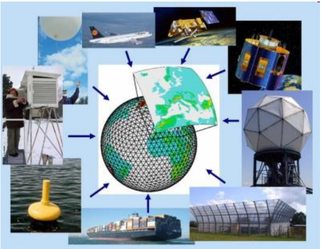
Data assimilation and Uncertainty Quantification: A Lagrangian Interacting Particle Perspective

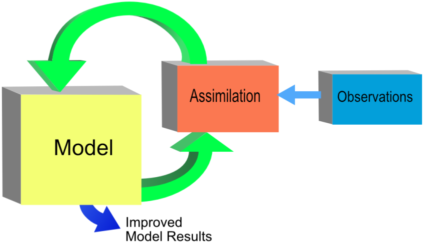
Sebastian Reich (www.sfb1294.de)

Universität Potsdam/ University of Reading

SIAM UQ 2018

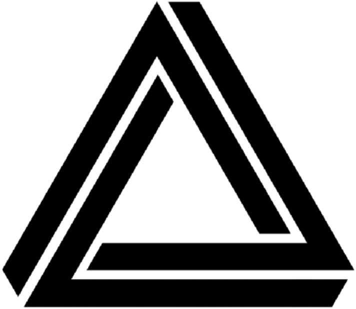
Numerical weather prediction





- **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

Mean Field Equations

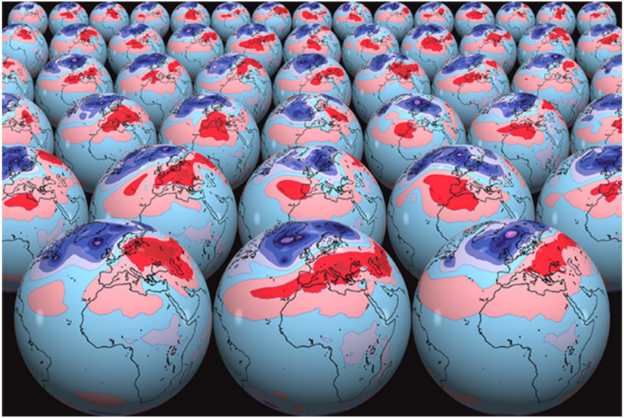


Interacting Particle Systems

Coupling of Measures

Ensemble prediction system with M members:

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t^i = f(Z_t^i), \qquad Z_0^i \sim \pi_0, \qquad i = 1, \ldots, M.$$



Source: ECMWF

Continuous–in–time **assimilation of precipitation data** y_t:

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t^i = f(Z_t^i) + \alpha_1 Q_t (Z_t^i - \overline{Z}_t) + \alpha_2 K_t (y_t - h(Z_t^i))$$

Additional terms:

- Inflation: $\alpha_1 > 0$, $Q_t \in \mathbb{R}^{N_z \times N_z}$ spd,

$$\overline{Z}_t = \frac{1}{M} \sum_i Z_t^i$$

- Nudging: $\alpha_2 > 0$, gain matrix $K_t \in \mathbb{R}^{N_z \times N_y}$, forward operator h.

Forward SDE

$$dZ_t^+ = f_t(Z_t^+)dt + \gamma^{1/2} dW_t^+ ,$$

$X_0^+ \sim \pi_0$, $t \in [0, T]$, W_t^+ standard Brownian motion forward in time.

Generates **probability measure** $\mathbb{P}_{[0,T]}$ over $C([0,T], \mathbb{R}^{N_z})$ with **marginal densities** π_t, i.e. $Z_t \sim \pi_t$.

The same measure is generated by **backward SDE**

$$dZ_t^- = b_t(Z_t^-)dt + \gamma^{1/2} dW_t^- ,$$

W_t^- Brownian motion backward in time, $X_T^- \sim \pi_T$.

