)\right).$$

and $\mathbf{z} = x_{0:N}$, $T = \Delta t\, N$.

Limit $\Delta t \to 0$ for fixed $T = \Delta t\, N$.

Realizations of $\mathbf{Z}$ are a.s. continuous functions

$$\mathbf{z} = x_{[0,T]} \in C([0,T], \mathbb{R}^{N_x})$$

with measure $\mathbb{P}$ over $C([0,T], \mathbb{R}^{N_x})$ formally defined by

$$\lim_{N\to\infty} \left\{ \pi(x_0) \prod_{n=1}^{N} \pi_{\Delta t}(x_n | x_{n-1}) \right\} \to \mathbb{P}(\mathrm{d}x_{[0,T]})$$

Note. Two SDEs with different diffusion matrices Q lead to measures $\mathbb{P}$ which are mutually singular.

Discrete-in-time.

Forward model:

$$Y_n = h(X_n) + R^{1/2}\Sigma_n$$

$\Sigma_n \sim N(0, I)$, $n = 1, \ldots, N$.

Likelihood:

$$l(y_{1:N}|x_{0:N}) \propto \exp\left(-\frac{1}{2}\sum_{n=1}^{N}(y_n - h(x_n))^{\mathsf{T}} R^{-1}(y_n - h(x_n))\right).$$

Continuous-in-time.

Forward model:

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + R^{1/2} V_t\,,$$

V_t standard Brownian motion, $Y_0 =$, $t \in [0, T]$.

Likelihood:

$$l(y_{[0,T]}|x_{[0,T]}) \propto \exp\left(-\frac{1}{2}\int_0^T \left(h_t^{\mathsf{T}} R^{-1} h_t \mathrm{d}t - 2h_t^{\mathsf{T}} R^{-1} \mathrm{d}y_t\right)\right)$$

with $h_t = h(x_t)$.

Bayes' formula:

Discrete-in-time model and observations:

$$\nu^*(x_{0:N}) \propto l(y_{1:N}|x_{0:N})\,\pi(x_{0:N})$$

Continuous-in-time model and observations:

$$\frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{P}} \propto l(y_{[0,T]}|x_{[0,T]})\,.$$

Remark. Normalising constants (i.e. evidence) $\bar{l}(y_{1:N})$ and $\bar{l}(y_{[0,T]})$, respectively, are important for model comparison.

Posterior measure II

We are typically interested in marginal distributions only.

(a) Smoothing/reanalysis: distribution of x_t / x_n given all the data:

$$X_{t|T} \sim \nu_{t|T} \qquad X_{n|T} \sim \nu_{n|T} \,.$$

(b) Filtering: distribution of x_t / x_n given all the data up to t / n:

$$X_{t|t} \sim \nu_{t|t} \qquad X_{n|n} \sim \nu_{n|n} \,.$$

(c) Prediction: distribution of x_t / x_n given data up to $\tau < t / k < n$:

$$X_{t|\tau} \sim \nu_{t|\tau} \qquad X_{n|k} \sim \nu_{n|k} \,.$$

Posterior measure III

Consider the modified SDE

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}\tilde{W}_t, \qquad X_0(\omega) = x_0\,, \tag{1}$$

with

$$\tilde{W}_t = W_t + \int_0^t u_s \mathrm{d}s\,.$$

Theorem (Girsanov).

Measure $\mathbb{P}$ introduced by (1) with $u_s \equiv 0$.

Measure $\mathbb{Q}^u$ introduced for any $u_s \neq 0$ such that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t |u_s|^2 \mathrm{d}s\right)\right] < \infty\,.$$

$\mathbb{Q}^u$ is absolutely continuous wrt $\mathbb{P}$ with Radon-Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}^u}{\mathrm{d}\mathbb{P}}\Big|_{W_{[0,t]}} = \exp Z_t^u\,, \qquad Z_t^u = \int_0^t u_s^{\mathsf{T}}\mathrm{d}W_s + \frac{1}{2}\int_0^t |u_s|^2\mathrm{d}s\,.$$

Posterior measure IV

Application of Girsanov to data assimilation with underlying SDE models:

Find a control law u and a change of the initial measure π_0 such that

$$\mathbb{Q}^u \approx \mathbb{Q}^* \,.$$

See below and Part IV on proposals steps.

Remarks.

(i) A good choice of the model error diffusion matrix Q in

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t$$

is crucial (smoothing vs. prediction).

(ii) The filtering problem also leads to control-type formulations; but they are motivated differently. See feedback particle filter later in this part.

Kalman-Bucy filter I

Linear SDE:

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t\,, \qquad X_0 \sim \mathrm{N}(\bar{x}_0, P_0)\,.$$

Linear forward model:

$$\mathrm{d}Y_t = HX_t\,\mathrm{d}t + \mathrm{d}V_t\,, \qquad Y_0 = 0\,.$$

Prior and posterior distributions are Gaussian:

$$\begin{aligned} \text{Signal}: &\quad (\bar{x}_t, P_t), \\ \text{Filtering}: &\quad (\bar{x}_{t|t}, P_{t|t}), \\ \text{Smoothing}: &\quad (\bar{x}_{t|T}, P_{t|T}) \end{aligned}$$

Evolution equations for

Signal:

$$\frac{\mathrm{d}\bar{x}_t}{\mathrm{d}t} = A\bar{x}_t\,,$$
$$\frac{\mathrm{d}P_t}{\mathrm{d}t} = AP_t + P_tA^{\mathsf{T}} + Q\,,$$

$\bar{x}_0 = \bar{x}_{0|0}$, $P_0 = P_{0|0}$ given.

