Nonlinear Assimilation via Score-based Sequential Langevin Sampling

Chenguang Duan

Wuhan University

Joint work with Zhao Ding, Yuling Jiao, Jerry Zhijian Yang, Cheng Yuan, Pingwen Zhang

July 18, 2025

Outline

Applications of Data Assimilation

Numerical weather prediction (NWP) uses **physical models** of the atmosphere and oceans to predict the weather based on the **measurement data**.

Physical models

Governing equations of atmosphere:

- conservation of momentum
- conservation of mass
- conservation of energy
- the equation of state for ideal gases
- conservation equation for water mass
- . . .

Measurement data

Problem Formulation

Dynamics model: A discrete time unobserved stochastic process $(\mathbf{X}_k)_{k\geq 1}$ satisfying

$$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k) \qquad \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) = p(\mathbf{X}_{k+1}|\mathbf{X}_k)$$

- $\mathcal{F}_k$ is a known time-dependent model.
- $\mathbf{V}_k$ is a random variable with known distribution.

Problem Formulation

Dynamics model: A discrete time unobserved stochastic process $(\mathbf{X}_k)_{k\geq 1}$ satisfying

$$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k) \qquad \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) = p(\mathbf{X}_{k+1}|\mathbf{X}_k)$$

- $\mathcal{F}_k$ is a known time-dependent model.
- $\mathbf{V}_k$ is a random variable with known distribution.

Measurement model: A discrete time observed stochastic process $(\mathbf{Y}_k)_{k\geq 1}$ satisfying

$$\mathbf{Y}_k = \mathcal{G}_k(\mathbf{X}_k, \mathbf{W}_k) \qquad g_k(\mathbf{Y}_k|\mathbf{X}_k) = p(\mathbf{Y}_k|\mathbf{X}_k)$$

- $\mathcal{G}_k$ is a known time-dependent measurement operator.
- $\mathbf{W}_k$ is a random measurement noise with known distribution.

Problem Formulation

Dynamics model: A discrete time unobserved stochastic process $(\mathbf{X}_k)_{k\geq 1}$ satisfying

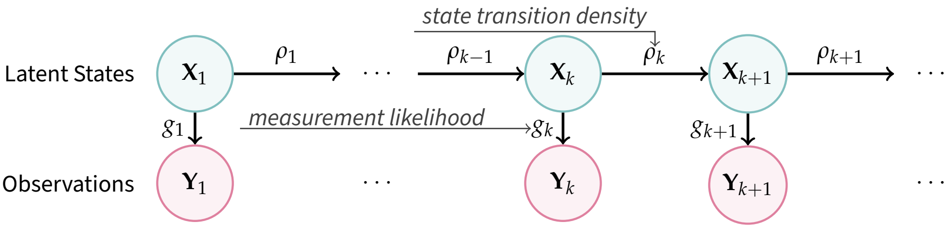
$$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k) \qquad \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) = p(\mathbf{X}_{k+1}|\mathbf{X}_k)$$

- $\mathcal{F}_k$ is a known time-dependent model.
- $\mathbf{V}_k$ is a random variable with known distribution.

Measurement model: A discrete time observed stochastic process $(\mathbf{Y}_k)_{k\geq 1}$ satisfying

$$\mathbf{Y}_k = \mathcal{G}_k(\mathbf{X}_k, \mathbf{W}_k) \qquad g_k(\mathbf{Y}_k|\mathbf{X}_k) = p(\mathbf{Y}_k|\mathbf{X}_k)$$

- $\mathcal{G}_k$ is a known time-dependent measurement operator.
- $\mathbf{W}_k$ is a random measurement noise with known distribution.



Problem Formulation

Dynamics model: A discrete time unobserved stochastic process $(\mathbf{X}_k)_{k\geq 1}$ satisfying

$$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k) \qquad \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) = p(\mathbf{X}_{k+1}|\mathbf{X}_k)$$

- $\mathcal{F}_k$ is a known time-dependent model.
- $\mathbf{V}_k$ is a random variable with known distribution.

Measurement model: A discrete time observed stochastic process $(\mathbf{Y}_k)_{k\geq 1}$ satisfying

$$\mathbf{Y}_k = \mathcal{G}_k(\mathbf{X}_k, \mathbf{W}_k) \qquad g_k(\mathbf{Y}_k|\mathbf{X}_k) = p(\mathbf{Y}_k|\mathbf{X}_k)$$

- $\mathcal{G}_k$ is a known time-dependent measurement operator.
- $\mathbf{W}_k$ is a random measurement noise with known distribution.

Data Assimilation

The goal of the data assimilation is to estimate the posterior distribution of the latent state $\mathbf{X}_{k+1}$ given all available measurements $\mathbf{Y}_{[k+1]}$, that is, $\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})$ for each $k \geq 1$.