It holds that

$$b_t(z) = f_t(z) - \gamma \nabla_z \log \pi_t(z) .$$

Fokker-Planck equation for marginals:

$$\begin{aligned}\partial_t \pi_t &= -\nabla_z \cdot (\pi_t f_t) + \frac{\gamma}{2}\Delta_z \pi_t \\ &= -\nabla_z \cdot (\pi_t b_t) - \frac{\gamma}{2}\Delta_z \pi_t \\ &= -\nabla_z \cdot (\pi_t u_t)\end{aligned}$$

with

$$u_t(z) = \frac{1}{2}(f_t(z) + b_t(z)) = f_t(z) - \frac{\gamma}{2}\nabla_z \log \pi_t(z)\,.$$

Replace forward SDE by **mean field equation**

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t = f_t(Z_t) - \frac{\gamma}{2}\nabla_z \log \pi_t(Z_t)\,, \qquad Z_0 \sim \pi_0\,.$$

Remark. Generates path measure $\mathbb{Q}_{[0,T]}$ which is different from SDE measure $\mathbb{P}_{[0,T]}$; only marginals π_t agree!

Lagrangian interacting particles (Gaussian approximation to π_t):

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t^i = f_t(Z_t^i) + \frac{\gamma}{2}(P_t)^{-1}(Z_t^i - \overline{Z}_t),$$

$Z_0^i \sim \pi_0$, $i = 1, \ldots, M$, empirical covariance matrix

$$P_t = \frac{1}{M-1} \sum_i (Z_t^i - \overline{Z}_t)(Z_t^i - \overline{Z}_t)^{\mathrm{T}}.$$

Connection to **inflation**:

$$Q_t = (P_t)^{-1}, \qquad \alpha_1 = \gamma/2.$$

- Other approximations of π_t, e.g. kernel methods, possible.
- Also used for **Bayesian inference**:

$$\mathrm{d}X_s = \nabla_x \log \pi(X_s|y)\,\mathrm{d}s + \sqrt{2}\mathrm{d}W_s$$

with $X_0 \sim \pi$ (prior), i.e., replace SDE by the mean-field system

$$\frac{\mathrm{d}}{\mathrm{d}s} X_s = u_s(X_s) := \nabla_x \log \frac{\pi(X_s|y)}{\pi_s(X_s)}.$$

References

- Daley, R., Atmospheric Data Analysis, Cambridge University Press, 1991.
- Nelson, E., Dynamical Theories of Brownian Motion, Princeton University Press, 1967.
- del Moral, P., Mean field simulation for Monte Carlo integration, CRC Press, 2013.
- SR, Cotter, J., Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.
- Law, K. et al., Data Assimilation – A Mathematical Introduction, Springer, 2015.
- Qiang Liu, Stein Variational Gradient Descent as Gradient Flow, arXiv:1704.07520v2, 2017.

Recap: Assimilation of precipitation data

Continuous–in–time **assimilation of precipitation data** y_t:

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t^i = f(Z_t^i) + \alpha_1 Q_t (Z_t^i - \overline{Z}_t) + \alpha_2 K_t (y_t - h(Z_t^i))$$

- Inflation: $\alpha_1 > 0$, $Q_t \in \mathbb{R}^{N_z \times N_z}$ spd,

$$\overline{Z}_t = \frac{1}{M} \sum_i Z_t^i$$

- Nudging: $\alpha_2 > 0$, gain matrix $K_t \in \mathbb{R}^{N_z \times N_y}$, forward operator h.

SDEs & nudging I

Given a **likelihood function**

$$L(z_{[0,T]}) := \exp\left(-\int_0^T V_t(z_t)\mathrm{d}t\right).$$

For example

$$V_t(z) = \frac{\beta}{2}\|h(z) - y_t\|^2 .$$

Bayes theorem (Radon–Nikodym):

$$\frac{\mathrm{d}\widehat{\mathbb{P}}_{[0,T]}}{\mathrm{d}\mathbb{P}_{[0,T]}}(z_{[0,T]}) := \frac{L(z_{[0,T]})}{\mathbb{P}_{[0,T]}[L]} .$$

The measure $\widehat{\mathbb{P}}_{[0,T]}$ solves the **filtering/smoothing problem** of SDE inference.