Filter:

$$\mathrm{d}\bar{x}_{t|t} = A\bar{x}_{t|t}\,\mathrm{d}t - K_t(H\bar{x}_{t|t}\,\mathrm{d}t - \mathrm{d}Y_t)\,,$$
$$\frac{\mathrm{d}P_{t|t}}{\mathrm{d}t} = AP_{t|t} + P_{t|t}A^{\mathsf{T}} + Q - K_tHP_{t|t}$$

with Kalman gain matrix

$$K_t = P_{t|t}H^{\mathsf{T}}\,.$$

Smoother:

Given the filter solution $(\bar{x}_{t|t}, P_{t|t})$, $t \in [0, T]$, solve **backward in time**

$$\frac{\mathrm{d}\bar{x}_{t|T}}{\mathrm{d}t} = A\bar{x}_{t|T} + QP_{t|t}^{-1}(\bar{x}_{t|T} - \bar{x}_{t|t}),$$
$$\frac{\mathrm{d}P_{t|T}}{\mathrm{d}t} = AP_{t|T} + P_{t|T}A^{\mathsf{T}} + Q + QP_{t|t}^{-1}P_{t|T} + P_{t|T}P_{t|t}^{-1}Q$$

for given $\bar{x}_{T,T}$ and $P_{T|T}$ at $t = T$.

Kalman-McKean-Vlasov equations

Interacting particle McKean-Vlasov representation of Kalman filter/smoother equations:

Signal:

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t, \qquad X_0 \sim \pi_0\,.$$

Filter:

$$\mathrm{d}X_{t|t} = AX_{t|t}\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t - K_t\left(\frac{1}{2}\left\{HX_{t|t} + H\bar{x}_{t|t}\right\}\mathrm{d}t - \mathrm{d}Y_t\right)$$

with $\bar{x}_{t|t} = \mathbb{E}[X_{t|t}]$, $K_t = P_{t|t}H^{\mathsf{T}}$, and $P_{t|t} = \mathbb{E}[(X_{t|t} - \bar{x}_{t|t})(X_{t|t} - \bar{x}_{t|t})^{\mathsf{T}}]$.

Smoother:

$$\mathrm{d}X_{t|T} = AX_{t|T}\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t + QP_{t|t}^{-1}(X_{t|T} - \bar{x}_{t|t})$$

with $X_{T|T} \sim \pi_{T|T}$.

Signal:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t, \qquad X_0 \sim \pi_0\,.$$

Feedback particle filter:

$$\mathrm{d}X_{t|t} = f(X_{t|t})\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t - K_t \circ \left(\frac{1}{2}\left\{h(X_{t|t}) + \bar{h}_{t|t}\right\}\mathrm{d}t - \mathrm{d}Y_t\right)$$

with the Kalman gain now implicitly defined by

$$\nabla_x \cdot (\pi_{t|t} K_t) = \pi_{t|t}(h - \bar{h}_{t|t})\,.$$

Smoother extension of feedback particle filter:

$$\mathrm{d}X_{t|T} = f(X_{t|T})\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t - Q\nabla_x \log \pi_{t|t}(X_{t|T})\,\mathrm{d}t$$

with $X_{T|T}$ given.

Forward optimal control formulation:

$$\mathrm{d}\tilde{X}_t = f(\tilde{X}_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t + Q\left\{\nabla_x \log \frac{\pi_{t|T}}{\pi_{t|t}}(\tilde{X}_t)\right\}\mathrm{d}t$$

with $\pi_{0|T}$ given by the backward smoother formulation.

The (time-dependent) control law is given by

$$u_t(x) = Q^{1/2}\,\nabla_x \log \frac{\pi_{t|T}}{\pi_{t|t}}(x)\,.$$

Remark. Time-averaged controls

$$\bar{u}(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T u_t(x)\,\mathrm{d}t$$

provide systematic model correction terms.

References

A. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970

Kody Law, Andrew Stuart and Konstantinos Zygalakis, Data Assimilation: A Mathematical Introduction, Springer, 2015

Greg Pavliotis, Stochastic Processes and Applications, Springer, 2014

Sebastian Reich and Colin Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015

Amir Taghvaei, Jana de Wiljes, Prashant Mehta and Sebastian Reich, Kalman Filter and Its Modern Extensions for the Continuous-Time Nonlinear Filtering Problem, J. Dyn. Sys. Meas., Control, 140, 2017, 030904

Carsten Hartmann, L. Richter, Christof Schütte and W. Zhang, Variational characterization of free energy: theory and algorithms, Entropy, 19, 2017, 626–653.

Kai Bergemann and Sebastian Reich, An ensemble Kalman-Bucy filter for continuous data assimilation, Meteorologische Zeitschrift, 21, 2012, 213–219.

McKean-Vlasov representation of Kalman-Bucy filter for continuous-time signal and forward models:

$$\mathrm{d}X_{t|t} = AX_{t|t}\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t - K_t\left(\frac{1}{2}\left\{HX_{t|t} + H\bar{x}_{t|t}\right\}\mathrm{d}t - \mathrm{d}Y_t\right)$$

with $\bar{x}_{t|t} = \mathbb{E}[X_{t|t}]$, $K_t = P_{t|t}H^{\mathsf{T}}$, and $P_{t|t} = \mathbb{E}[(X_{t|t} - \bar{x}_{t|t})(X_{t|t} - \bar{x}_{t|t})^{\mathsf{T}}]$.

Monte Carlo approximation: for $i = 1, \ldots, M$

$$\mathrm{d}x^i_{t|t} = Ax^i_{t|t}\,\mathrm{d}t + Q^{1/2}\mathrm{d}W^i_t - K^M_t\left(\frac{1}{2}\left\{Hx^i_{t|t} + H\bar{x}^M_{t|t}\right\}\mathrm{d}t - \mathrm{d}Y_t\right)$$

with

$$\bar{x}^M_{t|t} = \frac{1}{M}x^i_{t|t}, \qquad P^M_{t|t} = \frac{1}{M-1}\sum_{i=1}^{M}(x^i_{t,t} - \bar{x}^M_{t|t})(x^i_{t,t} - \bar{x}^M_{t|t})^{\mathsf{T}}$$

and $K^M_t = P^M_{t|t}H^{\mathsf{T}}$.