Recursive Bayesian Filtering Framework

State-Space Model

Dynamics model	$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k)$	$\longleftrightarrow$	State transition density	$\rho_k(\mathbf{X}_{k+1}\|\mathbf{X}_k)$
Measurement model	$\mathbf{Y}_{k+1} = \mathcal{G}_{k+1}(\mathbf{X}_{k+1}, \mathbf{W}_k)$	$\longleftrightarrow$	Measurement likelihood	$g_{k+1}(\mathbf{Y}_{k+1}\|\mathbf{X}_{k+1})$

$$
\begin{aligned}
&\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]}) \\
&\propto p(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}, \mathbf{Y}_{[k]})q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]}) \\
&= p(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}, \mathbf{Y}_{[k]}) \int p(\mathbf{X}_{k+1}|\mathbf{X}_k, \mathbf{Y}_{[k]})\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_k \\
&= g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}) \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k)\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_k
\end{aligned}
$$

- Bayes' rule
- Chapman-Kolmogorov identity
- $\mathbf{Y}_{k+1} \perp\!\!\!\perp \mathbf{Y}_{[k]}|\mathbf{X}_{k+1}$, and $\mathbf{X}_{k+1} \perp\!\!\!\perp \mathbf{Y}_{[k]}|\mathbf{X}_k$

Recursive Bayesian Filtering Framework

State-Space Model

Dynamics model	$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k)$	$\longleftrightarrow$	State transition density	$\rho_k(\mathbf{X}_{k+1} \mid \mathbf{X}_k)$
Measurement model	$\mathbf{Y}_{k+1} = \mathcal{G}_{k+1}(\mathbf{X}_{k+1}, \mathbf{W}_k)$	$\longleftrightarrow$	Measurement likelihood	$g_{k+1}(\mathbf{Y}_{k+1} \mid \mathbf{X}_{k+1})$

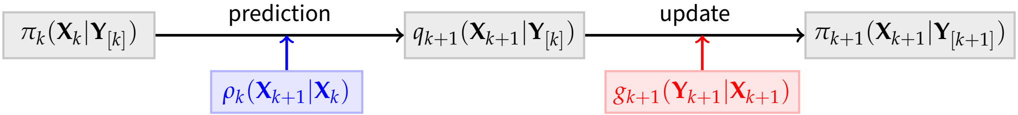
$$\overbrace{\pi_{k+1}(\mathbf{X}_{k+1} \mid \mathbf{Y}_{[k+1]})}^{\text{current posterior}} \propto \underbrace{g_{k+1}(\mathbf{Y}_{k+1} \mid \mathbf{X}_{k+1})}_{\text{measurement likelihood}} \underbrace{\int \overbrace{\rho_k(\mathbf{X}_{k+1} \mid \mathbf{X}_k)}^{\text{state transition}} \overbrace{\pi_k(\mathbf{X}_k \mid \mathbf{Y}_{[k]})}^{\text{previous posterior}} \, \mathrm{d}\mathbf{X}_k}_{\text{current prior}}$$

Recursive Bayesian Filtering Framework

State-Space Model

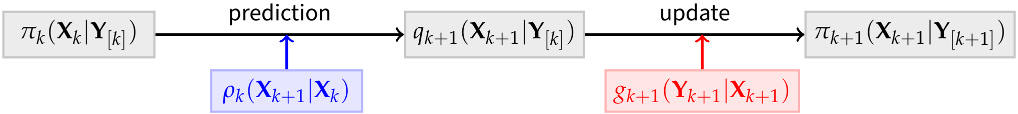
Dynamics model	$\mathbf{X}_{k+1} = \mathcal{F}_k(\mathbf{X}_k, \mathbf{V}_k)$	$\longleftrightarrow$	State transition density	$\rho_k(\mathbf{X}_{k+1}\|\mathbf{X}_k)$
Measurement model	$\mathbf{Y}_{k+1} = \mathcal{G}_{k+1}(\mathbf{X}_{k+1}, \mathbf{W}_k)$	$\longleftrightarrow$	Measurement likelihood	$g_{k+1}(\mathbf{Y}_{k+1}\|\mathbf{X}_{k+1})$

$$\overbrace{\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})}^{\text{current posterior}} \propto \underbrace{g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1})}_{\text{measurement likelihood}} \underbrace{\int \overbrace{\rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k)}^{\text{state transition}} \overbrace{\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})}^{\text{previous posterior}} \, \mathrm{d}\mathbf{X}_k}_{\text{current prior}}$$



Classical Methods for Data Assimilation

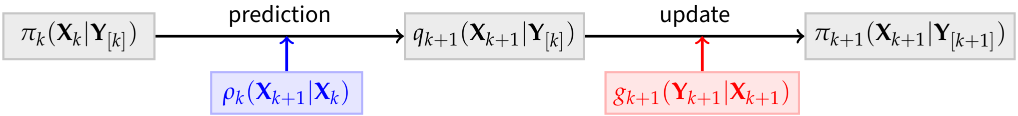
$$\underbrace{\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})}_{\text{current posterior}} \propto g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}) \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) \underbrace{\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})}_{\text{previous posterior}} \, \mathrm{d}\mathbf{X}_k$$



- How to estimate the prediction distribution $q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]})$?
- How to sample from the posterior distribution $\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})$ given the prediction and likelihood?

