Data Assimilation: Interacting particle & mean-field formulations

Sebastian Reich

University of Potsdam & SFB 1294 Data Assimilation

September 29, 2023

Content

- Part I. Introduction to data assimilation (DA)
  - What is DA?
  - Examples
  - Applications and challenges
- Part II. The ensemble Kalman filter (EnKF)
  - Standard Kalman filter
  - Mean-field EnKF formulation & Monte Carlo implementation
  - Why the EnKF?
  - Why not?
- Part III. Control & DA
  - Coupling of measures
  - Homotopy and data-driven control

Reference: Ensemble Kalman Methods: A Mean Field Perspective; Edoardo Calvello, SR, Andrew M. Stuart, arXiv:2209.11371.

Data assimilation: Two sets of <u>(random) variables</u> (RV) $\{v_n\}$ and $\{\hat{v}_n\}$, $n \geq 0$, which are <u>coupled</u> through the following two alternating steps

(i) **Prediction**:

$$v_n \to \hat{v}_{n+1}$$

(ii) **Analysis**:

$$\hat{v}_{n+1} \to v_{n+1}.$$

Remarks

- v_n & $\hat{v}_n$ typically take values in $\mathbb{R}^d$.
- Both <u>continuous and discrete time</u> DA formulations exist.
- The iteration index n can represent physical or algorithmic time.

Probabilistic approach versus optimisation

Probabilistic: Sequence of RVs v_n and $\hat{v}_n$ with associated PDFs:

$$v_n \sim \mu_n \propto e^{-\Psi_n}$$
$$\hat{v}_n \sim \hat{\mu}_n \propto e^{-\hat{\Psi}_n}$$

Work at level of RVs and stochastic processes.

Probabilistic: Sequence of RVs v_n and $\hat{v}_n$ with associated PDFs:

$$v_n \sim \mu_n \propto e^{-\Psi_n}$$

$$\hat{v}_n \sim \hat{\mu}_n \propto e^{-\hat{\Psi}_n}$$

Work at level of RVs and stochastic processes.

Optimisation: Sequence of cost functions Φ_n and $\hat{\Phi}_n$ and induced <u>deterministic variables</u>

$$v_n = \arg\min_{v} \Phi_n(v)$$

$$\hat{v}_n = \arg\min_{\hat{v}} \hat{\Phi}_n(\hat{v})$$

Work at level of cost functions (cost-to-go) and optimisation.

Probabilistic: Sequence of RVs v_n and $\hat{v}_n$ with associated PDFs:

$$v_n \sim \mu_n \propto e^{-\Psi_n}$$
$$\hat{v}_n \sim \hat{\mu}_n \propto e^{-\hat{\Psi}_n}$$

Work at level of RVs and stochastic processes.

Optimisation: Sequence of cost functions Φ_n and $\hat{\Phi}_n$ and induced deterministic variables

$$v_n = \arg\min_{v} \Phi_n(v)$$
$$\hat{v}_n = \arg\min_{\hat{v}} \hat{\Phi}_n(\hat{v})$$

Work at level of cost functions (cost-to-go) and optimisation.

Remark. In linear Gaussian setting $\Phi_n = \Psi_n$ and $\hat{\Phi}_n = \hat{\Psi}_n$. Otherwise both approaches lead to different results, in general.

SFB 1294 Data Assimilation

A) **Markov process**: Stochastic process $\{v_n\}$ defined by

$$v_{n+1} = F(v_n, \xi_n)$$

with $\{\xi_n\}$ i.i.d. Gaussian and $F(v, \xi)$ a given map.

Task: Predict v_n given v_0.

A) **Markov process**: Stochastic process $\{v_n\}$ defined by

$$v_{n+1} = F(v_n, \xi_n)$$

with $\{\xi_n\}$ i.i.d. Gaussian and $F(v, \xi)$ a given map.

Task: Predict v_n given v_0.

B) **Bayesian inverse problem**: Observation model

$$y = h(\hat{v}) + \eta$$

with η Gaussian measurement error and $h(\hat{v})$ a given forward map.

Task: Infer $\hat{v}$ given observed $y = y^\dagger$ and a prior $\hat{\mu}$.

SFB 1294 Data Assimilation

DA formulation: Given a sequence of observations $\{y_n^\dagger\}$, $n \geq 1$:

(i) **Prediction step**: Markov process with $v_n \sim \mu_n$

$$\hat{v}_{n+1} = F(v_n, \xi_n).$$

DA formulation: Given a sequence of observations $\{y_n^\dagger\}$, $n \geq 1$:

(i) **Prediction step**: Markov process with $v_n \sim \mu_n$

$$\hat{v}_{n+1} = F(v_n, \xi_n).$$

(ii) **Analysis step**: Bayesian inference with prior $\hat{v}_{n+1} \sim \hat{\mu}_{n+1}$, negative log-likelihood

$$l_{n+1}(v) = \frac{1}{2}\left(h(v) - y_{n+1}^\dagger\right)^{\mathrm{T}} R^{-1}\left(h(v) - y_{n+1}^\dagger\right),$$

and posterior

$$v_{n+1} \sim \mu_{n+1}, \qquad \mu_{n+1}(v) = \frac{e^{-l_{n+1}(v)}\hat{\mu}_{n+1}(v)}{\hat{\mu}_{n+1}\left[e^{-l_{n+1}}\right]}.$$

Optimisation: Given a cost function $U(v)$ find the minimiser

$$v^\dagger = \arg\min_{v \in \mathbb{R}^d} U(v).$$

Optimisation: Given a cost function $U(v)$ find the minimiser

$$v^\dagger = \arg\min_{v\in\mathbb{R}^d} U(v).$$

Gradient descent:

$$v_{n+1} = v_n - \tau \nabla_v U(v_n)$$

with step-sizes $\tau > 0$.

Optimisation: Given a cost function $U(v)$ find the minimiser

$$v^{\dagger} = \arg\min_{v \in \mathbb{R}^d} U(v).$$

Gradient descent:

$$v_{n+1} = v_n - \tau \nabla_v U(v_n)$$

with step-sizes $\tau > 0$.