Extension to continuous-time nonlinear signal and forward models: for $i = 1, \ldots, M$

$$\mathrm{d}x^i_{t|t} = f(x^i_{t|t})\,\mathrm{d}t + Q^{1/2}\mathrm{d}W^i_t - K^M_t\left(\frac{1}{2}\left\{h(x_{t|t})^i + \bar{h}^M_{t|t}\right\}\mathrm{d}t - \mathrm{d}Y_t\right)$$

with

$$\bar{h}^M_{t|t} = \frac{1}{M}\sum_{i=1}^{M} h(x^i_{t|t})$$

and

$$K^M_t = \frac{1}{M-1}\sum_{i=1}^{M}(x^i_{t|t} - \bar{x}^M_{t|t})(h(x^i_{t|t}) - \bar{h}^M_{t|t})^{\mathsf{T}}.$$

Alternative formulation

$$\mathrm{d}x^i_{t|t} = f(x^i_{t|t})\,\mathrm{d}t + Q^{1/2}\mathrm{d}W^i_t - K^M_t\left(\left\{h(x_{t|t})^i\mathrm{d}t + \mathrm{d}V^i_t\right\} - \mathrm{d}Y_t\right)$$

McKean-Vlasov representation of Kalman smoother:

$$\mathrm{d}X_{t|T} = AX_{t|T}\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t + QP_{t|t}^{-1}(X_{t|T} - \bar{x}_{t|t})$$

with $X_{T|T} \sim \pi_{T|T}$.

Monte Carlo approximation: for $i = 1, \ldots, M$

$$\mathrm{d}x_{t|T}^i = Ax_{t|T}^i\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t^i + Q(P_{t|t}^M)^{-1}(x_{t|T}^i - \bar{x}_{t|t})$$

with $X_{T|T}^i$ given by Monte Carlo approximation to Kalman-Bucy filter.

Ensemble Kalman-Bucy smoother:

$$\mathrm{d}x_{t|T}^i = f(x_{t|T}^i)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t^i + Q(P_{t|t}^M)^{-1}(x_{t|T}^i - \bar{x}_{t|t})$$

with $X_{T|T}^i$ at $t = T$ given by ensemble Kalman-Bucy filter.

Ensemble Kalman filter I

Discrete-time observations

$$y_n = Hx_n + R^{1/2}\Sigma_n$$

$\Sigma_n \sim N(0, I)$, $n = 1, \ldots, N$.

Required is an update from the forecast

$$X_\mathrm{f} := X_{t_n|t_{n-1}}$$

to the analysis

$$X_\mathrm{a} := X_{t_n|t_n}$$

at time t_n

Notation:

$$\begin{aligned} \text{mean}: &\quad \bar{x}_\mathsf{f} := \mathbb{E}[X_\mathsf{f}], \quad \bar{x}_\mathsf{a} := \mathbb{E}[X_\mathsf{a}] \\ \text{deviation}: &\quad \Delta X_\mathsf{f} := X_\mathsf{f} - \bar{x}_\mathsf{f}, \quad \Delta X_\mathsf{a} := X_\mathsf{a} - \bar{x}_\mathsf{a} \\ \text{covariance matrix}: &\quad P_\mathsf{f} := \mathbb{E}[\Delta X_\mathsf{f} \Delta X_\mathsf{f}^\mathsf{T}], \quad P_\mathsf{a} := \mathbb{E}[\Delta X_\mathsf{a} \Delta X_\mathsf{a}^\mathsf{T}] \end{aligned}$$

Ensemble Kalman filter (EnKF) produces X^a such that

$$\begin{aligned} \text{mean update}: &\quad \bar{x}_\mathsf{a} = \bar{x}_\mathsf{f} - K(H\bar{x}_\mathsf{f} - y_n) \\ \text{covariance update}: &\quad P_\mathsf{a} = P_\mathsf{f} - KHP_\mathsf{f} \end{aligned}$$

with Kalman gain matrix

$$K = P_\mathsf{f} H^\mathsf{T} (HP_\mathsf{f} H^\mathsf{T} + R)^{-1} .$$

Remarks. (i) Neither X_f nor X_a need to be Gaussian random variables.
(ii) Stated conditions in red do not determine X_a uniquely.

Implementation:

Forecast ensemble x_{f}^i, $i = 1, \ldots, M$:

$$\text{empirical mean}: \qquad \bar{x}_{\mathrm{f}}^M = \frac{1}{M} \sum_{i=1}^{M} x_{\mathrm{f}}^i$$

$$\text{emprical covariance matrix}: \qquad P_{\mathrm{f}}^M := \frac{1}{M-1} \sum_{i=1}^{M} x_{\mathrm{f}}^i (x_{\mathrm{f}}^i - \bar{x}_{\mathrm{f}}^M)^{\mathsf{T}}$$

$$\text{Kalman gain matrix}: \qquad K^M := P_{\mathrm{f}}^M H^{\mathsf{T}} (H P_{\mathrm{f}}^M H^{\mathsf{T}} + R)^{-1}$$

Stochastic EnKF:

$$x_{\mathrm{a}}^j = \bar{x}_{\mathrm{f}}^j - K^M \left(\left\{ H x_{\mathrm{f}}^j + \eta^j \right\} - y_n \right), \qquad \eta^i \sim \mathrm{N}(0, R)\,.$$

$j = 1, \ldots, M$.

Remark. There are many other variants of the EnKF.