Classical Methods for Data Assimilation

$$\underbrace{\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})}_{\text{current posterior}} \propto g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}) \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k) \underbrace{\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})}_{\text{previous posterior}} \, \mathrm{d}\mathbf{X}_k$$



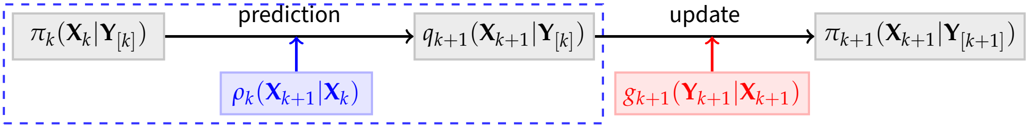
- How to estimate the prediction distribution $q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]})$?
- How to sample from the posterior distribution $\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]})$ given the prediction and likelihood?

	$\pi_k(\mathbf{X}_k\|\mathbf{Y}_{[k]})$	$\rho_k(\mathbf{X}_{k+1}\|\mathbf{X}_k)$	$g_{k+1}(\mathbf{Y}_{k+1}\|\mathbf{X}_{k+1})$	Limitations
Kalman Filter	Gaussian approximation	Gaussian	Gaussian	linear + Gaussian
Particle Filter	particle approximation	–	–	particle degeneracy

Outline

Prediction and Score Matching

Prediction density $$q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]}) = \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k)\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_k$$



How to estimate the score function $\nabla \log q_{k+1}(\cdot|\mathbf{Y}_{[k]})$ *of the prediction distribution?*

$$q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]}) = \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k)\pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_k$$

- Given: $\widehat{\mathbf{X}}_k^1, \dots, \widehat{\mathbf{X}}_k^n \sim^{\text{i.i.d.}} \widehat{\pi}_k(\mathbf{X}_k|\mathbf{Y}_{[k]}) \approx \pi_k(\mathbf{X}_k|\mathbf{Y}_{[k]})$.
- Prediction with **dynamics model**: $\underline{\mathbf{X}}_{k+1}^i = \mathcal{F}_k(\widehat{\mathbf{X}}_k^i, \mathbf{V}_k^i)$, for each $1 \leq i \leq n$.

$$\underline{\mathbf{X}}_{k+1}^i \sim \widehat{q}_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]}) := \int \rho_k(\mathbf{X}_{k+1}|\mathbf{X}_k)\widehat{\pi}_k(\mathbf{X}_k|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_k \approx q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]}).$$

- Denoising **score matching**:

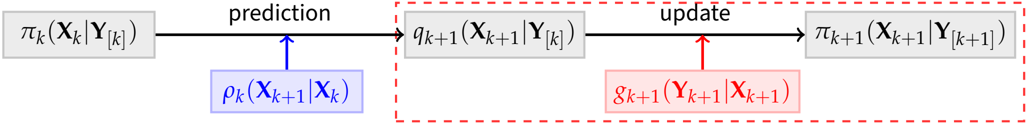
$$\widehat{\mathbf{s}}_{k+1}(\cdot|\mathbf{Y}_{[k]}) = \operatorname*{arg\,min}_{\mathbf{s}:\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}^d} L_{k+1}(\mathbf{s}) = \mathbb{E}_{\underline{\mathbf{X}}_{k+1}\sim\widehat{q}_{k+1}(\cdot|\mathbf{Y}_{[k]})}\mathbb{E}_{\boldsymbol{\varepsilon}\sim N(0,\mathbf{I}_d)}\left[\|\sigma\mathbf{s}(\underline{\mathbf{X}}_{k+1}+\sigma\boldsymbol{\varepsilon},\mathbf{Y}_{[k]})+\boldsymbol{\varepsilon}\|_2^2\right]$$

which is the score function of the Gaussian smoothed prediction distribution

$$\widehat{\mathbf{s}}_{k+1}(\cdot|\mathbf{Y}_{[k]}) := \nabla\log\left(\int N(\cdot;\mathbf{X}_{k+1},\sigma^2\mathbf{I}_d)\widehat{q}_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]})\mathrm{d}\mathbf{X}_{k+1}\right) \approx \nabla\log\widehat{q}_{k+1}(\cdot|\mathbf{Y}_{[k]})$$