Remarks

- Requires derivatives.
- Local minima.
- Not affine invariant (unlike Newton's method).

Probabilistic optimisation: Given a step-size $\tau > 0$ & $v_0 \sim \mu_0$,

(i) **Prediction step**: $v_n \sim \mu_n$

$$\hat{v}_{n+1} = v_n + \sqrt{\epsilon\tau}\,\xi_n, \quad \epsilon > 0,$$

Probabilistic optimisation: Given a step-size $\tau > 0$ & $v_0 \sim \mu_0$,

(i) **Prediction step**: $v_n \sim \mu_n$

$$\hat{v}_{n+1} = v_n + \sqrt{\epsilon\tau}\,\xi_n, \quad \epsilon > 0,$$

(i) **Analysis step**: Prior $\hat{v}_{n+1} \sim \hat{\mu}_{n+1}$, negative log likelihood function $l(v) = U(v)$, and posterior

$$v_{n+1} \sim \mu_{n+1}, \qquad \mu_{n+1}(v) = \frac{e^{-\tau U(v)}\hat{\mu}_{n+1}(v)}{\hat{\mu}_{n+1}\left[e^{-\tau U}\right]}.$$

Remark: Desired

$$\lim_{n\to\infty} v_n \to v^\dagger.$$

Alternative **affine invariant prediction step**:

$$\hat{v}_{n+1} = v_n + \frac{\epsilon\tau}{2}(v_n - m_n^v), \qquad \epsilon > 0$$

with mean

$$m_n^v := \mu_n[v] = \mathbb{E}_n[v].$$

Remarks:

(i) mean update

$$\hat{m}_{n+1}^v = m_n^v.$$

(ii) covariance matrix update

$$\hat{C}_{n+1}^{vv} = \left(1 + \frac{\epsilon\tau}{2}\right)^2 C_n^{vv}.$$

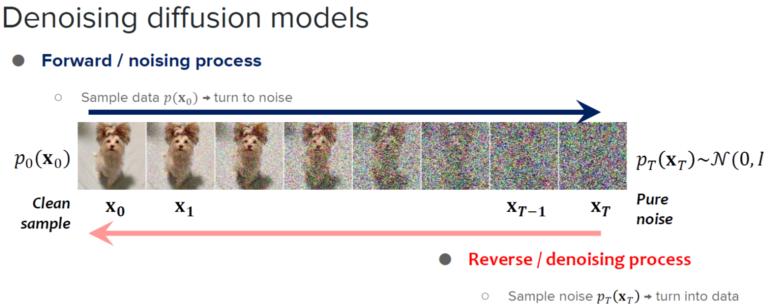


Figure: Courtesy: scholar.havard.edu

Diffusion modelling: Given samples $x_0^{(j)}$, $j = 1, \ldots, J$, from an unknown data distribution p_0.

Noising step: $t \in [0, 1]$,

$$\mathrm{d}x_t = -\frac{1}{2}\beta(t)x_t \mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t, \qquad x_0 = x_0^{(j)} \sim p_0,$$

with transition kernel

$$p_t(x_t|x_0) = \mathrm{N}(\sqrt{\alpha_t}x_0, v_t I), \;\; \alpha_t = \exp\left(-\int_0^t \beta(s)\mathrm{d}s\right), \;\; v_t = 1 - \alpha_t$$

Diffusion modelling: Given samples $x_0^{(j)}$, $j = 1, \ldots, J$, from an unknown data distribution p_0.

Noising step: $t \in [0, 1]$,

$$\mathrm{d}x_t = -\frac{1}{2}\beta(t)x_t\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t, \qquad x_0 = x_0^{(j)} \sim p_0,$$

with transition kernel

$$p_t(x_t|x_0) = \mathrm{N}(\sqrt{\alpha_t}x_0, v_t I), \ \alpha_t = \exp\left(-\int_0^t \beta(s)\mathrm{d}s\right), \ v_t = 1 - \alpha_t$$

Denoising step: $t \in [0, 1]$, $\mu_0 = p_T \approx \mathrm{N}(0, I)$, $\to \mu_1 \approx p_0$,

$$\mathrm{d}v_t = \frac{1}{2}\beta(t)v_t\mathrm{d}t + \beta(t)\nabla_v \log p_{T-t}(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t, \ v_0 \sim \mu_0.$$

SFB 1294 Data Assimilation

Conditional diffusion modelling: Sample from

$$p_0(x_0|y) \propto p(y|x_0)p_0(x_0),$$
$$-\log p(y|x) = \frac{1}{2}(y - Hv)^{\mathrm{T}} R^{-1}(Hv - y) + \text{const}.$$

Conditional diffusion modelling: Sample from

$$p_0(x_0|y) \propto p(y|x_0)p_0(x_0),$$
$$-\log p(y|x) = \frac{1}{2}(y - Hv)^{\mathrm{T}} R^{-1}(Hv - y) + \text{const.}$$

Prediction step. $\mu_0 = \mathrm{N}(0, I) \to v_1 \sim \hat{\mu}_1$:

$$\mathrm{d}v_t = \frac{1}{2}\beta(t)v_t\mathrm{d}t + \beta(t)\nabla_v \log p_{T-t}(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t$$

Analysis step. $\hat{\mu}_1 = p_0 \to \mu_1 = p_0(\cdot|y)$.

Mean-field perspective:

- All examples so far introduced two sequences of PDFs $\{\mu_n\}$ and $\{\hat{\mu}_n\}$, for $n \geq 0$. (Conditional diffusion model $n = 0, 1$.)

SFB 1294 Data Assimilation

Mean-field perspective:

- All examples so far introduced two sequences of PDFs $\{\mu_n\}$ and $\{\hat{\mu}_n\}$, for $n \geq 0$. (Conditional diffusion model $n = 0, 1$.)
- Those marginal PDFs do not define the desired stochastic processes in $\{v_n\}$ and $\{\hat{v}_n\}$ uniquely.