Rewrite of the stochastic EnKF:

$$
\begin{aligned}
x_{\mathrm{a}}^j &= x_{\mathrm{f}}^j - \sum_{i=1}^M x_{\mathrm{f}}^i \frac{1}{M-1} \left\{ (x_{\mathrm{f}}^i - \bar{x}_{\mathrm{f}}^M)^{\mathsf{T}} H^{\mathsf{T}} (H P_{\mathrm{f}}^M H^{\mathsf{T}} + R)(H x_{\mathrm{f}}^j + \eta^j - y_n) \right\} \\
&= x_{\mathrm{f}}^j - \sum_{i=1}^M x_{\mathrm{f}}^i s_{ij} \\
&= \sum_{i=1}^M x_{\mathrm{f}}^i \{\delta_{ij} - s_{ij}\}, \qquad (\delta_{ij} \text{ the Kronecker delta}) \\
&= \sum_{i=1}^M x_{\mathrm{f}}^i d_{ij}
\end{aligned}
$$

Remark. Different EnKF formulations lead to different d_{ij}'s. But they all satisfy

$$
\sum_{i=1}^M d_{ij} = 1 \,.
$$

Definition. The class of (linear) ensemble transform filters is defined by

$$x_{\mathrm{a}}^{j} = \sum_{i=1}^{M} x_{\mathrm{f}}^{i} d_{ij}$$

for appropriate coefficients d_{ij} satisfying

$$\sum_{i=1}^{M} d_{ij} = 1 \,.$$

Remark. Define

$$w_i := \sum_{j=1}^{M} d_{ij}$$

and note that

$$\bar{x}_{\mathrm{a}}^{M} = \frac{1}{M} \sum_{j=1}^{M} x_{\mathrm{a}}^{j} = \frac{1}{M} \sum_{i,j=1}^{M} x_{\mathrm{f}}^{i} d_{ij} = \frac{1}{M} \sum_{i=1}^{M} w_i \, x_{\mathrm{f}}^{i} \,.$$

(i) For a ensemble transform filter to be consistent, it should hold that

$$w_i \propto \exp\left\{-\frac{1}{2}(h(x_{\mathrm{f}}^j)-y_n)^{\mathsf{T}}R^{-1}(h(x_{\mathrm{f}}^j)-y_n)\right\} \qquad (\textit{importance weights})$$

subject to $\sum_{i=1}^{M} w_i = M$.

(ii) Absolute continuity of the posterior measure with respect to the prior measure suggests that any x_{a}^j should be in the convex hall formed by the prior ensemble $\{x_{\mathrm{f}}^j\}$.

This holds provided $d_{ij} \geq 0$ and $\sum_i d_{ij} = 1$ for all $i, j = 1, \ldots, M$.

Summary. A consistent ensemble transform filter should satisfy

$$d_{ij} \geq 0, \quad \sum_{i=1}^{M} d_{ij} = 1, \quad \sum_{j=1}^{M} d_{ij} = w_i \text{ (importance weights)}$$

The conditions

$$d_{ij} \geq 0, \quad \sum_{i=1}^{M} d_{ij} = 1, \quad \sum_{j=1}^{M} d_{ij} = w_i \text{ (importance weights)}$$

do not uniquely determine the coefficients d_{ij}.

The ensemble transform particle filter (ETPF) is based on

$$\{d_{ij}\} = \operatorname{argmax} \sum_{i=1}^{M} (x_{\mathrm{a}}^i - \bar{x}_{\mathrm{a}}^M)^{\mathsf{T}} (x_{\mathrm{f}}^i - \bar{x}_{\mathrm{f}}^M)$$

subject to the constraints stated above and

$$x_{\mathrm{a}}^j := \sum_{i=1}^{M} x_{\mathrm{f}}^i d_{ij}.$$

Remark. This is equivalent to a discrete optimal transport problem.

Lorenz-63 model, first component observed infrequently ($\Delta t = 0.12$) and with large measurement noise ($R = 8$):