Update via Langevin Sampling

Bayes' rule $$\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]}) \propto g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1}) q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]})$$



How to sample from the posterior distribution $\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})$ *given the prediction and likelihood?*

$$\pi_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k+1]}) \propto g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{X}_{k+1})q_{k+1}(\mathbf{X}_{k+1}|\mathbf{Y}_{[k]})$$

- Langevin dynamics

$$\begin{aligned}\mathrm{d}\mathbf{Z}_t &= \nabla \log \pi_{k+1}(\mathbf{Z}_t|\mathbf{Y}_{[k+1]})\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{B}_t, \\ &= \{\nabla \log g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{Z}_t) + \nabla \log q_{k+1}(\mathbf{Z}_t|\mathbf{Y}_{[k]})\}\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{B}_t, \quad \mathbf{Z}_0 \sim \widehat{q}_{k+1}(\cdot|\mathbf{Y}_{[k]}).\end{aligned}$$

  - Prediction score estimation
  - Euler-Maruyama discretization
- Score-based Langevin Monte Carlo

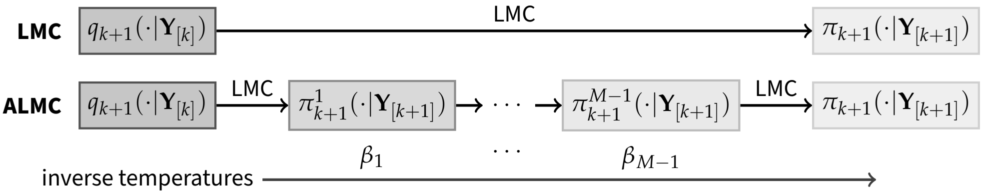
$$\widehat{\mathbf{Z}}_{(k+1)h} = \widehat{\mathbf{Z}}_{kh} + h\{\nabla \log g_{k+1}(\mathbf{Y}_{k+1}|\widehat{\mathbf{Z}}_{kh}) + \widehat{\mathbf{s}}_{k+1}(\widehat{\mathbf{Z}}_{kh}, \mathbf{Y}_{[k]})\} + \sqrt{2h}\boldsymbol{\xi}_k, \quad \widehat{\mathbf{Z}}_0 \sim \widehat{q}_{k+1}(\cdot|\mathbf{Y}_{[k]}).$$

Annealing Strategy

Interpolations between the prediction and posterior distribution:

$$\pi_{k+1}^{m}(\cdot|\mathbf{Y}_{[k+1]}) \propto \pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})^{\beta_m} q_{k+1}(\cdot|\mathbf{Y}_{[k]})^{1-\beta_m}, \quad 0 \leq m \leq M$$

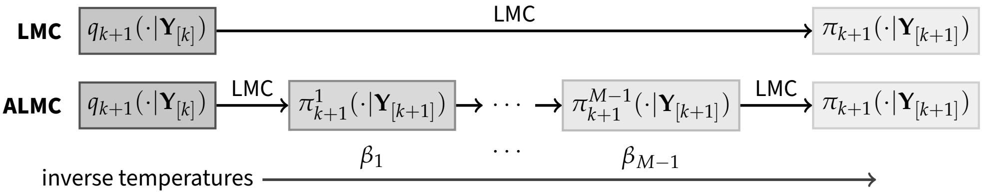
- Annealing schedule: $0 \equiv \beta_0 < \beta_1 < \cdots < \beta_M \equiv 1$.



Interpolations between the prediction and posterior distribution:

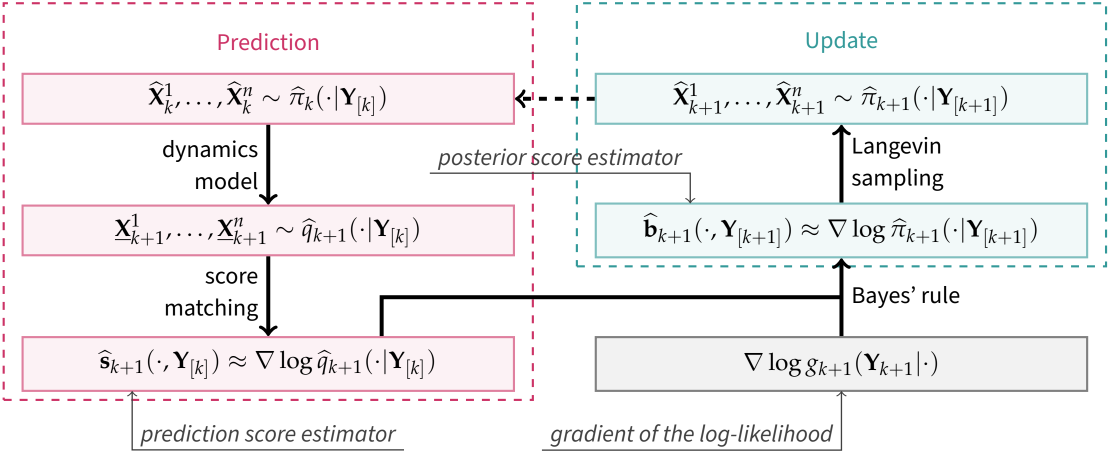
$$\pi_{k+1}^{m}(\cdot|\mathbf{Y}_{[k+1]}) \propto \pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})^{\beta_m} q_{k+1}(\cdot|\mathbf{Y}_{[k]})^{1-\beta_m}, \quad 0 \le m \le M$$

