Data assimilation – Schrödinger’s perspective

Sebastian Reich (www.sfb1294.de)

Universität Potsdam/ University of Reading

IMS NUS, August 30, 2018
Core components of DA

Mean Field Equations

Interacting Particle Systems

Coupling of Measures
ÜBER DIE UMKEHRUNG DER NATURGESETZE

VON

E. SCHröDINGER

Sonderausgabe aus den Sitzungsberichten der preußischen Akademie der Wissenschaften

Phys.-Math. Klasse, 1931. IX
Ensemble prediction system with $M$ members:

$$\frac{d}{dt}Z_t^i = f(Z_t^i), \quad Z_0^i \sim \pi_0, \quad i = 1, \ldots, M.$$
Ensemble prediction II

Source: The quiet revolution of numerical weather prediction, Nature, 2015
Continuous-in-time **assimilation of precipitation data** $y_t$:

\[
\frac{d}{dt} Z^i_t = f(Z^i_t) + \alpha_1 Q_t(Z^i_t - \overline{Z}_t) + \alpha_2 K_t(y_t - h(Z^i_t))
\]

**Additional terms:**

- **Inflation:** $\alpha_1 > 0$, $Q_t \in \mathbb{R}^{N_z \times N_z}$ spd,
  
  \[
  \overline{Z}_t = \frac{1}{M} \sum_i Z^i_t
  \]

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

**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 \( C([0, T], \mathbb{R}^N) \) with **marginal densities** \( \pi_t \), 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:

\[ \partial_t \pi_t = -\nabla_z \cdot (\pi_t f_t) + \frac{\gamma}{2} \Delta_z \pi_t = -\nabla_z \cdot (\pi_t b_t) - \frac{\gamma}{2} \Delta_z \pi_t \]

\[ = -\nabla_z \cdot (\pi_t u_t) \]

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

\[ \frac{d}{dt} Z_t = f_t(Z_t) - \frac{\gamma}{2} \nabla_z \log \pi_t(Z_t), \quad Z_0 \sim \pi_0. \]

Remark. Generates path measure \( \Phi_{[0,T]} \) which is different from SDE measure \( \mathbb{P}_{[0,T]} \); only marginals \( \pi_t \) agree!
Lagrangian interacting particles (Gaussian approximation to \( \pi_t \)):

\[
\frac{d}{dt} Z^i_t = f_t(Z^i_t) + \frac{\gamma}{2} (P_t)^{-1}(Z^i_t - \bar{Z}_t),
\]

\( Z^i_0 \sim \pi_0, \ i = 1, \ldots, M \), empirical covariance matrix

\[
P_t = \frac{1}{M-1} \sum_{i} (Z^i_t - \bar{Z}_t)(Z^i_t - \bar{Z}_t)^T.
\]

Connection to inflation:

\[
Q_t = (P_t)^{-1}, \quad \alpha_1 = \gamma/2.
\]
References

Recap: Assimilation of precipitation data

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

$$\frac{d}{dt} Z^i_t = f(Z^i_t) + \alpha_1 Q_t (Z^i_t - \bar{Z}_t) + \alpha_2 K_t (y_t - h(Z^i_t))$$

- **Inflation:** $\alpha_1 > 0$, $Q_t \in \mathbb{R}^{N_z \times N_z}$ spd,
  $$\bar{Z}_t = \frac{1}{M} \sum_i Z^i_t$$

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

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

For example

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

**Bayes theorem** (Radon–Nikodym):

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

The measure \(\mathbb{P}_{[0,T]}\) solves the **filtering/smoothing problem** of SDE inference.
Mean-field formulation:

$$d\hat{Z}_t = \left\{ f_t(\hat{Z}_t) + P_t \nabla_z \psi_t(\hat{Z}_t) \right\} dt + \sqrt{\gamma} dW_t$$

with the potential $\psi_t$ satisfying the elliptic PDE

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

$\hat{Z}_0 \sim \hat{\pi}_0 = \pi_0, \overline{V}_t = \hat{\pi}_t [V_t], P_t = \text{cov} (\hat{Z}_t)$. 
If $\hat{\pi}_t$ Gaussian and $h(z) = Hz$ in $V_t$, then

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

Compare to nudging scheme:

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

i.e., $K_t = P_t H^T$, $\beta = \alpha_2$, but innovation different.
Combining nudging and inflation

**Ensemble Kalman–Bucy filter** (Gaussian approximation to $\pi_t$):

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

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

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

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

**Remark. Feedback particle filter** for likelihood with

$$
V_t(z)dt \Rightarrow \frac{1}{2} \|h(z)\|^2 dt - h(z)^T dy_t.
$$
References

Discrete–time observations I

![Diagram showing the process of state observation, model updating, and data assimilation over time.](image-url)
Discrete–time observations:

\[ y_{tn} = h(Z_{tn}) + R^{1/2} \Xi_{tn}, \quad n = 1, \ldots, N. \]

Likelihood function:

\[
L(Z_{[0,T]}) := \exp \left( -\frac{1}{2} \sum_{n} (y_{tn} - h(z_{tn}))^\top R^{-1} (y_{tn} - h(z_{tn})) \right).
\]

Bayes:

\[
\frac{d\mathbb{P}_{[0,T]}(Z^+_{[0,T]})}{d\mathbb{P}_{[0,T]}} := \frac{L(Z^+_{[0,T]})}{\mathbb{P}_{[0,T]}[L]}.
\]
For simplicity: **single observation**, i.e.

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

But keep recursive nature of sequential DA in mind!

**Four main players:**

- **last analysis**: \( \pi_0 \)
- **forecast** based on last analysis: \( \pi_T \)
- **new analysis** at time \( t = T \) (Bayes, filtering distribution): \( \hat{\pi}_T \)
- **smoothing distribution** at \( t = 0 \): \( \hat{\pi}_0 \)
Example: Gaussian mixture I

$\pi_1$ is a Gaussian mixture, Gaussian likelihood, $\hat{\pi}_1$ weighted Gaussian mixture.
Example: Gaussian mixture II

Applied to smoothing PDF $\hat{\pi}_0$.  
Applied to $\pi_0$ directly!
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 ($\pi_0$) and the smoother ($\hat{\pi}_0$) is somewhat better.
Forward-backward smoother iteration:

- **Forward:**
  \[ d\hat{Z}^+_t = f(\hat{Z}^+_t)dt + \sqrt{\gamma}W^+_t, \]
  \[ Z^+_0 \sim \pi_0. \text{ Yields } \pi_t. \]

- **Backward:**
  \[ d\hat{Z}^-_t = f(\hat{Z}^-_t)dt - \gamma \nabla_z \log \pi_t(\hat{Z}^-_t)dt + \sqrt{\gamma}W^-_t, \]
  with \( \hat{Z}^-_T \sim \hat{\pi}_T \) and
  \[ \hat{\pi}_T(z) \propto L(z) \pi_T(z). \]
  \text{Yields } \hat{\pi}_t.

**Smoother:**

\[ d\hat{Z}^+_t = f(\hat{Z}^+_t)dt + \gamma \nabla_z \log \frac{\hat{\pi}_t}{\pi_t}(\hat{Z}^+_t)dt + \sqrt{\gamma}W^+_t \]

\[ \hat{Z}^+_0 \sim \hat{\pi}_0, \hat{Z}^+_T \sim \hat{\pi}_T. \]
Scalar Brownian dynamics under a double well potential ($\gamma = 0.5$):

The forward smoother SDE links the smoother measure $\hat{\pi}_0$ with $\hat{\pi}_T$.
Still requires transforming $\pi_0$ into $\hat{\pi}_0$ (but now at $t = 0$).
A different perspective on sequential DA:

**Schrödinger problem.** Find the measure \( \tilde{P}_{[0,T]} \) which minimises the Kullback-Leibler divergence

\[
\tilde{P}_{[0,T]} = \arg \inf_{Q \ll P} KL(Q_{[0,T]} || P_{[0,T]})
\]

subject to the constraints

\[
\tilde{\pi}_0 = q_0 = \pi_0, \quad \tilde{\pi}_T = q_T = \tilde{\pi}_T.
\]