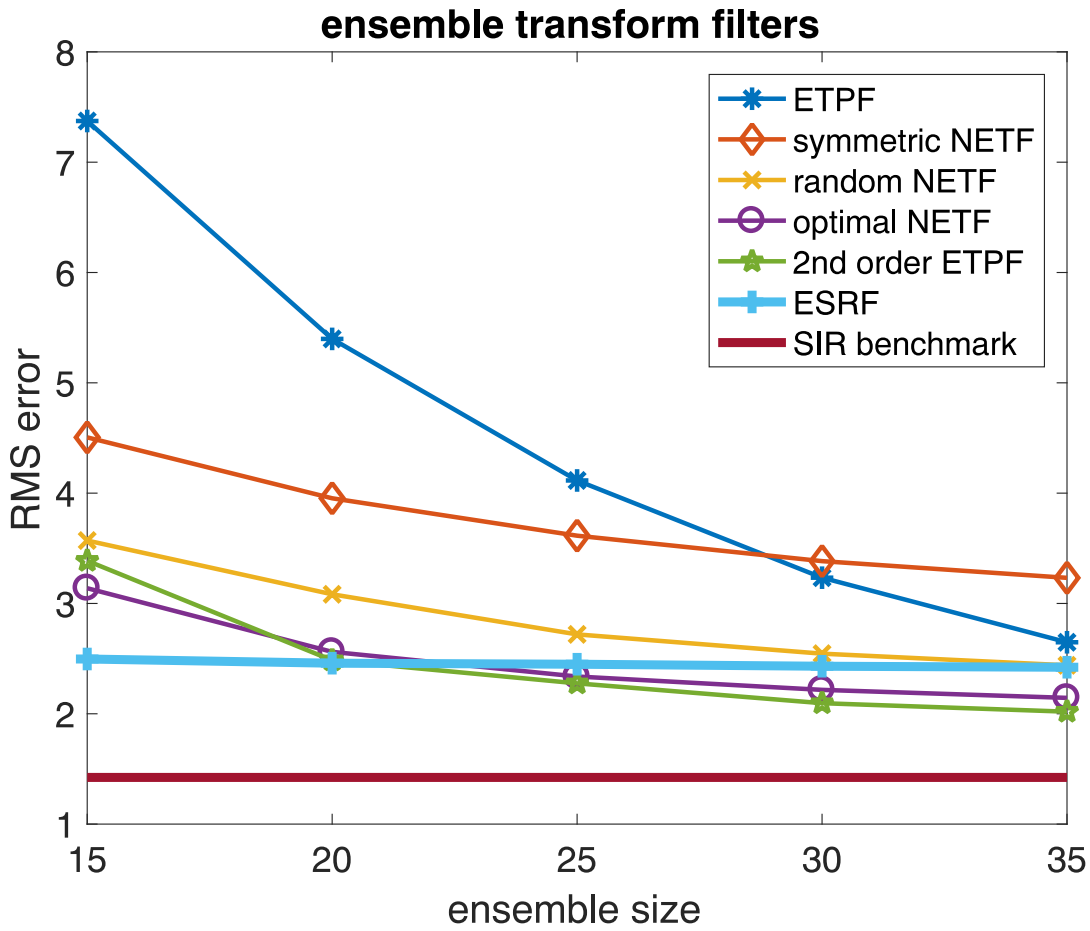


Figure: RMSEs for various second-order accurate LETFs compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, M.

Stability and accuracy

Data (scalar) at time t_n: $y_n \sim \mathrm{N}(y_{\mathrm{true}}, R)$

Analysis at time t_n: $\{y^i_{n|n}\}$

RMS error:

$$\mathrm{RMSE} := \left\{ \frac{1}{N} \sum_{n=1}^{N} (y_n - \bar{y}_{n|n})^2 \right\}^{1/2} < R^{1/2}$$

Ensemble spread:

$$\mathrm{VAR} := \frac{1}{N} \sum_{n=1}^{N} (y_{n|n} - \bar{y}_{n|n})^2 < R$$

Calibration and sharpness:

$$\mathrm{CRPS} := \frac{1}{N} \sum_{n=1}^{N} \int (F_{y_n}(y) - F_{y^i_{n|n}}(y))^2 \mathrm{d}y$$

References

G. Evensen, Data Assimilation. The Ensemble Kalman Filter, Springer, 2006

Kody Law, Andrew Stuart and Konstantinos Zygalakis, Data Assimilation: A Mathematical Introduction, Springer, 2015

Sebastian Reich and Colin Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015

Mark Asch, Marc Bocquet and Maelle Nodet, Data Assimilation. Methods, Algorithms, and Applications, SIAM, 2017.

Jana de Wiljes, Sebastian Reich, Wilhelm Stannat, Long-time stability and accuracy of the ensemble Kalman-Bucy filter for fully observed processes and small measurement noise, arXiv:1612.06065, 2017

Sebastian Reich, A nonparametric ensemble transform method for Bayesian inference, SIAM J. Sci. Comput., 35, 2013, A2013–A2024.

Tilmann Gneiting, Fadoua Balabdaoui, Adrian Raftery, Probabilistic forecasts, calibration and sharpness, J. Royal Stats. Soc., Series B, 69, 2007, 243–268.

Importance sampling leads to weighted particle approximation of the posterior measure:

$$\mathbb{Q}^* \approx \frac{1}{M} \sum_{i=1}^{M} w_i \, \delta(\mathbf{z} - \mathbf{z}^i)$$

with $\mathbf{z}^i = \mathbf{Z}^i(\omega) \sim \mathbb{P}$ and

$$w_i = L(\mathbf{z}^i) := \frac{l(y|\mathbf{z}^i)}{\bar{l}(y)} \, .$$

It holds that

$$\rho := \frac{\mathbb{P}[L]^2}{\mathbb{P}[L^2]} = \frac{1}{\mathbb{P}[L^2]} \leq e^{-2D(\mathbb{Q}^*|\mathbb{P})} \, ,$$

which scales like

$$\rho \approx C \, e^{-N_y}$$

in case of N_y independent observations and the effective sample size (see Part I) decreases exponentially fast as $N_y \gg 1$.

Currently available remedies

Available approaches to beat the curse of dimensionality include:

- variational data assimilation
- localization
- ensemble inflation
- hybrid filter
- alternative proposal steps

Weak constraint 4DVar data assimilation:

$$x_{0:N|N}^{\mathrm{MAP}} = \operatorname{argmin} L(x_{0:N})$$

with

$$L(x_{0:N}) = \frac{1}{2}(x_0 - \bar{x}_0)^{\mathsf{T}} P_0 (x_0 - \bar{x}_0) + \frac{1}{2} \sum_{n=1}^{N} \left\{ a_n^{\mathsf{T}} Q^{-1} a_n + b_n^{\mathsf{T}} R^{-1} b_n \right\}$$

subject to

$$a_n := x_n - \mathcal{M}(x_{n-1}), \qquad b_n := h(x_n) - y_n \,.$$

Remark. Laplace approximation requires Hessian of L at $x_{0:N|N}^{\mathrm{MAP}}$, which can be obtained as a byproduct of quasi-Newton methods.

The Randomized Maximum Likelihood (RML) method is one method that combines ensemble and variational approaches.