- Annealing schedule: $0 \equiv \beta_0 < \beta_1 < \cdots < \beta_M \equiv 1$.



$$\begin{aligned}\nabla \log \pi_{k+1}^{m}(\cdot|\mathbf{Y}_{[k+1]}) &= \beta_m \nabla \log \pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]}) + (1-\beta_m)\nabla \log q_{k+1}(\cdot|\mathbf{Y}_{[k]}) \\ &= \beta_m \nabla \log g_{k+1}(\mathbf{Y}_{k+1}|\cdot) + \nabla \log q_{k+1}(\cdot|\mathbf{Y}_{[k]})\end{aligned}$$

Summary of Procedure



Algorithm 1: Score-based **S**equential **L**angevin **S**ampling for data assimilation.

Input: The observations $(\mathbf{Y}_k)_{k\in\mathbb{N}}$, the dynamics model $(\mathcal{F}_k)_{k\in\mathbb{N}}$, the likelihood $\{g_k(\mathbf{Y}_k|\cdot)\}_{k\in\mathbb{N}}$.

Output: A particle approximation $\widehat{\mathbf{X}}_{k+1}^1, \ldots, \widehat{\mathbf{X}}_{k+1}^n$ to the distribution $\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})$.

```
# Initial posterior sampling.
```
Sample from the gauss of the initial prior distribution $\underline{\mathbf{X}}_1^1, \ldots, \underline{\mathbf{X}}_1^n \sim^{\text{i.i.d.}} \widehat{q}_1$.

Estimate the score from $\{\underline{\mathbf{X}}_1^i\}_{i=1}^n$ by score matching $\widehat{\mathbf{s}}_1$.

Sample from the posterior $\widehat{\pi}_1(\cdot|\mathbf{y}_1)$ by Langevin sampling: $\widehat{\mathbf{X}}_1^1, \ldots, \widehat{\mathbf{X}}_1^n \leftarrow \mathrm{ALMC}(\underline{\mathbf{X}}_1^1, \ldots, \underline{\mathbf{X}}_1^n, \widehat{\mathbf{s}}_1, g_1(\mathbf{Y}_1|\cdot))$.

```
# Recursive posterior sampling.
```
for $\underline{k = 1, 2, \ldots}$ **do**

```
# Prediction step.
```
Run the dynamics model: $\underline{\mathbf{X}}_{k+1}^i \leftarrow \mathcal{F}_k(\widehat{\mathbf{X}}_k^i, \mathbf{V}_k^i)$ with $\mathbf{V}_k^i \sim p_{\mathbf{V}}$ for $1 \le i \le n$.

Estimate the prediction score from $\{\underline{\mathbf{X}}_{k+1}^i\}_{i=1}^n$ by score matching $\widehat{\mathbf{s}}_{k+1}(\cdot, \mathbf{Y}_{[k]})$.

```
# Update step.
```
Sample from the posterior $\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})$ by Langevin sampling:

$\widehat{\mathbf{X}}_{k+1}^1, \ldots, \widehat{\mathbf{X}}_{k+1}^n \leftarrow \mathrm{ALMC}(\underline{\mathbf{X}}_{k+1}^1, \ldots, \underline{\mathbf{X}}_{k+1}^n, \widehat{\mathbf{s}}_{k+1}(\cdot, \mathbf{Y}_{[k]}), g_{k+1}(\mathbf{Y}_{k+1}|\cdot))$.

end

return $\underline{\widehat{\mathbf{X}}_{k+1}^1, \ldots, \widehat{\mathbf{X}}_{k+1}^n}$