Mean-field perspective:

- All examples so far introduced two sequences of PDFs $\{\mu_n\}$ and $\{\hat{\mu}_n\}$, for $n \geq 0$. (Conditional diffusion model $n = 0, 1$.)
- Those marginal PDFs do not define the desired stochastic processes in $\{v_n\}$ and $\{\hat{v}_n\}$ uniquely.
- This conversion requires joint distributions.

SFB 1294 Data Assimilation

Mean-field perspective:

- All examples so far introduced two sequences of PDFs $\{\mu_n\}$ and $\{\hat{\mu}_n\}$, for $n \geq 0$. (Conditional diffusion model $n = 0, 1$.)
- Those marginal PDFs do not define the desired stochastic processes in $\{v_n\}$ and $\{\hat{v}_n\}$ uniquely.
- This conversion requires joint distributions.
- Consider mean-field processes of type

$$\textbf{Prediction}: \quad \hat{v}_{n+1} = \Phi_n(v_n, \mu_n, \xi_n),$$

$$\textbf{Analysis}: \quad v_{n+1} = \hat{\Phi}_n(\hat{v}_{n+1}, \hat{\mu}_{n+1}, \hat{\xi}_n),$$

for appropriate functions Φ_n, $\hat{\Phi}_n$; i.i.d. random variables ξ_n, $\hat{\xi}_n$; μ_n the law of v_n and $\hat{\mu}_n$ the law of $\hat{v}_n$.

Monte Carlo/particle implementations:

- Introduce J particles $v_{n,j}$ and $\hat{v}_{n,j}$, $j = 1, \ldots, J$, and their associated (random) empirical measures

$$\mu_n^J(v) = \frac{1}{J} \sum_{j=1}^{M} \delta(v - v_{n,j}), \qquad \hat{\mu}_n^J(v) = \frac{1}{J} \sum_{j=1}^{M} \delta(v - \hat{v}_{n,j})$$

at each iteration index $n \geq 0$.

Monte Carlo/particle implementations:

- Introduce J particles $v_{n,j}$ and $\hat{v}_{n,j}$, $j = 1, \dots, J$, and their associated (random) empirical measures

$$\mu_n^J(v) = \frac{1}{J}\sum_{j=1}^{M} \delta(v - v_{n,j}), \qquad \hat{\mu}_n^J(v) = \frac{1}{J}\sum_{j=1}^{M} \delta(v - \hat{v}_{n,j})$$

at each iteration index $n \geq 0$.

- Interacting particles update step: $n \geq 0$, $j = 1, \dots, M$

Prediction: $\hat{v}_{n+1,j} = \Phi_n^{\mathrm{a}}(v_{n,j}, \mu_n^J, \xi_{n,j}),$

Analysis: $v_{n+1,j} = \hat{\Phi}_n^{\mathrm{a}}(\hat{v}_{n+1,j}, \hat{\mu}_{n+1}^J, \hat{\xi}_{n,j}),$

with (generally) approximate maps Φ_n^{a} and $\hat{\Phi}_n^{\mathrm{a}}$.

Theory:

- Long time behaviour of μ_n and $\hat{\mu}_n$ as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on initial μ_0 as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on data $\{y_k^\dagger\}_{k=1}^n$.
- ...

Theory:

- Long time behaviour of μ_n and $\hat{\mu}_n$ as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on initial μ_0 as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on data $\{y_k^\dagger\}_{k=1}^n$.
- ...

Algorithms:

- Two types of approximation errors:
  - Monte Carlo sampling error
  - Approximate prediction and analysis steps Φ_n^{a} and $\hat{\Phi}_n^{\mathrm{a}}$, respectively.
- Error propagation and stability as $n \to \infty$.
- ...

Theory:

- Long time behaviour of μ_n and $\hat{\mu}_n$ as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on initial μ_0 as $n \to \infty$.
- Dependence of μ_n and $\hat{\mu}_n$ on data $\{y_k^\dagger\}_{k=1}^n$.
- ...

Algorithms:

- Two types of approximation errors:
  - Monte Carlo sampling error
  - Approximate prediction and analysis steps Φ_n^{a} and $\hat{\Phi}_n^{\mathrm{a}}$, respectively.
- Error propagation and stability as $n \to \infty$.
- ...

Remark. Related but also fundamentally different to sequential Monte-Carlo methods.

Content

- Part I. Introduction to data assimilation (DA)
  - What is DA?
  - Examples
  - Applications and challenges
- Part II. The ensemble Kalman filter (EnKF)
  - Standard Kalman filter
  - Mean-field EnKF formulation & Monte Carlo implementation
  - Why the EnKF?
  - Why not?
- Part III. Control & DA
  - Coupling of measures
  - Homotopy and data-driven control

Reference: Ensemble Kalman Methods: A Mean Field Perspective; Edoardo Calvello, SR, Andrew M. Stuart, arXiv:2209.11371.

Stochastic dynamical system:

$$v_{n+1} = Fv_n + b + B\xi_n,$$

$\xi_n \sim \mathrm{N}(0, I)$ i.i.d.

Stochastic dynamical system:

$$v_{n+1} = F v_n + b + B\xi_n,$$

$\xi_n \sim \mathrm{N}(0, I)$ i.i.d.

Observation model:

$$y = H\hat{v} + R^{1/2}\eta,$$

$\eta \sim \mathrm{N}(0, I)$ and prior distribution $\hat{v} \sim \mathrm{N}(\hat{m}^v, \hat{C}^{vv})$.

