On the discretization of the Onsager-Machlup functional

Nozomi Sugiura

JAMSTEC

01Mar2017
Main claim

- For nonlinear smoothing problem:

\[ dx = f(x) \, dt + \sigma \, dW \ \xrightarrow{\text{discretize}} \]

\[ x_n = x_{n-1} + f(x_{n-1}) \delta_t + \sigma \sqrt{\delta_t} \xi_{n-1}, \quad \xi_{n-1} \sim \mathcal{N}(0, I), \quad (1) \]

\[ x_0 \sim \mathcal{N}(x_b, \sigma_b^2 I), \quad y_n|x_n \sim \mathcal{N}(x_n, \sigma_o^2 I), \quad (2) \]

- A proper setting of the cost function is \(^a\)

\[ J = \frac{1}{2\sigma_b^2} (x_0 - x_b)^2 + \sum_n \frac{1}{2\sigma_o^2} (x_n - y_n)^2 \]

\[ + \delta_t \sum_n \left[ \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1}) \right)^2 + \frac{1}{2} \text{div} f(x_{n-1}) \right]. \quad (3) \]

- But, the divergence term should be excluded in MCMC.

\(^a\) \(x \in \mathbb{R}^D\): state variable, \(W\): \(D\)-dim Wiener process, \(t\): time, \(n\): time index, \(\delta_t\): time increment, \(x_b\): background value, \(\sigma_b\): background std., \(y\): observational data, \(\sigma_o\): observational std., \(\sigma\): model error std., \([\sigma] = [x] T^{-1/2}\).
Two formulas for prior

To the system: \( x_n = x_{n-1} + f(x_{n-1}) \delta t + \sigma \sqrt{\delta t} \xi_{n-1} \), we can assign one of the following prior probabilities:

1. The probability \( \pi_{\text{path}} \) for a rough path\(^b\) \( \{ \omega \, | \, x_n(\omega) \}_{0 \leq n \leq N} \) in \( \text{Itô} \) form:

\[
\pi_{\text{path}} \propto \exp \left\{ -\delta t \sum_n \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1}) \right)^2 \right\}.
\]

(e.g., Zinn-Justin, 2002)

\( \implies \) The choice of stochastic integral (\( \text{Itô} \) or Stratonovich) matters.

2. The probability \( \pi_{\text{tube}} \) for a smooth tube\(^c\) that represents its neighboring paths \( \Omega = \{ \omega \, | \, (\forall n) \, | \, x_n - x_n(\omega) | < \epsilon \} \):

\[
\pi_{\text{tube}} \propto \exp \left\{ -\delta t \sum_n \left[ \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1/2}) \right)^2 + \frac{1}{2} \text{div} f(x_{n-1/2}) \right] \right\}.
\]

(e.g., Ikeda and Watanabe, 1981)

\(^b\)For a rough (zigzag) path, \( x_n - x_{n-1} = O(\delta t^{1/2}) \).

\(^c\)For a smooth tube, \( x_n - x_{n-1} = O(\delta t) \).
Two algorithms for DA

Using the unnormalized posterior probability \( P_\bullet = \pi_\bullet \times \text{(likelihood)} = e^{-J} \), we can apply one of the following data assimilation algorithms:

1. **MCMC** (Cotter et al., 2013) samples the paths \( X^{(k)} = \{x_n(\omega_k)\}_{0 \leq n \leq N} \) according to the distribution \( P_{\text{path}} \) by iterating \((\alpha > 0)\):

\[
X^{(k+1)} = X^{(k)} + \alpha \nabla \ln P_{\text{path}} + \sqrt{2\alpha} \xi, \quad \xi \sim \mathcal{N}(0, 1)^{D(N+1)},
\]

with the Metropolis rejection step for adjustment.

\[ \implies \text{We get an ensemble of sample paths according to the posterior probability.} \]

2. **4D-Var** seeks the most probable tube \( X = \{x_n\}_{0 \leq n \leq N} \) by iterating \((\alpha > 0)\):

\[
X^{(k+1)} = X^{(k)} + \alpha \nabla \ln P_{\text{tube}}.
\]

\[ \implies \text{We get the maximum a posteriori (MAP) estimate for the tubes.} \]
Four schemes for OM