Outline

Assumptions

- **A1. Lipschitz score.** For each $k \in \mathbb{N}$, the posterior score is λ-Lipschitz on $\mathbb{R}^d$, that is, for each $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$,

$$\|\nabla \log \pi_{k+1}(\mathbf{x}_1|\mathbf{Y}_{[k+1]}) - \nabla_{\mathbf{x}} \log \pi_{k+1}(\mathbf{x}_2|\mathbf{Y}_{[k+1]})\|_2 \leq \lambda \|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$

- **A2. Log-Sobolev inequality.** For each $k \in \mathbb{N}$, the posterior distribution satisfies a log-Sobolev inequality with constant C_{LSI}, that is, for each function $f \in C_0^\infty(\mathbb{R}^d)$,

$$\mathrm{Ent}(f^2) \leq 2C_{\mathrm{LSI}} \mathbb{E}\left[\|\nabla f\|_2^2\right],$$

  where the entropy is defined as $\mathrm{Ent}(g) = \mathbb{E}[g \log g] - \mathbb{E}[g] \log \mathbb{E}[g]$, and the expectation is taken with respect to the posterior distribution $\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})$.

- **A3. Boundedness and condition number.** For each $k \in \mathbb{N}$,
  - there exists $B \geq 1$ such that $\rho_k(\mathbf{x}|\mathbf{X}_k), \|\nabla \rho_k(\mathbf{x}|\mathbf{X}_k)\|_\infty \leq B$ for $\mathbf{x} \in \mathbb{R}^d$,
  - there exists $D \geq 1$ such that $q_{k+1}(\mathbf{x}|\mathbf{Y}_{[k]}) \geq D^{-1}$ for $\mathbf{x} \in \Omega_{k+1} := \mathrm{supp}(q_{k+1}(\cdot|\mathbf{Y}_{[k]}))$,
  - $q_1(\mathbf{x}), \|\nabla q_1(\mathbf{x})\|_\infty \leq B$ and $q_1(\mathbf{x}) \geq D^{-1}$ for $\mathbf{x} \in \Omega_1 := \mathrm{supp}(q_1)$, and
  - there exists $\kappa > 0$ such that

$$\frac{\sup_{\mathbf{x}} g_1(\mathbf{Y}_1|\mathbf{x})}{\int g_1(\mathbf{Y}_1|\mathbf{x}) q_1(\mathbf{x}) \mathrm{d}\mathbf{x}}, \frac{\sup_{\mathbf{x}} g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{x})}{\int g_{k+1}(\mathbf{Y}_{k+1}|\mathbf{x}) q_{k+1}(\mathbf{x}|\mathbf{Y}_{[k]}) \mathrm{d}\mathbf{x}} \leq \kappa.$$

- **A4. Error of score matching.** There exists a score matching tolerance $\Delta \in (0, 1)$ such that

$$\mathbb{E}_{\underline{\mathbf{X}}_1}\left[\|\nabla_{\mathbf{x}} \log \widehat{q}_1(\underline{\mathbf{X}}_1) - \widehat{\mathbf{s}}_1(\underline{\mathbf{X}}_1)\|_2^2\right] \leq \Delta^2,$$

$$\mathbb{E}_{\underline{\mathbf{X}}_{k+1}}\left[\|\nabla_{\mathbf{x}} \log \widehat{q}_{k+1}(\underline{\mathbf{X}}_{k+1}|\mathbf{Y}_{[k]}) - \widehat{\mathbf{s}}_{k+1}(\underline{\mathbf{X}}_{k+1}, \mathbf{Y}_{[k]})\|_2^2\right] \leq \Delta^2,$$

for each $k \in \mathbb{N}$. Here the expectation $\mathbb{E}_{\underline{\mathbf{X}}_1}[\cdot]$ is taken with respect to $\underline{\mathbf{X}}_1 \sim \widehat{q}_1$, and the expectation $\mathbb{E}_{\underline{\mathbf{X}}_{k+1}}[\cdot]$ is taken with respect to $\underline{\mathbf{X}}_{k+1} \sim \widehat{q}_{k+1}(\cdot|\mathbf{Y}_{[k]})$.

Convergence Analysis for Posterior Sampling

Under Assumptions **A1** to **A4**. Then for each $k \in \mathbb{N}$ and each terminal time $T = Kh$,

$$\underbrace{\|\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]}) - \widehat{\pi}_{k+1}(\cdot|\mathbf{Y}_{[k+1]})\|_{\mathrm{TV}}^2}_{\text{posterior error}} \lesssim \underbrace{\exp\left(-\frac{T}{5C_{\mathrm{LSI}}}\right)\eta_\chi^2}_{\text{convergence of LD}} + \underbrace{dC_{\mathrm{LSI}}\lambda^2 h}_{\text{discretization error}} + \underbrace{\sqrt{\kappa B^3 D^3}(T + C_{\mathrm{LSI}}\eta_\chi)\Delta}_{\text{score estimation error}}$$
$$+ B^4 D^4 T \underbrace{\|\pi_k(\cdot|\mathbf{Y}_{[k]}) - \widehat{\pi}_k(\cdot|\mathbf{Y}_{[k]})\|_{\mathrm{TV}}^2}_{\text{prior error}},$$