Mean–field formulation:

$$\mathrm{d}\widehat{Z}_t = \left\{ f_t(\widehat{Z}_t) + P_t \nabla_z \psi_t(\widehat{Z}_t) \right\} \mathrm{d}t + \sqrt{\gamma}\,\mathrm{d}W_t$$

with the potential ψ_t satisfying the elliptic PDE

$$\nabla_z \cdot (\widehat{\pi}_t P_t \nabla_z \psi_t) = \widehat{\pi}_t (V_t - \overline{V}_t)$$

$\widehat{Z}_0 \sim \widehat{\pi}_0 = \pi_0$, $\overline{V}_t = \widehat{\pi}_t[V_t]$, $P_t = \mathrm{cov}\,(\widehat{Z}_t)$.

If $\widehat{\pi}_t$ Gaussian and $h(z) = Hz$ in V_t, then

$$\nabla_z \psi_t(z) = \beta H^T \left(y_t - \frac{Hz + H\overline{Z}_t}{2} \right).$$

Compare to **nudging** scheme:

$$\alpha_2 K_t (y_t - Hz),$$

i.e., $K_t = P_t H^{\mathrm{T}}$, $\beta = \alpha_2$, but **innovation** different.

Combining nudging and inflation

Ensemble Kalman–Bucy filter (Gaussian approximation to π_t):

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t^i = f_t(Z_t^i) + \frac{\gamma}{2}(P_t)^{-1}(Z_t^i - \overline{Z}_t) + \beta K_t \left(y_t - \frac{h(Z_t^i) + \overline{h}_t}{2} \right)$$

$Z_0^i \sim \pi_0$, $i = 1, \ldots, M$, empirical covariance matrices

$$P_t = \frac{1}{M-1} \sum_i (Z_t^i - \overline{Z}_t)(Z_t^i - \overline{Z}_t)^{\mathrm{T}},$$

$$K_t = \frac{1}{M-1} \sum_i (Z_t^i - \overline{Z}_t)(h(Z_t^i) - \overline{h}_t)^{\mathrm{T}}.$$

Remark. Feedback particle filter for likelihood with

$$V_t(z)\mathrm{d}t \Rightarrow \frac{1}{2}\|h(z)\|^2 \mathrm{d}t - h(z)^{\mathrm{T}} \mathrm{d}y_t$$

(mean–field equations for Kushner-Zakai-Stratonovitch equation).

References

- Moser, J., On the volume elements on a manifold, Trans. Amer. Math. Soc., 1965.
- SR, A dynamical systems framework for intermittent data assimilation, BIT, 2010.
- Daum, F., Huang, J., Particle filter for nonlinear filters, ICASSP, IEEE, 2011.
- Bergemann, K, SR, An ensemble Kalman–Bucy filter for continuous–time data assimilation, Meteorolog. Zeitschrift, 2012.
- Yang, T. et al., Feedback particle filter, IEEE Trans. Autom. Control, 2013.
- Taghvaei, A. et al., Kalman filter and its modern extensions for the continuous–time nonlinear filtering problem, J. Dyn. Sys., Meas., and Control, 2018.
- de Wiljes, et al., Long–time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise, SIAM J. Appl. Dyn. Sys., 2018.

Discrete–time observations:

$$y_{t_n} = h(Z_{t_n}) + R^{1/2}\,\Xi_{t_n}, \qquad n = 1, \ldots, N\,.$$

Likelihood function:

$$L(z_{[0,T]}) := \exp\left(-\frac{1}{2}\sum_n (y_{t_n} - h(z_{t_n}))^{\mathrm{T}} R^{-1} (y_{t_n} - h(z_{t_n}))\right).$$

Bayes:

$$\frac{\mathrm{d}\widehat{\mathbb{P}}_{[0,T]}}{\mathrm{d}\mathbb{P}_{[0,T]}}(Z^+_{[0,T]}) := \frac{L(Z^+_{[0,T]})}{\mathbb{P}_{[0,T]}[L]}\,.$$

The measure $\widehat{\mathbb{P}}_{[0,T]}$ solves the **filtering/smoothing problem** of SDE inference.