Idee. Perturb cost functional $J(x_{0:N})$ in the following manner:

$$\begin{aligned} \text{Initial conditions}: &\quad \bar{x}_0 + \xi_0^i, \quad \xi_0^i \sim \mathrm{N}(0, P_0) \\ \text{Model errors}: &\quad a_n - \xi_n^i, \quad \xi_n^i \sim \mathrm{N}(0, Q) \\ \text{Measurement errors}: &\quad b_n + \eta_n^i, \quad \eta_n^i \sim \mathrm{N}(0, R) \end{aligned}$$

This leads to

$$x_{0:N|N}^i = \operatorname{argmin} L^i(x_{0:N})$$

with a_n and b_n as defined before and

$$L^i(x_{0:N}) = \frac{1}{2}(x_0 - \bar{x}_0 - \xi_0^i)^{\mathsf{T}} P_0 (x_0 - \bar{x}_0 - \xi_0^i) + \frac{1}{2}\sum_{n=1}^{N}(a_n - \xi_n^i)^{\mathsf{T}} Q^{-1}(a_n - \xi_n^i)$$

$$+ \frac{1}{2}\sum_{n=1}^{N}(b_n + \eta_n^i)^{\mathsf{T}} R^{-1}(b_n + \eta_n^i)$$

Remark. Exact sampling for linear $\mathcal{M}$ and h.

Localization I

States x are spatially dependent. To emphasise this aspect we temporarily switch to notion:

$$x_n \in C(\mathbb{R}^3, \mathbb{R}) \;\rightarrow\; u(x, t_n) \in \mathbb{R}, \quad x \in \mathbb{R}^3$$

Observations at location $x_l \in \mathbb{R}^3$:

$$y_{n,l} = u_{\text{truth}}(x_l, t_n) + R_l^{1/2} \xi_{n,l}, \qquad \xi_{n,l} \sim \mathrm{N}(0, I)\,.$$

Standard EnKF/ ensemble transform filters lead to

$$u_{\mathrm{a}}^j(x) = \sum_{i=1}^{M} u_{\mathrm{f}}^i(x)\, d_{ij} \qquad \forall x \in \mathbb{R}^3\,.$$

Two concepts of localization:

- domain or B-localization
- observation or R-localization

R-localization for EnKF/ ensemble transform filter:

$$u_{\mathrm{a}}^j(x) = \sum_{i=1}^{M} u_{\mathrm{f}}^i(x)\, d_{ij}(x) \qquad \forall x \in \mathbb{R}^3 .$$

Spatially-dependent coefficients $d_{ij}(x)$ depend only on observations in the vicinity of x.

This is achieved through

$$\frac{1}{R_l(x)} := \frac{\rho(x - x_l)}{R_l}$$

with $\rho(0) = 1$ and $\rho(x) \to 0$ as $|x| \to 0$.

E.g. importance weights:

$$w_i(x) \propto \exp\left(-\sum_l \frac{1}{2R_l(x)} (y_{n,l} - u^i(x_l, t_n))^2 \right).$$

for updating $u_{\mathrm{f}}^i(x)$, $i = 1, \ldots, M$.

Ensemble inflation

Multiplicative inflation:

$$x_{\mathrm{f}}^i \to x_{\mathrm{f}}^i + \alpha(x_{\mathrm{f}}^i - \bar{x}_{\mathrm{f}}^M), \qquad \alpha > 0\,.$$

Equivalent to forward Euler discretization of

$$\frac{\mathrm{d}}{\mathrm{d}t} x^i = (x^i - \bar{x}^M), \qquad i = 1, \ldots, M,$$

with step-size $\alpha > 0$.

Statistically equivalent to Euler-Maruyama discretization of SDE

$$\mathrm{d}X = P\,\mathrm{d}W\,, \qquad P = \mathbb{E}[(X - \bar{x})(X - \bar{x})^{\mathsf{T}}],\,,$$

with step-size $\alpha > 0$ as ensemble size $M \to \infty$.

Compare to Brownian motion where P is replaced by a constant matrix Q.

Hybrid filter I

Aim: Bridge EnKF and particle filters in an adaptive manner.

Idee: Decompose likelihood function $l(y_n|x_\mathrm{f})$

$$l(y_n|x_\mathrm{f}) = l(y_n|x_\mathrm{f})^\alpha \, l(y_n|x_\mathrm{f})^{1-\alpha} = l_1(y_n|x_\mathrm{f}) \, l_2(y_n|x_\mathrm{f})$$

with $\alpha \in [0,1]$. Bayes' formula becomes

$$\frac{\mathrm{d}\mathbb{Q}_1}{\mathrm{d}\mathbb{P}}(x_\mathrm{f}) \propto l_1(y_n|x_\mathrm{f})$$
$$\frac{\mathrm{d}\mathbb{Q}_2}{\mathrm{d}\mathbb{Q}_1}(x_\mathrm{f}) \propto l_2(y_n|x_\mathrm{f})$$

It holds that $\mathbb{Q}^* = \mathbb{Q}_2$.

Apply ensemble transform particle filter to the first inference problem and EnKF to the second (or vice verse).

Denote the filter coefficients by $d_{ij,1}$ and $d_{ij,2}$, respectively.

Resulting ensemble transform filter is of the form:

$$x_{\mathrm{a}}^j = \sum_{i=1}^{M} \sum_{k=1}^{M} x_{\mathrm{f}}^i \, d_{ik,1} \, d_{kj,2} = \sum_{i=1}^{M} x_{\mathrm{f}}^i \, d_{ij}$$

with

$$d_{ij} = \sum_{k=1}^{M} d_{ik,1} \, d_{kj,2} \, .$$

Question: How to choose bridging parameter $\alpha \in [0, 1]$? Currently used:

$$\text{effective sample size}: \quad M_{\mathrm{eff}} = \frac{M}{\frac{1}{M} \sum_i w_i^2} \geq c M, \quad w_i \propto l_1(y_n | x_{\mathrm{f}}^i) = l(y_n | x_{\mathrm{f}}^i)^\alpha$$

$c < 1$ a given threshold value, $\sum_{i=1}^{M} w_i = M$.