where the step size h and the initial distribution $\pi_k^0(\cdot|\mathbf{Y}_{[k]})$ satisfies

$$h \lesssim \frac{1}{dC_{\mathrm{LSI}}\lambda^2}, \quad \chi^2(\pi_{k+1}^0(\cdot|\mathbf{Y}_{[k+1]})\|\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})) \leq \eta_\chi^2.$$

Convergence Analysis for Posterior Sampling

Under Assumptions **A1** to **A4**. Then for each $k \in \mathbb{N}$ and each terminal time $T = Kh$,

$$\underbrace{\|\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]}) - \widehat{\pi}_{k+1}(\cdot|\mathbf{Y}_{[k+1]})\|_{\mathrm{TV}}^2}_{\text{posterior error}} \lesssim \underbrace{\exp\left(-\frac{T}{5C_{\mathrm{LSI}}}\right)\eta_\chi^2}_{\text{convergence of LD}} + \underbrace{dC_{\mathrm{LSI}}\lambda^2 h}_{\text{discretization error}} + \underbrace{\sqrt{\kappa B^3 D^3}(T + C_{\mathrm{LSI}}\eta_\chi)\Delta}_{\text{score estimation error}}$$

$$+ B^4 D^4 T \underbrace{\|\pi_k(\cdot|\mathbf{Y}_{[k]}) - \widehat{\pi}_k(\cdot|\mathbf{Y}_{[k]})\|_{\mathrm{TV}}^2}_{\text{prior error}},$$

where the step size h and the initial distribution $\pi_k^0(\cdot|\mathbf{Y}_{[k]})$ satisfies

$$h \lesssim \frac{1}{dC_{\mathrm{LSI}}\lambda^2}, \quad \chi^2(\pi_{k+1}^0(\cdot|\mathbf{Y}_{[k+1]})\|\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]})) \leq \eta_\chi^2.$$

- Early-stopping: Trade-off between the convergence of Langevin dynamics and score estimation error.
- Condition number: How the score matching error Δ effects the posterior error.

Convergence Analysis for Assimilation

Under Assumptions **A1** to **A4**. Suppose that the initial prior error satisfy

$$\varepsilon_{\text{init}} := \|\nabla_{\mathbf{x}} \log q_1 - \nabla_{\mathbf{x}} \log \widehat{q}_1\|_{L^\infty(\mathbb{R}^d)}.$$

Then for each time step $k \in \mathbb{N}$ and error tolerance $\varepsilon \in (0,1)$,

$$\|\pi_{k+1}(\cdot|\mathbf{Y}_{[k+1]}) - \widehat{\pi}_{k+1}(\cdot|\mathbf{Y}_{[k+1]})\|_{\text{TV}} \leq \widetilde{\mathcal{O}}(\varepsilon_{\text{init}} + \varepsilon),$$

where the $\widetilde{\mathcal{O}}$ notation omits logarithmic factors of $\varepsilon_{\text{init}}$ and ε, and the constant behind the $\widetilde{\mathcal{O}}$ notation is independent of $\varepsilon_{\text{init}}$ and ε. Moreover, the step size h, the number of the Langevin iterations K and the score matching error Δ satisfy:

$$h = \Theta\Big(\frac{\varepsilon^2}{dC_{\text{LSI}}\lambda^2}\Big), \quad K = \Theta\Big(\frac{d\lambda^2 C_{\text{LSI}}^2}{\varepsilon^2} \log\Big(\frac{\eta_\chi^2}{\varepsilon^2}\Big)\Big),$$

$$\Delta = \Theta\Big(\frac{\varepsilon^2}{\sqrt{\kappa B^3 D^3} C_{\text{LSI}}} \frac{1}{\log(\varepsilon^{-2}\eta_\chi^2) + \eta_\chi}\Big).$$

Outline

Double-Well Potential

Dynamics model:

$$dX_t = -\nabla U(X_t)dt + \beta dB_t$$

- Double-well potential: $U(x) = x^4 - 2x^2$.