Gaussian DA:

$$v_n \sim \mathrm{N}(m_n^v, C_n^{vv}), \qquad \hat{v}_n \sim \mathrm{N}(\hat{m}_n^v, \hat{C}_n^{vv}).$$

Gaussian DA:

$$v_n \sim \mathrm{N}(m_n^v, C_n^{vv}), \qquad \hat{v}_n \sim \mathrm{N}(\hat{m}_n^v, \hat{C}_n^{vv}).$$

Prediction step:

$$\hat{m}_{n+1}^v = F m_n^v + b,$$
$$\hat{C}_{n+1}^{vv} = F C_n^{vv} F^{\mathrm{T}} + B B^{\mathrm{T}}.$$

Gaussian DA:

$$v_n \sim \mathrm{N}(m_n^v, C_n^{vv}), \qquad \hat{v}_n \sim \mathrm{N}(\hat{m}_n^v, \hat{C}_n^{vv}).$$

Prediction step:

$$\hat{m}_{n+1}^v = F m_n^v + b,$$
$$\hat{C}_{n+1}^{vv} = F C_n^{vv} F^{\mathrm{T}} + B B^{\mathrm{T}}.$$

Analysis step: Given data $y_{n+1}^{\dagger}$

$$m_{n+1}^v = \hat{m}_{n+1}^v + K_{n+1}(y_{n+1}^{\dagger} - H \hat{m}_{n+1}^v),$$
$$C_{n+1}^{vv} = \hat{C}_{n+1}^{vv} - K_{n+1} H \hat{C}_{n+1}^{vv}.$$

with Kalman gain

$$K_{n+1} = \hat{C}_{n+1}^{vy} (\hat{C}_n^{yy})^{-1} = \hat{C}_{n+1}^{vv} H^{\mathrm{T}} (H \hat{C}_{n+1}^{vv} H^{\mathrm{T}} + R)^{-1}.$$

Gaussian mean-field DA:

$$v_n \to \hat{v}_{n+1}, \qquad \hat{v}_{n+1} \to v_{n+1}.$$

SFB 1294 Data Assimilation

Gaussian mean-field DA:

$$v_n \to \hat{v}_{n+1}, \qquad \hat{v}_{n+1} \to v_{n+1}.$$

Prediction & output step: $v_n \sim \mu_n$ & ξ_n, η_n i.i.d. $\mathrm{N}(0, I)$

$$\hat{v}_{n+1} = Fv_n + b + B\xi_n,$$
$$\hat{y}_{n+1} = H\hat{v}_{n+1} + R^{1/2}\eta_n.$$

Gaussian mean-field DA:

$$v_n \to \hat{v}_{n+1}, \qquad \hat{v}_{n+1} \to v_{n+1}.$$

Prediction & output step: $v_n \sim \mu_n$ & ξ_n, η_n i.i.d. $\mathrm{N}(0, I)$

$$\hat{v}_{n+1} = Fv_n + b + B\xi_n,$$
$$\hat{y}_{n+1} = H\hat{v}_{n+1} + R^{1/2}\eta_n.$$

Analysis step: Given data $y^\dagger_{n+1}$, linear mean-field update

$$v_{n+1} = \hat{v}_{n+1} + K_{n+1}(y^\dagger_{n+1} - \hat{y}_{n+1})$$

with Kalman gain

$$K_{n+1} = \hat{C}^{vy}_{n+1}\left(\hat{C}^{yy}_{n+1}\right)^{-1} = \hat{C}_{n+1}H^{\mathrm{T}}(H\hat{C}_{n+1}H^{\mathrm{T}} + R)^{-1}.$$

EnKF: J particles $v_{n,j}$, $\hat{v}_{n,j}$, $j = 1, \ldots, J$, empirical estimators

$$\mathrm{m}_n^v := \mu_n^J[\hat{v}] = \mathbb{E}_n^J[\hat{v}], \ \hat{\mathrm{C}}_n^{vy} := \mathbb{E}_n^J\left[(\hat{v} - \mathbb{E}_n^J[\hat{v}])(\hat{y} - \mathbb{E}_n^J[\hat{y}])^{\mathrm{T}}\right], \text{ etc.}$$

EnKF: J particles $v_{n,j}$, $\hat{v}_{n,j}$, $j = 1, \ldots, J$, empirical estimators

$$\mathrm{m}_n^v := \mu_n^J[\hat{v}] = \mathbb{E}_n^J[\hat{v}], \ \hat{\mathrm{C}}_n^{vy} := \mathbb{E}_n^J\left[(\hat{v} - \mathbb{E}_n^J[\hat{v}])(\hat{y} - \mathbb{E}_n^J[\hat{y}])^{\mathrm{T}}\right], \text{ etc.}$$

Prediction & output step: $j = 1, \ldots, J$, $\xi_{n,j}$, $\eta_{n,j}$ i.i.d. $\mathrm{N}(0, I)$

$$\hat{v}_{n+1,j} = Fv_{n,j} + b + B^{\mathrm{T}}\xi_{n,j},$$
$$\hat{y}_{n+1,j} = H\hat{v}_{n+1,j} + R^{1/2}\eta_{n,j}.$$

SFB 1294 Data Assimilation

EnKF: J particles $v_{n,j}$, $\hat{v}_{n,j}$, $j = 1, \dots, J$, empirical estimators

$$\mathrm{m}_n^v := \mu_n^J[\hat{v}] = \mathbb{E}_n^J[\hat{v}],\ \hat{\mathrm{C}}_n^{vy} := \mathbb{E}_n^J\left[(\hat{v} - \mathbb{E}_n^J[\hat{v}])(\hat{y} - \mathbb{E}_n^J[\hat{y}])^{\mathrm{T}}\right],\ \text{etc.}$$

Prediction & output step: $j = 1, \dots, J$, $\xi_{n,j}$, $\eta_{n,j}$ i.i.d. $\mathrm{N}(0, I)$

$$\hat{v}_{n+1,j} = F v_{n,j} + b + B^{\mathrm{T}} \xi_{n,j},$$
$$\hat{y}_{n+1,j} = H \hat{v}_{n+1,j} + R^{1/2} \eta_{n,j}.$$

EnKF analysis step: Given data $y_{n+1}^{\dagger}$, $j = 1, \dots, J$,

$$v_{n+1,j} = \hat{v}_{n+1,j} + \mathrm{K}_{n+1}(y_{n+1}^{\dagger} - \hat{y}_{n+1,j})$$

with (empirical) Kalman gain

$$\mathrm{K}_{n+1} := \hat{\mathrm{C}}_{n+1}^{vy} \left(\hat{\mathrm{C}}_{n+1}^{yy}\right)^{-1}.$$

Applications of the EnKF:

- Often used with $J \ll d$ when $d \gg 1$ (model reduction). Additional computational hacks required (localisation, inflation).