Hybrid filter III

Hybrid filter: $\mathbf{D} := \mathbf{D}_{\mathrm{ESRF}}(\alpha)\, \mathbf{D}_{\mathrm{ETPF}}(1-\alpha)$.

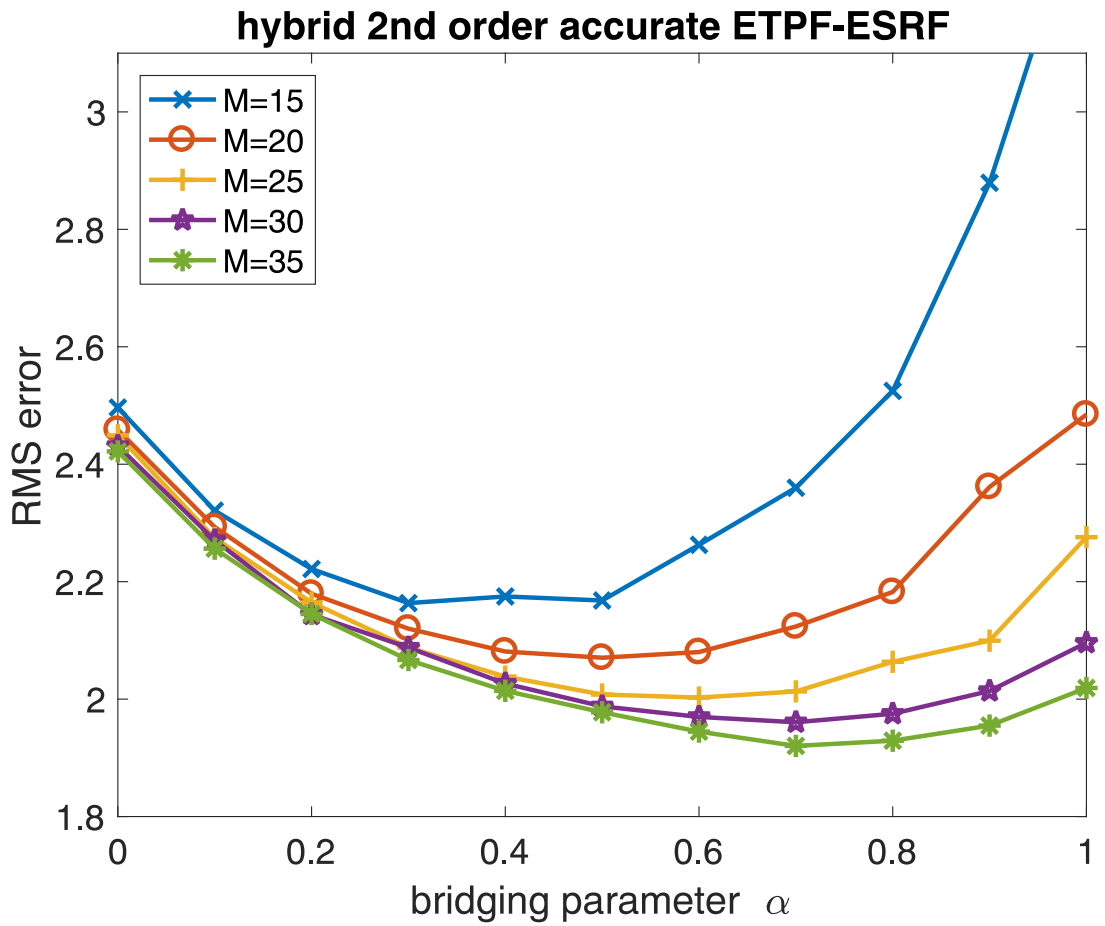


Figure: RMSEs for hybrid ESRF ($\alpha = 0$) and 2nd-order corrected LETF/ETPF ($\alpha = 1$) as a function of the sample size, M.

Hybrid filter IV

Lorenz-96 model, discretized nonlinear advection equation, 40 grid points, every second observed.
Hybrid filter $\mathbf{P} := \mathbf{P}_{\text{LETKF}}(\alpha)\, \mathbf{P}_{\text{ETPF}}(1-\alpha)$ + localization.