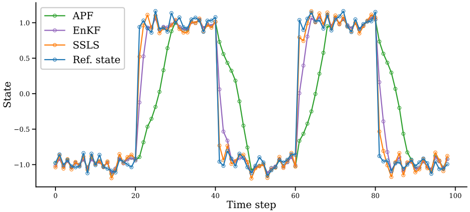
Measurement model:

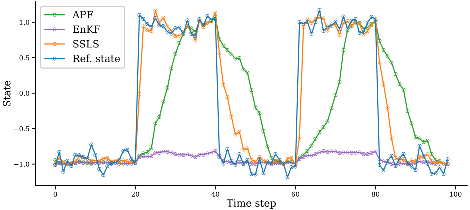
- Linear measurement model:

$$Y_k = X_k + \sigma W_k$$

- Nonlinear measurement model:

$$Y_k = \exp(X_k - \gamma_k) + \sigma W_k$$





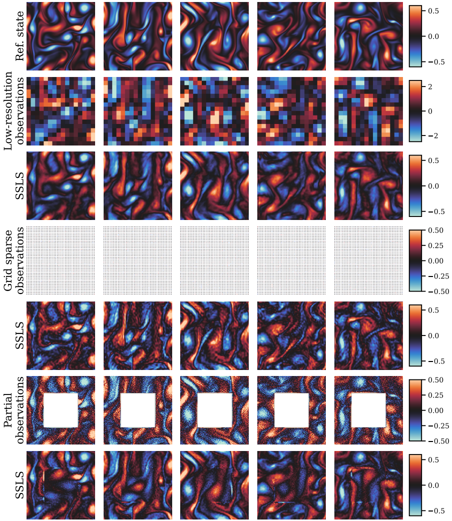
Kolmogorov flow

Dynamics model:

$$\begin{cases} \partial_t \mathbf{u} = -(\mathbf{u} \cdot \nabla)\mathbf{u} + \frac{1}{\mathrm{Re}} \nabla^2 \mathbf{u} - \frac{1}{\rho} \nabla p + \mathbf{F}, \\ 0 = \nabla \cdot \mathbf{u}. \end{cases}$$

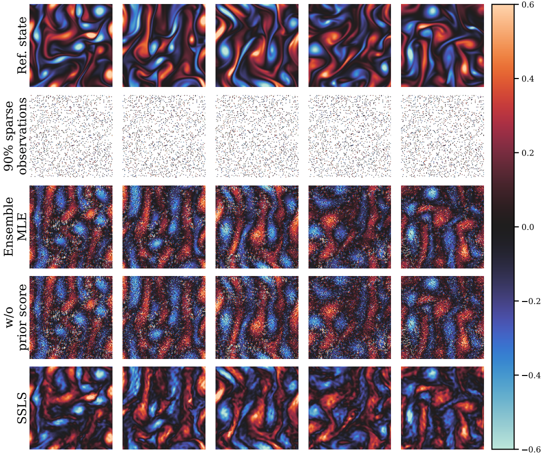
Measurement model:

- Low-resolution observation
- Sparse observation
- Partial observation

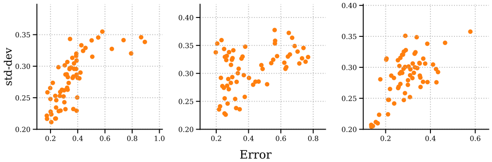


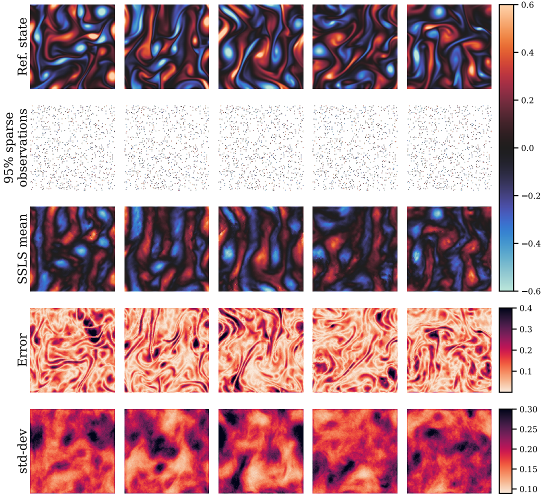
Ablation study: influence of the prediction score

- Score-based Langevin sampling
- Langevin sampling without prediction score
- Ensemble maximum likelihood estimation



Uncertainty quantification





Outline

Concluding Remarks

- Improvement of the computational efficiency.
- Provable non-log-concave posterior sampling.
- Refined error bounds for long-time assimilation.
- Applications in the real data.

Reference: Zhao Ding, Chenguang Duan, Yuling Jiao, Jerry Zhijian Yang, Cheng Yuan, and Pingwen Zhang. Nonlinear Assimilation via Score-based Sequential Langevin Sampling. arXiv:2411.13443.

Thanks for your attention!

Homepage:
https://chenguangduan.github.io/
Google Scholar:
https://scholar.google.com/citations?user=RpmGgyMAAAAJ
Email: cgduan.math@gmail.com