The prior probability has a form $\pi \propto \exp \left( -\delta_t \sum_n \widetilde{OM} \right)$, where $\widetilde{OM}$ is called the Onsager-Machlup functional (Onsager and Machlup, 1953). We are going to test four discretization schemes:

1. **E [Euler]** (e.g., Zinn-Justin, 2002):
   $$\widetilde{OM}_E \equiv \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1}) \right)^2,$$

2. **ED [Euler with div]**:
   $$\widetilde{OM}_{ED} \equiv \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1}) \right)^2 + \frac{1}{2} \text{div} f(x_{n-1}),$$

3. **T [Trapezoidal]**:
   $$\widetilde{OM}_T \equiv \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1/2}) \right)^2,$$

4. **TD [Trapezoidal with div]** (e.g., Ikeda and Watanabe, 1981; Apte et al., 2007):
   $$\widetilde{OM}_{TD} \equiv \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta t} - f(x_{n-1/2}) \right)^2 + \frac{1}{2} \text{div} f(x_{n-1/2}),$$

where, $f(x_{n-1/2}) \simeq (f(x_n) + f(x_{n-1}))/2$. 
We can also solve it by particle smoother:

1. Generate samples of initial and model errors, and integrate $M$ copies of model using them to obtain a Monte-Carlo approximation of the prior distribution:

$$P(X) \simeq \frac{1}{M} \sum_{m=1}^{M} \prod_{n=0}^{N} \delta(X_n - x_n^{(m)}).$$

2. Reweight it according to the Bayes’ theorem:\

$$P(Y|X) \propto \exp \left( -\frac{1}{2\sigma_y^2} (y - x_N)^2 \right),$$  \hspace{1cm} (6)

$$P(X|Y) = \frac{P(X)P(Y|X)}{\int P(X)P(Y|X)dX}$$  \hspace{1cm} (7)

$$= \sum_{m=1}^{M} \prod_{n=0}^{N} \delta(X_n - x_n^{(m)}) \frac{w^{(m)}}{\sum_{m=1}^{M} w^{(m)}},$$  \hspace{1cm} (8)

$$w^{(m)} \equiv \exp \left( -\frac{1}{2\sigma_y^2} (y - x_N^{(m)})^2 \right).$$  \hspace{1cm} (9)

This method does not involve the computation of prior probability $\pi$.\d\Suppose obs. is only at $n = N$.\d
Find the probability distribution of paths:

\[ dx = \tanh(x)dt + dW, \quad x(t = 0) \sim \mathcal{N}(0, 0.16), \quad (10) \]

subject to an observation \( y \):

\[ y|x(t = 5) \sim \mathcal{N}(x(t = 5), 0.16), \quad y = 1.5. \quad (11) \]

In this case, \( \text{div } f(x) = \frac{1}{\cosh^2(x)} \) imposes a penalty for small \( x \).

<table>
<thead>
<tr>
<th>Solver</th>
<th>OM Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle Smoother</td>
<td>(no need)</td>
</tr>
<tr>
<td>MCMC</td>
<td>E, ED, T, TD</td>
</tr>
<tr>
<td>4D-Var</td>
<td>E, ED, T, TD</td>
</tr>
</tbody>
</table>
MCMC Results vs PS (Hyperbolic model)

Probability density of paths normalized on each time slice.

Reference solution by Particle Smoother

MCMC with Scheme E or TD

MCMC with Scheme ED

MCMC with Scheme T
Expected path derived by MCMC (Hyperbolic model)

Mean orbits (MCMC)

PS: Reference solution by Particle Smoother, E: Euler without div term, ED: Euler with div term, T: Trapezoidal without div term, TD: Trapezoidal with div term.

→ When you solve it by MCMC, schemes E and TD offer proper expected path.
PS: Reference solution by Particle Smoother, E: Euler without div term, ED: Euler with div term, T: Trapezoidal without div term, TD: Trapezoidal with div term.