For simplicity: **single observation**, i.e.

$$N = 1, \quad R = I, \quad t_1 = T, \qquad L(z) = \frac{1}{2}\|y_T - h(z)\|^2 .$$

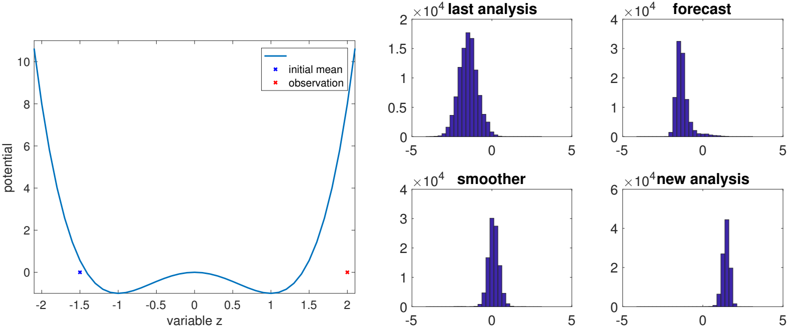
But keep recursive nature of sequential DA in mind!

Four **main players**:

- last analysis: π_0
- forecast based on last analysis: π_T
- new analysis at time $t = T$ (Bayes, filtering distribution): $\widehat{\pi}_T$
- smoothing distribution at $t = 0$: $\widehat{\pi}_0$

Standard sequential DA leads to a **discontinuous change** in distributions at observation time $t = T$ from π_T to $\widehat{\pi}_T$.

Scalar Brownian dynamics under a double well potential ($\gamma = 0.5$):



The forecast and the new analysis at $T = 0.5$ are nearly singular with respect to each other.

The relation between the last analysis (π_0) and the smoother ($\widehat{\pi}_0$) is somewhat better. Exploited in optimal proposal density/auxiliary particle filters.

Forward-backward smoother iteration:

- **Forward**:

$$dZ_t^+ = f(Z_t^+)dt + \sqrt{\gamma}W_t^+ ,$$

  $Z_0^+ \sim \pi_0$. **Yields** π_t.

- **Backward**:

$$d\widehat{Z}_t^- = f(\widehat{Z}_t^-)dt - \gamma\nabla_z \log \pi_t(\widehat{Z}_t^-)dt + \sqrt{\gamma}W_t^- ,$$

  with $\widehat{Z}_T^- \sim \widehat{\pi}_T$ and

$$\widehat{\pi}_T(z) \propto L(z)\pi_T(z) .$$

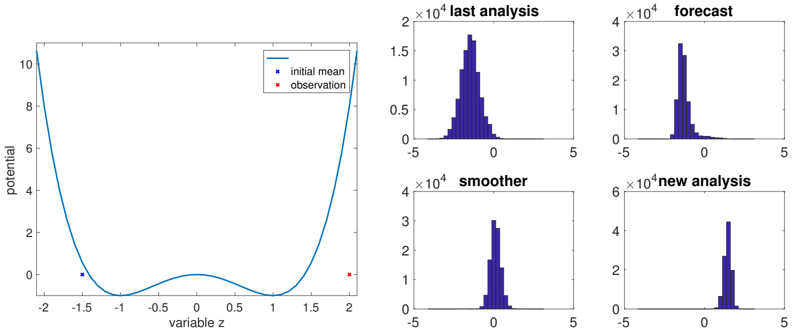
  Yields $\widehat{\pi}_t$.

Smoother:

$$d\widehat{Z}_t^+ = f(\widehat{Z}_t^+)dt + \gamma\nabla_z \log \frac{\widehat{\pi}_t}{\pi_t}(\widehat{Z}_t^+)dt + \sqrt{\gamma}W_t^+$$

$\widehat{Z}_0^+ \sim \widehat{\pi}_0$, $\widehat{Z}_T^+ \sim \widehat{\pi}_T$.