SFB 1294 Data Assimilation

Applications of the EnKF:

- Often used with $J \ll d$ when $d \gg 1$ (model reduction). Additional computational hacks required (localisation, inflation).
- Can be extended to nonlinear DA problems.
  - **Prediction & output step**:

$$\hat{v}_{n+1} = \Phi(v_n, \xi_n),$$
$$\hat{y}_{n+1} = h(\hat{v}_{n+1}) + R^{1/2}\eta_n.$$

  - **EnKF analysis step**: unaltered.

Applications of the EnKF:

- Often used with $J \ll d$ when $d \gg 1$ (model reduction). Additional computational hacks required (localisation, inflation).
- Can be extended to nonlinear DA problems.
  - **Prediction & output step**:

$$\hat{v}_{n+1} = \Phi(v_n, \xi_n),$$
$$\hat{y}_{n+1} = h(\hat{v}_{n+1}) + R^{1/2}\eta_n.$$

  - **EnKF analysis step**: unaltered.
- The EnKF is affine invariant (like Newton's method) and derivative-free.

SFB 1294 Data Assimilation

Applications of the EnKF:

- Often used with $J \ll d$ when $d \gg 1$ (model reduction). Additional computational hacks required (localisation, inflation).
- Can be extended to nonlinear DA problems.
  - **Prediction & output step**:

$$\hat{v}_{n+1} = \Phi(v_n, \xi_n),$$
$$\hat{y}_{n+1} = h(\hat{v}_{n+1}) + R^{1/2}\eta_n.$$

  - **EnKF analysis step**: unaltered.
- The EnKF is affine invariant (like Newton's method) and derivative-free.
- **But why like this? Best Unbiased Linear Estimator** (BLUE) .

Prediction step: $(\hat{v}_{n+1}, \hat{y}_{n+1})$ with potentially unknown distribution (generative model)

Best linear (point) estimator for $\hat{v}_{n+1}$ given $\hat{y}_{n+1}$:

$$e(y) = b + Ky$$

with

$$(b, K) = \arg\min \mathbb{E}\left[\|\hat{v}_{n+1} - e(\hat{y}_{n+1})\|^2\right].$$

Plug-in estimator: Given samples $\{(\hat{v}_{n,j}, \hat{y}_{n,j})\}$

$$\mathrm{K} = \hat{\mathrm{C}}^{vy}_{n+1}(\hat{\mathrm{C}}^{yy}_{n+1})^{-1}$$
$$\mathrm{b} = \hat{\mathrm{m}}^{v}_{n+1} - \mathrm{K}\hat{\mathrm{m}}^{y}_{n+1}$$

Remark. Corresponds to update of the ensemble mean in an EnKF.

Data generation: $v^\dagger_{n+1}$ true value (fixed),

$$y^\dagger_{n+1} = Hv^\dagger_{n+1} + R^{1/2}\eta^\dagger_{n+1}, \qquad \eta^\dagger_{n+1} \sim \mathrm{N}(0, I).$$

Data generation: $v^\dagger_{n+1}$ true value (fixed),

$$y^\dagger_{n+1} = Hv^\dagger_{n+1} + R^{1/2}\eta^\dagger_{n+1}, \qquad \eta^\dagger_{n+1} \sim \mathrm{N}(0, I).$$

Bias of BLUE:

$$\mathbb{E}^\dagger\left[e(y^\dagger_{n+1}) - v^\dagger_{n+1}\right] = (I - KH)(\hat{m}^v_{n+1} - v^\dagger_{n+1}).$$

Covariance of BLUE:

$$\mathrm{var}^\dagger[e(y^\dagger_{n+1})] = KRK^{\mathrm{T}}.$$

Data generation: $v^\dagger_{n+1}$ true value (fixed),

$$y^\dagger_{n+1} = Hv^\dagger_{n+1} + R^{1/2}\eta^\dagger_{n+1}, \qquad \eta^\dagger_{n+1} \sim \mathrm{N}(0, I).$$

Bias of BLUE:

$$\mathbb{E}^\dagger\left[e(y^\dagger_{n+1}) - v^\dagger_{n+1}\right] = (I - KH)(\hat{m}^v_{n+1} - v^\dagger_{n+1}).$$

Covariance of BLUE:

$$\mathrm{var}^\dagger[e(y^\dagger_{n+1})] = KRK^{\mathrm{T}}.$$

Remark. Classical bias-variance trade-off of frequentist uncertainty quantification (UQ).

EnKF: Propagation of frequentist UQ

Challenge: In DA we need more than an estimator. We also need to propagate its uncertainty!

EnKF: Propagation of frequentist UQ

SFB 1294 Data Assimilation

Challenge: In DA we need more than an estimator. We also need to propagate its uncertainty!

Solution: Introduce random variable v_{n+1} such that

$$
\begin{aligned}
m^{v}_{n+1} &= e(y^{\dagger}_{n+1}),\\
C^{vv}_{n+1} &\geq KRK^{\mathrm{T}},\\
C^{vv}_{n+1} &\leq \hat{C}^{vv}_{n+1}.
\end{aligned}
$$

Mean-field EnKF:

$$
\begin{aligned}
\hat{y}_{n+1} &= H\hat{v}_{n+1} + R^{1/2}\eta_{n+1},\\
v_{n+1} &= \hat{v}_{n+1} + K(y^{\dagger}_{n+1} - \hat{y}_{n+1}).
\end{aligned}
$$

since

$$
C^{vv}_{n+1} = (I - KH)\hat{C}^{vv}_{n+1}(I - KH)^{\mathrm{T}} + KRK^{\mathrm{T}} = (I - KH)\hat{C}^{vv}_{n+1}.
$$

SFB 1294 Data Assimilation

Drift-diffusion model

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t$$

$\sigma > 0$, w_t Brownian motion.