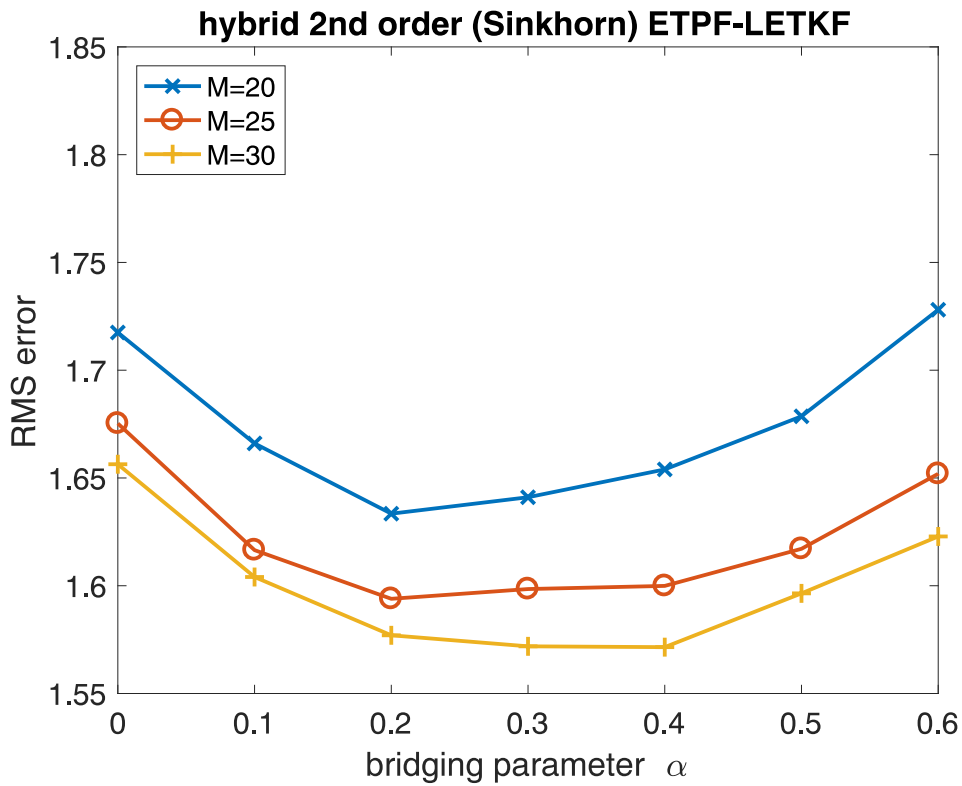


Figure: RMSE for hybrid LETKF ($\alpha = 0$) and 2nd-order corrected LETF/ETPF ($\alpha = 1$).

Alternative Proposal steps I

Standard filter algorithms use model dynamics

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t$$

to produce forecasts x_{f}^i at time t_n given an analysis x_{a}^i at time t_{n-1}.

Alternatively, one can try to find controls u_s^i, $s \in [t_{n-1}, t_n]$ and use

$$\mathrm{d}\tilde{X}_t = f(\tilde{X}_t)\,\mathrm{d}t + Q^{1/2}u_t^i\,\mathrm{d}t + Q^{1/2}\mathrm{d}W_t$$

to produce forecasts $\tilde{x}_{\mathrm{f}}^i$ at time t_n given an analysis x_{a}^i at time t_{n-1}.

Denote the resulting proposal distribution at t_n by

$$q(x_{\mathrm{f}}|u, x_{\mathrm{a}})\,.$$

Alternative Proposal steps II

Target density:

$$\begin{aligned}\pi(y_n, x_\mathrm{f}, x_\mathrm{a}) &\propto l(y_n|x_\mathrm{f})\, q(x_\mathrm{f}|0, x_\mathrm{a})\, \pi_{n-1|n-1}(x_\mathrm{a}) \\ &= \pi(x_\mathrm{f}|x_\mathrm{a}, y_n)\, \pi(y_n|x_\mathrm{a}) \pi_{n-1|n-1}(x_\mathrm{a})\end{aligned}$$

Proposal density:

$$q(x_\mathrm{f}|u, x_\mathrm{a}) \pi_{n-1|n-1}(x_\mathrm{a})$$

Importance sampling:

$$\begin{aligned}w_i &\propto \frac{l(y_n|x_\mathrm{f}^i)\, q(x_\mathrm{f}^i|0, x_\mathrm{a}^i)\, \pi_{n-1|n-1}(x_\mathrm{a}^i)}{q(x_\mathrm{f}^i|u^i, x_\mathrm{a}^i) \pi_{n-1|n-1}(x_\mathrm{a}^i)} \\ &\propto \frac{\pi(x_\mathrm{f}^i|x_\mathrm{a}^i, y_n)\, \pi(y_n|x_\mathrm{a}^i)}{q(x_\mathrm{f}^i|u^i, x_\mathrm{a}^i)}\end{aligned}$$

Alternative Proposal steps III

Recall from Part II:

$$\frac{\mathrm{d}\mathbb{Q}^u}{\mathrm{d}\mathbb{P}}_{|W_{[t_{n-1},t_n]}} = \exp Z_t^u \quad \Rightarrow \quad \frac{q_{\mathrm{f}}(\cdot|u, x_{\mathrm{a}})}{q_{\mathrm{f}}(\cdot|0, x_{\mathrm{a}})}_{|W_{[t_{n-1},t_n]}}$$

with

$$Z_t^u = \int_{t_{n-1}}^{t_n} u_s^{\mathsf{T}} \mathrm{d}W_s + \frac{1}{2}\int_{t_{n-1}}^{t_n} |u_s|^2 \mathrm{d}s\,.$$

Optimal choice:

$$q(x_{\mathrm{f}}^i|u^i, x_{\mathrm{a}}^i) = \pi(x_{\mathrm{f}}^i|x_{\mathrm{a}}^i, y_n)$$

with importance weights

$$w_i \propto \pi(y_n|x_{\mathrm{a}})\,.$$

I.e. optimal control problem for finding u_s^i, $s \in [t_{n-1}|t_n]$ given $X_{t_n} = x_a^i$ and a change of measure at t_{n-1} from $\pi_{n-1|n-1}$ to $\pi_{n-1|n}$.

References

Peter Jan van Leeuwen, Yuan Cheng and Sebastian Reich, Frontiers in Applied Dynamical Systems: Reviews and Tutorials 2: Nonlinear Data Assimilation, Springer, 2015.

Nawinda Chustagulprom, Sebastian Reich, and Maria Reinhardt, A hybrid ensemble transform particle filter for nonlinear spatially extended dynamical systems, SIAM/ASA J UQ, 4, 2016, 592–608.

Paul Fearnhead and Hans R. Künsch, Particle Filters and Data Assimilation, arXiv:1709.04196, 2017.

Walter Acevedo, Jana de Wiljes, and Sebastian Reich, Second-order accurate ensemble transform particle filters, SIAM J. Sci. Comput., 39, 2017, A1834–A1850.

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, A.M. Stuart, Importance sampling: Intrinsic dimension and computational cost, Statistical Science, 32, 2017, 405–431.

R.N. Bannister, A review of operational methods of variational and ensemble-variational data assimilation, Q.J. Royal Meteorol. Soc., 143, 2017, 607–633.