The measure \( \tilde{P}_{[0,T]} \) is generated by a **controlled SDE**

\[
d\tilde{Z}_t^+ = f(\tilde{Z}_t^+)dt + u_t(\tilde{Z}_t^+)dt + \sqrt{\gamma}dW_t^+.
\]
Find an initial distribution \( \phi_0^+ \) and its evolution \( \phi_t^+ \) under the forward SDE

\[
dZ_t^+ = f(Z_t^+) \, dt + \sqrt{\gamma} dW_t^+
\]

such that the associated backward SDE

\[
dZ_t^- = \left( f(Z_t^-) - \gamma \nabla_z \log \phi_t^+(Z_t^-) \right) \, dt + \sqrt{\gamma} dW_t^-
\]

with final condition \( \phi_T^- := \hat{\pi}_T \) leads to marginals \( \phi_t^- \) such that

\[
\phi_0^- = \pi_0.
\]

Then the control

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

solves the Schrödinger problem.
Example I: Lorenz–63

Starting point:

(i) Euler-Maruyama discretization

\[ Z_{n+1} = Z_n + f(Z_n)\Delta t + (\gamma \Delta t)^{1/2} \Xi_n \]

giving rise to transition kernel

\[ q_+(z_{n+1}|z_n). \]

(ii) Given \( M \) samples \( z^i_n \) from \( Z_n \) and \( M \) samples \( z^i_{n+1} \) from \( Z_{n+1} \) define the \( M \times M \) matrix \( Q \) by

\[ q_{ij} = q_+(z^i_{n+1}|z^j_n). \]

(iii) Observations at time \( t_{n+1} \) lead to importance weights \( w^i_{n+1} \).
Example II: Lorenz–63

Schrödinger problem:

\[ P^* = \arg \min_{P \in \Pi} \text{KL}(P \| Q), \quad \text{KL}(P \| Q) := \sum_{i,j=1}^{M} p_{ij} \log \frac{p_{ij}}{q_{ij}}, \]

subject to

\[ \Pi = \left\{ P \in \mathbb{R}^{M \times M} : p_{ij} \geq 0, \sum_{i=1}^{M} p_{ij} = 1/M, \sum_{j=1}^{M} p_{ij} = w_i \right\} \]

Application to DA

a) Schrödinger resample: Use \( P^* \) to sample from \( \hat{\pi}_{n+1} \)

\[ \mathbb{P} [ \hat{Z}_{n+1}^j = Z_{n+1}^j ] = M p^*_{ij} . \]

b) Schrödinger transform: Set

\[ \hat{Z}_{n+1}^j = M \sum_{i=1}^{M} z_{n+1}^i p^*_{ij} + (\gamma \Delta t) \Xi^j . \]
Stochastic Lorenz–63, observations at every time-step, observation error scaled with inverse time-step.
References


**Importance sampling (IS):** Available realizations $Z_T^i \sim \pi_T$ with importance weights

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

**Optimal transport (OT):** Instead of resampling, find coupling/transformation

$$\hat{Z}_T = \nabla_Z \psi(Z_T),$$

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

More abstractly,

$$\hat{Z}_T(a) = \int Z_T(a') \delta(a' - \nabla_a \psi(a)) da',$$

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

Replace the integral by a sum and formally write

\[ \hat{Z}_T^i = \sum_{i=1}^{M} Z_T^i d_{ij} \]

Requires

\[ \sum_{i=1}^{M} d_{ij} = 1, \quad \frac{1}{M} \sum_{j} d_{ij} = w^i. \]

Select an "optimal" transformation through maximising correlation

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

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

\[ \frac{1}{M-1} \sum_{i=1}^{M} (\hat{Z}_T^i - \hat{Z}_T)(\hat{Z}_T^i - \hat{Z}_T)^T = \sum_{i=1}^{M} w^i (Z_T^i - \hat{Z}_T)(Z_T^i - \hat{Z}_T)^T \]

(Nonlinear Ensemble Transform Filter).
Numerical example I

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

![Graph showing RMSEs for various second-order accurate LETF filters compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, $M$.]

**Figure:** RMSEs for various second-order accurate LETF filters compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, $M$. 
Numerical example II

Hybrid filter: \( P := P_{ESRF}(\alpha) P_{ETPF}(1-\alpha) \).

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

Universität Potsdam/ University of Reading 35
References

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.
"Miss Peterson, may I go home? I can't assimilate any more data today."

(source: J.B. Handelsman, New Yorker, 05/31/1969)
Collaborators

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