Drift-diffusion model

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t$$

$\sigma > 0$, w_t Brownian motion.

Given:

i) **Initial conditions**

$$v_0 \sim \mu_0,$$

ii) **Observation at time $t = T > 0$**

$$y_T = Hv_T + \eta_T, \qquad \eta_T \sim \mathrm{N}(0, R),$$

H the (linear) forward model.

Two-dimensional nonlinear diffusion: $v = (v_1, v_2)^\top$.

Drift term:

$$f(v) = -\nabla V(v), \quad V(v) = \frac{\lambda_1}{2}\left(v_2 - 2 + \beta(v_1)^2\right)^2 + \frac{\lambda_2}{2}\left(\frac{(v_1)^4}{2} - (v_1)^2\right)$$

with parameters $\lambda_1 = 2000$, $\lambda_2 = 5$, and $\beta = 1/5$, diffusion constant $\sigma = 1$.

Two-dimensional nonlinear diffusion: $v = (v_1, v_2)^\top$.

Drift term:

$$f(v) = -\nabla V(v), \quad V(v) = \frac{\lambda_1}{2}\left(v_2 - 2 + \beta (v_1)^2\right)^2 + \frac{\lambda_2}{2}\left(\frac{(v_1)^4}{2} - (v_1)^2\right)$$

with parameters $\lambda_1 = 2000$, $\lambda_2 = 5$, and $\beta = 1/5$, diffusion constant $\sigma = 1$.

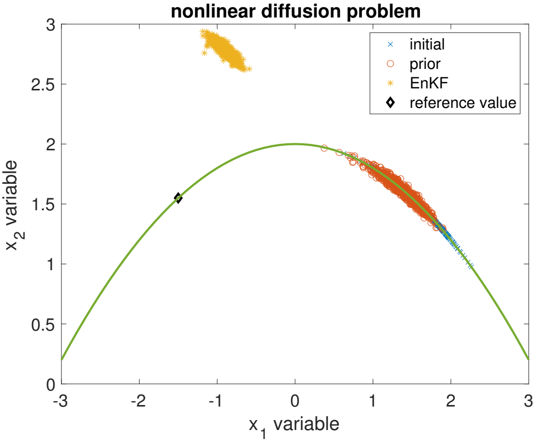
Initial distribution:

$$v_{1,0} \sim \mathrm{N}(1.5, 0.0625), \qquad v_{2,0} = 2 - \beta v_{1,0}^2.$$

Observation:

$$H = (1\ 0), \quad T = 1, \quad R = 0.01, \quad y_T^\dagger = -1.5.$$

Why going beyond EnKF?



Drift-diffusion model

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + g_t(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t.$$

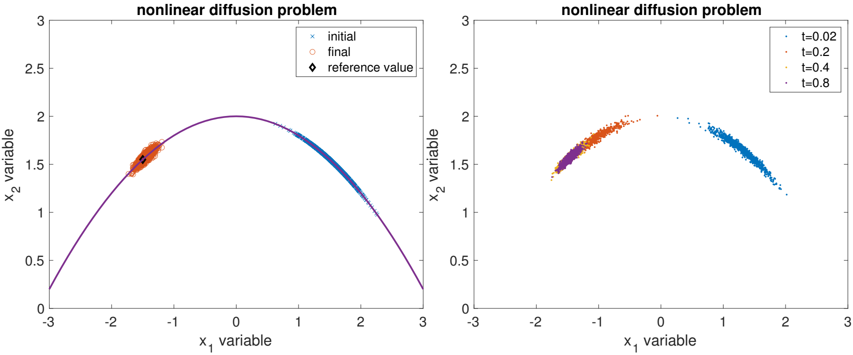


Figure: Left panel: Initial and final particle positions under the controlled evolution process. Right panel: Particle positions at intermediate times

Content

SFB 1294 Data Assimilation

- Part I. Introduction to data assimilation (DA)
  - What is DA?
  - Examples
  - Applications and challenges
- Part II. The ensemble Kalman filter (EnKF)
  - Standard Kalman filter
  - Mean-field EnKF formulation & Monte Carlo implementation
  - Why the EnKF & why not?
- Part III. Control & DA
  - Coupling of measures
  - Homotopy and data-driven control

References:

SR, Data assimilation: A dynamic homotopy-based coupling approach, arXiv:2209.05279, to appear STUOD Proceedings, 2023.

Chen, Y. et al, Gradient Flows for Sampling: Mean-Field Models, Gaussian Approximations and Affine Invariance, arXiv:2302.11024.

Prior (prediction): Random variable $\hat{v}$ with

$$\hat{v} \sim \hat{\mu}$$

negative log-likelihood:

nonl. regression: $$I(v) = \frac{1}{2}(h(v) - y^\dagger)^{\mathrm{T}} R^{-1}(h(v) - y^\dagger),$$

h nonlinear forward map, R error covariance matrix, $y^\dagger \in \mathbb{R}^{N_y}$ the data.

Bayesian posterior (analysis):

$$\mu(v) := \frac{e^{-I(v)}\,\hat{\mu}(v)}{\hat{\mu}\,[e^{-I}]}.$$

Find joint distribution π with desired marginals, i.e.,

$$(\hat{v}, v) \sim \pi$$

such that

$$\pi(\hat{v}, v) = \pi(v|\hat{v})\,\hat{\mu}(\hat{v}) = \pi(\hat{v}|v)\mu(v). \tag{1}$$

The set of all such joint distributions is denoted by Π.

Find joint distribution π with desired marginals, i.e.,

$$(\hat{v}, v) \sim \pi$$

such that

$$\pi(\hat{v}, v) = \pi(v|\hat{v})\,\hat{\mu}(\hat{v}) = \pi(\hat{v}|v)\mu(v). \tag{1}$$

The set of all such joint distributions is denoted by Π.

Catch: Π is not empty: $\pi(\hat{v}, v) = \hat{\mu}(\hat{v})\mu(v) \in \Pi$.