Scalar Brownian dynamics under a double well potential ($\gamma = 0.5$):



The forward smoother SDE links the smoother measure $\widehat{\pi}_0$ with $\widehat{\pi}_T$.

Still requires transforming π_0 into $\widehat{\pi}_0$ (but now at $t = 0$).

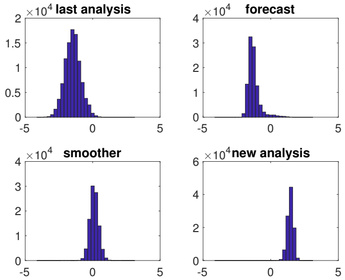
A different perspective on sequential DA:

Schrödinger problem. Find the measure $\widetilde{\mathbb{P}}_{[0,T]}$ which minimises the Kullback-Leibler divergence

$$\widetilde{\mathbb{P}}_{[0,T]} = \arg \inf_{\mathbb{Q} \ll \mathbb{P}} \mathrm{KL}(\mathbb{Q}_{[0,T]} || \mathbb{P}_{[0,T]})$$

subject to the constraints

$$\widetilde{\pi}_0 = q_0 = \pi_0, \qquad \widetilde{\pi}_T = q_T = \widehat{\pi}_T\,.$$



The measure $\widetilde{\mathbb{P}}_{[0,T]}$ is generated by a **controlled SDE**

$$\mathrm{d}\widetilde{Z}_t^+ = f(\widetilde{Z}_t^+)\mathrm{d}t + u_t(\widetilde{Z}_t^+)\mathrm{d}t + \sqrt{\gamma}\mathrm{d}W_t^+\,.$$

Find an initial distribution ϕ_0^+ and its evolution ϕ_t^+ under the forward SDE

$$\mathrm{d}Z_t^+ = f(Z_t^+)\,\mathrm{d}t + \sqrt{\gamma}\mathrm{d}W_t^+$$

such that the associated backward SDE

$$\mathrm{d}Z_t^- = \Big(f(Z_t^-) - \gamma\nabla_z \log \phi_t^+(Z_t^-)\Big)\,\mathrm{d}t + \sqrt{\gamma}\mathrm{d}W_t^-$$

with final condition $\phi_T^- := \widehat{\pi}_T$ leads to marginals ϕ_t^- such that

$$\phi_0^- = \pi_0\,.$$

Then the control

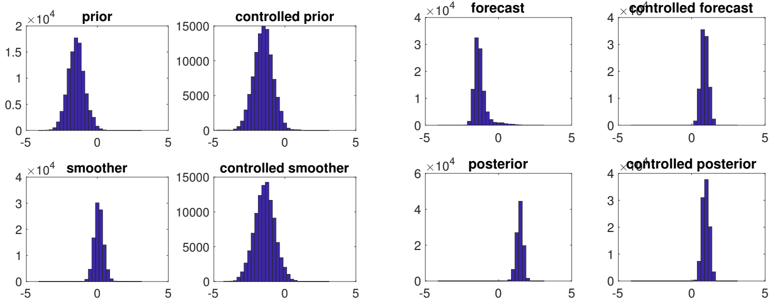
$$u_t = \gamma\nabla_z \log \frac{\phi_t^-}{\phi_t^+}$$

solves the Schrödinger problem.

Remark. *Just* smoothing with respect to a modified initial distribution $\phi_0^+ \neq \pi_0$!