→ When you solve it by 4D-Var, schemes with div term (ED, TD) offer proper MAP estimate.
Example B (Rössler model)

Nonlinear smoothing problem for Rössler model

Find the probability distribution of paths:

\[
\begin{align*}
    dx_1 &= (-x_2 - x_3)dt + \sigma dW_1, \\
    dx_2 &= (x_1 + ax_2)dt + \sigma dW_2, \\
    dx_3 &= (b + x_1x_3 - cx_3)dt + \sigma dW_3,
\end{align*}
\]  

(12)

\[x(t = 0) \sim \mathcal{N}(x_b, 0.04I),\]  

(13)

subject to an observation \(y\):

\[y|x(t = 0.4) \sim \mathcal{N}(x(t = 0.4), 0.04I).\]  

(14)

where, \((a, b, c) = (0.2, 0.2, 6), \ \sigma = 2, \ x_b = (2.0659834, -0.2977757, 2.0526298)^T, \\
y = (2.5597086, 0.5412736, 0.6110939)^T.\]

In this case, \(\text{div } f(x) = x_1 + a - c\) imposes a penalty for large \(x_1\).
PS: Reference solution by Particle Smoother, \( E \): Euler without div term, \( ED \): Euler with div term, \( T \): Trapezoidal without div term, \( TD \): Trapezoidal with div term.

\( \rightarrow \) When you solve it by MCMC, schemes \( E \) and \( TD \) offer proper expected path.
Most probable tube derived by 4D-Var (Rössler model)

PS: Reference solution by Particle Smoother, \( E \): Euler without div term, \( ED \): Euler with div term, \( T \): Trapezoidal without div term, \( TD \): Trapezoidal with div term.

→ When you solve it by 4D-Var, schemes with div term (ED, TD) offer proper MAP estimate.

\[ \chi(t) = \arg \max_{\mathbf{X}(t)} P[\Omega_{\mathbf{X}(t)} | Y], \] 

where \( \Omega \) represents the tube centered at \( \mathbf{X}(t) \) with radius 0.03.
We examined the discretization schemes of the Onsager-Machlup function, \( \frac{1}{2\sigma^2} \left( \frac{dx}{dt} - f(x) \right)^2 + \frac{1}{2} \text{div}(f) \), using hyperbolic model and Rössler model.

Consistent with literature, the following discretization schemes turn out to be correct (✓):

### For sampling by MCMC

<table>
<thead>
<tr>
<th></th>
<th>with div ( f )</th>
<th>without div ( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler scheme</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Trapezoidal scheme</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

### For MAP estimate by 4D-Var

<table>
<thead>
<tr>
<th></th>
<th>with div ( f )</th>
<th>without div ( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler scheme</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Trapezoidal scheme</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
For nonlinear smoothing problem:

\[ dx = f(x) dt + \sigma dW \]

\[ x_n = x_{n-1} + f(x_{n-1}) \delta_t + \sigma \sqrt{\delta_t} \xi_{n-1}, \quad \xi_{n-1} \sim \mathcal{N}(0, I), \]

\[ x_0 \sim \mathcal{N}(x_b, \sigma_b^2 I), \quad y_n|x_n \sim \mathcal{N}(x_n, \sigma_o^2 I), \]

A proper setting of the cost function is \( f \)

\[ J = \frac{1}{2\sigma_b^2} (x_0 - x_b)^2 + \sum_n \frac{1}{2\sigma_o^2} (x_n - y_n)^2 \]

\[ + \delta_t \sum_n \left[ \frac{1}{2\sigma^2} \left( \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1}) \right)^2 + \frac{1}{2} \text{div} f(x_{n-1}) \right]. \]

But, the divergence term should be excluded in MCMC.

\( f \): time index, \( \delta_t \): time increment, \( x_b \): background value, \( \sigma_b \): background std.,
\( y \): observational data, \( \sigma_o \): observational std., \( \sigma \): model error std. \( \forall [\sigma] = [x] T^{-1/2} \).


Scaling of the terms

Taylor expansion of $f(x_{n-1})$ term around $x_{n-1/2}$:

$$OM \simeq \sum \delta_t \left\{ \sigma^{-2} \left[ \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1/2}) - (x_n - x_{n-1}) \frac{\partial f}{\partial x}(x_{n-1/2}) \right]^2 + \text{div}(f) \right\}$$

$$= \delta_t \left\{ \sigma^{-2} (\text{noise} + \text{shift})^2 + \text{divergence} \right\}.$$

noise $\equiv \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1/2})$, shift $\equiv (x_n - x_{n-1}) \frac{\partial f}{\partial x}(x_{n-1/2})$, divergence $\equiv \text{div}(f)$.