Selection of a unique π^* via:

- **optimal transportation** (minimise expected distance between $\hat{v}$ and v)
- **Schrödinger bridges** (minimise the Kullback–Leibler divergence to some reference measure)
- **homotopy** (Fisher–Rao gradient flow)

Coupling via homotopy I

Idea:[1] Define transport map T as the time-one flow map of a (mean-field) ODE

$$\frac{\mathrm{d}}{\mathrm{d}\tau} v_\tau = f_\tau(v_\tau), \qquad v_0 \sim \hat{\mu}, \quad \tau \in [0,1].$$

Liouville equation for $v_\tau \sim \pi_\tau$:

$$\partial_\tau \pi_\tau = -\nabla \cdot (\pi_\tau f_\tau)$$

Problem: How to fix f_τ appropriately given $\pi_0 = \hat{\mu}$ and $\pi_1 = \mu$?

[1] Daum & Huang, ICASSP, 2011; SR, BIT, 2011, see also normalising flows in ML

Homotopy Bayes:

$$\pi_\tau(v) \propto e^{-\tau l(v)} \hat{\mu}(v).$$

implying

$$\partial_\tau \pi_\tau = -\pi_\tau \left(l - \pi_\tau[l]\right)$$

Fisher–Rao gradient flow:

$$\partial_\tau \pi_\tau = -\text{grad}\, \mathcal{E}(\pi_\tau)$$

with functional

$$\mathcal{E}(\pi) = \int l(v)\pi(v)\mathrm{d}v.$$

Mean-field equations in the parameters $v \in \mathbb{R}^d$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau} v_\tau = f(v_\tau; \pi_\tau)$$
$$\nabla \cdot (\pi_\tau f) = \operatorname{grad} \mathcal{E}(\pi_\tau),$$

where π_τ denotes the law of v_τ and $v_0 \sim \hat{\mu}$.

Remark: Mean-field drift term $f(v; \pi)$ is **not** uniquely determined.

SFB 1294 Data Assimilation

Mean-field equations in the parameters $v \in \mathbb{R}^d$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau} v_\tau = f(v_\tau; \pi_\tau)$$
$$\nabla \cdot (\pi_\tau f) = \operatorname{grad} \mathcal{E}(\pi_\tau),$$

where π_τ denotes the law of v_τ and $v_0 \sim \hat{\mu}$.

Remark: Mean-field drift term $f(v;\pi)$ is **not** uniquely determined.

Affine invariant choices: Find potential function ψ such that

$$f(v;\pi) = (\nabla^2 \psi)^{-1} \nabla \psi(v;\pi) \qquad \text{or} \qquad f(v;\pi) = C^{vv} \nabla \psi(v;\pi)$$

with covariance matrix C^{vv}.

Drift-diffusion model

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t$$

$\sigma > 0$, w_t Brownian motion.

Drift-diffusion model

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t$$

$\sigma > 0$, w_t Brownian motion.

Given:

i) **Initial conditions**

$$v_0 \sim \mu_0,$$

ii) **Observations at time $t = T > 0$**

$$y_T = Hv_T + \eta_T, \qquad \eta_T \sim \mathrm{N}(0, R),$$

H the (linear) forward model.

SFB 1294 Data Assimilation

Controlled SDE

$$\mathrm{d}\tilde{v}_t = f(\tilde{v}_t)\mathrm{d}t + g_t(\tilde{v}_t)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}w_t$$

for appropriate control $g_t(v)$, $\tilde{v}_t \sim \tilde{\mu}_t$.

SFB 1294 Data Assimilation

Controlled SDE

$$\mathrm{d}\tilde{v}_t = f(\tilde{v}_t)\mathrm{d}t + g_t(\tilde{v}_t)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}w_t$$

for appropriate control $g_t(v)$, $\tilde{v}_t \sim \tilde{\mu}_t$.

Schrödinger bridge problem: Find control g_t minimizing Kullback-Leibler divergence

$$\mathrm{KL}(\hat{\mu}|\tilde{\mu}) = \int_0^T \int_{\mathbb{R}^d} \hat{\mu}_t(v)(\log \hat{\mu}_t(v) - \log \tilde{\mu}_t(v))\mathrm{d}v\,\mathrm{d}t$$

subject to

$$\tilde{\mu}_0 = \mu_0, \qquad \tilde{\mu}_T = \mu_T.$$

Homotopy:

$$\tilde{\mu}_t(v) = Z_t^{-1} e^{-\frac{t}{T} l(v)} \hat{\mu}_t(v)$$

with $Z_t = \int e^{-\frac{t}{T} l(v)} \hat{\mu}_t(v) \mathrm{d}x$. Obviously

$$\tilde{\mu}_0 = \mu_0, \qquad \tilde{\mu}_T = \mu_T.$$

Homotopy:

$$\tilde{\mu}_t(v) = Z_t^{-1} e^{-\frac{t}{T} I(v)} \hat{\mu}_t(v)$$

with $Z_t = \int e^{-\frac{t}{T} I(v)} \hat{\mu}_t(v) \mathrm{d}x$. Obviously

$$\tilde{\mu}_0 = \mu_0, \qquad \tilde{\mu}_T = \mu_T.$$

Given $v_t \sim \hat{\mu}_t$:

$$\mathrm{d}v_t = f(v_t)\mathrm{d}t + (2\sigma)^{1/2}\mathrm{d}w_t, \qquad v_0 \sim \mu_0$$

Desired $v_t^{\mathrm{h}} \sim \pi_t^{\mathrm{h}}$:

$$\mathrm{d}\tilde{v}_t = f(\tilde{v}_t)\mathrm{d}t - \frac{\sigma t}{T} \nabla I(\tilde{v}_t)\mathrm{d}t + g_t(\tilde{v}_t)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}w_t$$

for **appropriate control** $g_t(v)$.

Desired $\tilde{v}_t \sim \tilde{\mu}_t$:

$$\mathrm{d}\tilde{v}_t = f(\tilde{v}_t)\mathrm{d}t - \frac{2\sigma t}{T}\nabla l(\tilde{v}_t)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}w_t + g_t(\tilde{v}_t)\mathrm{d}t$$

for **appropriate control** $g_t(v)$.