Example

Double well potential, all densities are approximated as Gaussian (**linear, time-dependent control term**), ten iterations:



initial time

final time

Remarks

- **Smoothing**:

$$\phi_0^+ = \pi_0 \quad \& \quad \phi_T^- = \widehat{\pi}_T \quad \Rightarrow \quad \widehat{\pi}_t = \phi_t^-$$

  Schrödinger:

$$\phi_0^- = \pi_0 \quad \& \quad \phi_T^- = \widehat{\pi}_T \quad \Rightarrow \quad \phi_t^+ \;/\; u_t$$

- Link to **Sinkhorn** and **Robbins & Monro** iterations: If

$$\pi_0(z) = \frac{1}{M}\sum_{i=1}^{M} \delta(z - z_0^i)$$

  then

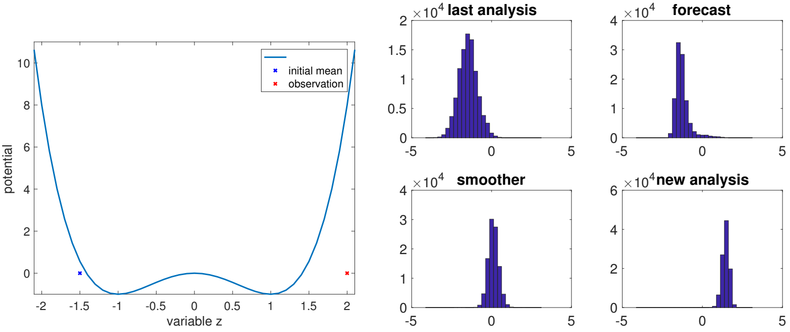
$$\phi_0^+(z) = \sum_{i=1}^{M} \alpha_i \delta(z - z_0^i), \qquad \sum_{i=1}^{M} \alpha_i = 1$$

  leading to a Sinkhorn fixed point iteration in the weights $\{\alpha_i\}$ which involves taking expectation with respect to $\widehat{\pi}_T$.

References

- Schrödinger, E., Über die Umkehrung der Naturgesetze, Sitzungsberichte der Preuss. Phys. Math. Klasse, 1932.
- Yongxin Chen et al., On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, J. Optim. Theory and Appl., 2016.
- Fearnhead, P., Künsch, H.R., Particle filters and data assimilation, Annual Review of Statistics and Its Application, 2018.
- Guarniero et al., The iterated auxiliary particle filter, J. American Statist. Assoc., 2017.
- van Leeuwen, P.-J., Nonlinear data assimilation, Frontiers in Applied Dynamical Systems, Vol. 2, 2015.
- Peyre, G., Cuturi, M., Computational Optimal Transport, arXiv:1803.00567, 2018.
- SR, Data assimilation, Acta Numerica, 2019.

We cannot, in general, implement the Schrödinger approach to sequential DA exactly.



Coupling of measures II

Available realisations $Z_T^i \sim \pi_T$ with **importance weights**

$$w^i \propto \frac{\widehat{\pi}_T}{\pi_T}(Z_T^i)\,.$$

Instead of **resampling**, find **coupling/transformation**

$$\widehat{Z}_T = \nabla_z \psi(Z_T)\,,$$

$Z_T \sim \pi_T$ and $\widehat{Z} \sim \widehat{\pi}_T$.

More abstractly,

$$\widehat{Z}_T(a) = \int Z_T(a')\delta(a' - \nabla_a \psi(a))\mathrm{d}a'\,,$$

where A is some random reference variable. For example, $A = Z_T$.

Replace the integral by a sum and formally write

$$\hat{Z}_T^j = \sum_{i=1}^{M} Z_T^i d_{ij}$$

with $a' \to i$, $Z_T(a') \to Z_T^i$, $a \to j$, $\hat{Z}_T(a) \to \hat{Z}_T^j$, $\delta(a' - \nabla_a \psi(a)) \to d_{ij}$. Need

$$\sum_{i=1}^{M} d_{ij} = 1, \qquad \frac{1}{M}\sum_{j} d_{ij} = w^i .$$

Select an "optimal" transformation through maximising correlation

$$V(D) = \frac{1}{M}\sum_{ij} d_{ij} Z_T^i \cdot Z_T^j = \frac{1}{M}\sum_{j} \hat{Z}_T^j \cdot Z_T^j .$$

In addition, either $d_{ij} \geq 0$ (**Ensemble Transform Particle Filter**) or

$$\frac{1}{M-1}\sum_{i=1}^{M} (\hat{Z}_T^i - \bar{\hat{z}}_T)(\hat{Z}_T^i - \bar{\hat{z}}_T)^{\mathrm{T}} = \sum_{i=1}^{M} w^i (Z_T^i - \bar{\hat{z}}_T)(Z_T^i - \bar{\hat{z}}_T)^{\mathrm{T}}$$

(**Nonlinear Ensemble Transform Filter**).