We assume order-one fluctuations: $\sigma = O(1)$.

- For a sample path of stochastic process, the scaling $x_n - x_{n-1} = O(\delta_t^{1/2})$ leads to

$$OM = \sum \delta_t \left\{ \sigma^{-2} \left( \frac{\text{noise}^2}{\delta_t^{-1}} + \text{noise} \times \text{shift} + \frac{\text{shift}^2}{\delta_t} \right) + \text{divergence} \right\}. \tag{15}$$

- The shift term induces a Jacobian that coincides with the divergence term in TD scheme (Zinn-Justin, 2002).
Scaling of the terms

Taylor expansion of $f(x_{n-1})$ term around $x_{n-1/2}$:

$$OM \simeq \sum_n \delta_t \left\{ \sigma^{-2} \left[ \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1/2}) - (x_n - x_{n-1}) \frac{\partial f}{\partial x}(x_{n-1/2}) \right]^2 + \text{div} (f) \right\}$$

$$= \delta_t \left\{ \sigma^{-2} (\text{noise} + \text{shift})^2 + \text{divergence} \right\}.$$  

\text{noise} \equiv \frac{x_n - x_{n-1}}{\delta_t} - f(x_{n-1/2}), \quad \text{shift} \equiv (x_n - x_{n-1}) \frac{\partial f}{\partial x}(x_{n-1/2}), \quad \text{divergence} \equiv \text{div} (f).$

We assume order-one fluctuations: $\sigma = O(1)$.

- For a smooth tube, the scaling $x_n - x_{n-1} = O(\delta_t)$ leads to

$$OM = \sum \delta_t \left\{ \sigma^{-2} \left( \underbrace{\text{noise}^2}_{1} + \underbrace{\text{noise} \times \text{shift}}_{\delta_t} + \underbrace{\text{shift}^2}_{\delta_t^2} \right) + \underbrace{\text{divergence}}_{1} \right\}. \quad (15)$$

- The \textbf{shift} term is negligible, but the \textbf{divergence} term is not.
gray: path trajectories, **MAP**: the most probable tube (estimated by 4D-Var)

The particle smoother is performed with $3 \times 10^{12}$ particles.
gray: path trajectories, **MAP**: the most probable tube (estimated by 4D-Var)

The particle smoother is performed with $5.1 \times 10^{10}$ particles.
Invariant distribution in MCMC

The solution $x$ of Stochastic differential equation for state variable $X$:

$$dX_s = b(X_s)ds + \sigma(X_s)dW_s$$

has a probability distribution $u(x, s)$ that evolves according to:

$$\frac{\partial u}{\partial s} = -\sum_i \frac{\partial}{\partial x_i} (b_i u) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (V_{ij} u).$$

(17)

where, $V = \sigma \sigma^T$. In particular, if we take a probability distribution $\pi(x)$ and set $b_i = \frac{1}{2} \frac{\partial}{\partial x_i} \log \pi$, $V = I$, this reads

$$\frac{\partial u}{\partial s} = -\frac{1}{2} \sum_i \frac{\partial}{\partial x_i} \left( \frac{\partial \log \pi}{\partial x_i} u \right) + \frac{1}{2} \sum_i \frac{\partial^2 u}{\partial x_i^2} = -\frac{1}{2} \sum_i \frac{\partial}{\partial x_i} \left( \frac{\partial \pi}{\partial x_i} u \right) + \frac{1}{2} \sum_i \frac{\partial^2 u}{\partial x_i^2},$$

(18)

Once, the distribution $u(x, s_0) = \pi(x)$ is attained, it remains as it is $u(x, s) = \pi(x)$. We call it as “Stochastic process $dX_s = \left( \frac{1}{2} \nabla \log \pi \right) ds + dW_s$ has the invariant distribution $\pi$,”

This suggests the discretized process:

$$X^{(k+1)} = X^{(k)} + \alpha \nabla \log \pi + \sqrt{2\alpha} \xi, \quad \xi \sim N(0, 1)$$

(19)

generates a good candidate for sample paths, where $\nabla \log \pi = -\nabla J$ is the adjoint of the cost function.