Control g_t satisfies PDE

$$\nabla \cdot (\tilde{\mu}_t g_t) = \frac{1}{T}\tilde{\mu}_t \left(l + t\nabla l \cdot f - \frac{\sigma t^2}{T}\|\nabla l\|^2 \right) + \tilde{\mu}_t \frac{\dot{Z}_t}{Z_t}.$$

SFB 1294 Data Assimilation

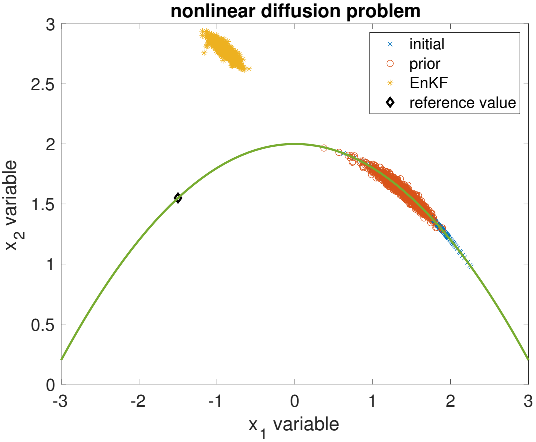
Constant gain approximation (EnKF-like):

$$g_t^{\mathrm{KF}}(v) = -K_t \left(\frac{1}{2} (Hv + Hm_t) - y_T^{\dagger} \right) - \frac{t}{2T} C_t^{vv} H^\top R^{-1} H \left(f(x) + \tilde{\mu}_t[f] \right)$$

with **Kalman gain**

$$K_t = \left\{ \frac{1}{T} C_t^{vv} + \frac{t}{T} C_t^{vf} - \frac{2\sigma t^2}{T^2} C_t^{vv} H^\top R^{-1} H \right\} H^\top R^{-1}.$$

Nonlinear example I



Nonlinear example II

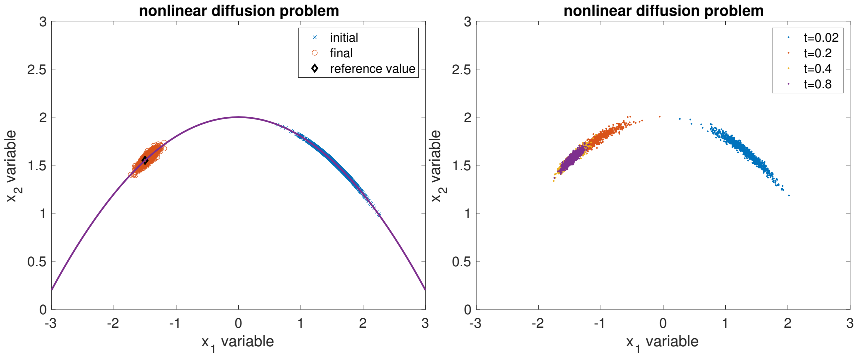


Figure: Left panel: Initial and final particle positions under the controlled evolution process. Right panel: Particle positions at intermediate times $t_k \in [0, 1]$.

Lorenz 63 model

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = f(X_t), \qquad f(x,y,z) = \begin{pmatrix} a(y-x) \\ x(b-z)-y \\ xy-cz \end{pmatrix} \tag{2}$$

with parameters $a = 10$, $b = 28$ and $c = 8/3$. The first component is observed.

$M/\Delta t_{\mathrm{obs}}$	5	10	15
0.05	0.5712/0.5457	0.5620/0.5475	0.5659/0.5496
0.10	0.8466/0.7735	0.8171/0.7627	0.8229/0.7707
0.12	0.9606/0.8645	0.9515/0.8621	0.9375/0.8615

Table: RMSE for the **ensemble square root filter** and the **homotopy approach** in terms of ensemble sizes $M \in \{5, 10, 15\}$ and observation intervals $\Delta t_{\mathrm{obs}} \in \{0.05, 0.1, 0.12\}$.

Prediction step. $\mu_0 = \mathrm{N}(0, I) \to v_1 \sim \hat{\mu}_1$:

$$\begin{aligned} \mathrm{d}v_t &= \frac{1}{2}\beta(t)v_t\mathrm{d}t + \beta(t)\nabla_v \log p_{T-t}(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t \\ &= f_t(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t \end{aligned}$$

Conditioning.

$$l(v_1) = \frac{1}{2}(Hv_1 - y)^{\mathrm{T}} R^{-1}(Hv_1 - y).$$

Prediction step. $\mu_0 = \mathrm{N}(0, I) \to v_1 \sim \hat{\mu}_1$:

$$\mathrm{d}v_t = \frac{1}{2}\beta(t)v_t\mathrm{d}t + \beta(t)\nabla_v \log p_{T-t}(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t$$
$$= f_t(v_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t$$

Conditioning.

$$l(v_1) = \frac{1}{2}(Hv_1 - y)^{\mathrm{T}} R^{-1}(Hv_1 - y).$$

Controlled diffusion.

$$\mathrm{d}\tilde{v}_t = f_t(\tilde{v}_t)\mathrm{d}t - \beta(t)tH^{\mathrm{T}}R^{-1}(H\tilde{v}_t - y) + g_t(\tilde{v}_t)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t$$

with $2\sigma \to \beta(t)$ and $T = 1$.

Pros:

- no initial approximation of μ_T needed
- no forward-backward iterations required
- avoids high-dimensional optimisation problem
- derivative-free
- applicable to conditional diffusion modelling

Homotopy approach III

Pros:

- no initial approximation of μ_T needed
- no forward-backward iterations required
- avoids high-dimensional optimisation problem
- derivative-free
- applicable to conditional diffusion modelling

Cons:

- numerical approximation of $\tilde{\mu}_t$ and of an elliptic PDE required
- EnKF-type implementation requires that $\tilde{\mu}_t$ remains close to Gaussian
- dynamics of the interacting particle system not understood

THE END