Numerical example I

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

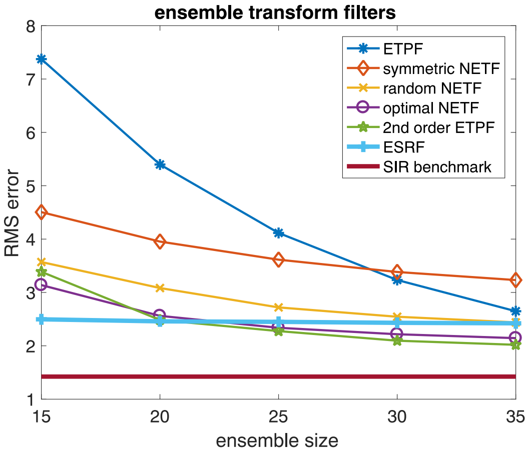


Figure: RMSEs for various second-order accurate LETFs compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, M.

Numerical example II

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

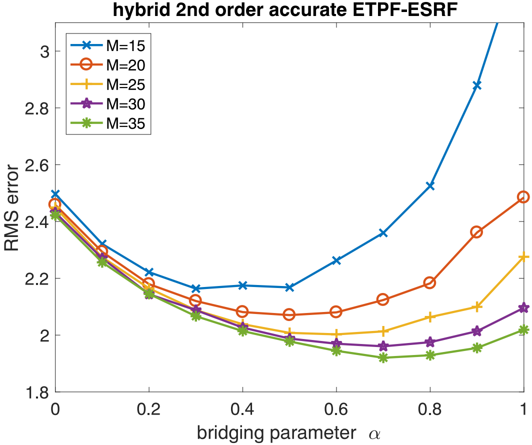


Figure: RMSEs for hybrid ESRF ($\alpha = 0$) and 2nd-order corrected NETF/ETPF ($\alpha = 1$) as a function of the sample size, M.

References

- Evensen, G., Data assimilation – the ensemble Kalman filter, Springer, 2009.
- SR, A nonparametric ensemble transform method for Bayesian inference, SIAM J. Sci. Comput., 2013.
- SR, Cotter, J., Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.
- Tödter J., Ahrens, B., A second–order exact ensemble square root filter for nonlinear data assimilation, Month. Weather Rev., 2015.
- Chustagulprom, N. et al., A hybrid ensemble transform particle filter for nonlinear and spatially extended systems, SIAM/ASA J. UQ, 2016.
- Acevedo, W., et al., Second–order accurate ensemble transform particle filters, SIAM J. Sci. Comput., 2017.
- Fearnhead, P., Künsch, H.R., Particle filters and data assimilation, Annual Review of Statistics and Its Application, 2018.

Summary

- Continuous–in–time DA naturally leads to various interacting particle systems.
- Schrödinger problem provides an "optimal" mathematical framework for sequential DA with discrete–in–time observations.
- Numerical implementation nontrivial; good drift corrections can be derived using Gaussian approximations or kernel methods.
- Coupling arguments are central to derivation of interacting particle systems.
- Relevant to rare event simulations, optimal control problems and derivative–free optimization.

Collaborators

- Walter Acevedo
- Kay Bergemann
- Yuan Cheng
- Nawinda Chustagulprom
- Colin Cotter
- Jana de Wiljes
- Prashant Mehta
- Wilhelm Stannat
- Amari Taghvaei