Continuous-discrete smoothing of diffusions

Applications to stochastic landmark deformation models

Frank van der Meulen
Delft Institute of Applied Mathematics
The Netherlands
AI seminar series University of Copenhagen, February 22, 2019

Overview

A problem in landmark deformation

General problem formulation

Guided diffusion processes

Computing the guiding term

Examples
  Twice integrated sine example
  Application to landmark deformation

Concluding remarks
A problem in landmark deformation
A problem in landmark deformation

- Suppose we wish to deform one (non-intersecting closed) curve to another one in a smooth way.
- Match initial landmarks configuration

\[ x_0 = (x_{1,0}, \ldots, x_{n,0}) \]

to final configuration

\[ x_1 = (x_{1,1}, \ldots, x_{n,1}). \]

- Assume each landmark is in \( \mathbb{R}^d \).
• Let \((t, x) \mapsto v_t(x)\) be time dependent velocity fields and define the flow \(t \mapsto \Phi^v_t\) by

\[
\frac{\partial \Phi}{\partial t} = v_t \circ \Phi.
\]

• Diffeomorphic matching problem:

\[
\left\{ \begin{array}{l}
\min \int_0^1 \|v_t\|_V^2 \, dt, \\
\text{subject to } \Phi^v_1(x_0) = x_1
\end{array} \right\} \quad v \in L^2([0, 1], V)
\]

Suppose \(v \in L^2([0, 1], V)\) where \(V\) is a Reproducing Kernel Hilbert Space with kernel \(K_V\).

**Theorem**

The solution \(v \in L^2([0, 1], V)\) exists and satisfies

\[
v_t(q) = \sum_{i=1}^{n} K_V(q, q_{i,t}) p_{i,t}
\]

where \(t \mapsto (q_{i,t}, p_{i,t})\) solves

\[
\begin{align*}
\frac{d}{dt} q_{i,t}^\alpha &= \frac{\partial H}{\partial p_i^\alpha}(q_t, p_t) \\
\frac{d}{dt} p_{i,t}^\alpha &= -\frac{\partial H}{\partial q_i^\alpha}(q_t, p_t)
\end{align*}
\]

\((i = 1, \ldots, n, \alpha = 1, \ldots d)\) with Hamiltonian

\[
H(q, p) = \frac{1}{2} \sum_{i,j=1}^{n} \langle p_i, p_j \rangle K_V(q_i - q_j).
\]
Stochastic Hamiltonian dynamics on shape manifold

**Key idea:** define Hamiltonian dynamics with intrinsic noise.

**Marsland and Shardlow (2017):**

\[
\begin{align*}
\frac{d}{dt} q_{i, t} & = \frac{\partial H}{\partial p_i} (q_t, p_t) \\
\frac{d}{dt} p_{i, t} & = - \frac{\partial H}{\partial q_i} (q_t, p_t) + \sigma dW_{i, t}.
\end{align*}
\]

**Arnaudon, Holm and Sommer (2017):**

\[
\begin{align*}
\frac{d}{dt} q_{i, t} & = \frac{\partial H}{\partial p_i} (q_t, p_t) + \sum_{j=1}^{J} \sigma^i_j (q_{i, t}) \circ dW^j_t \\
\frac{d}{dt} p_{i, t} & = - \frac{\partial H}{\partial q_i} (q_t, p_t) - \sum_{j=1}^{J} \sum_{\beta=1}^{d} \frac{\partial \sigma^j_\beta (q_{i, t})}{\partial q_i} p_{i, t}^\beta \circ dW^j_t
\end{align*}
\]
Example: Marsland-Shardlow model - positions

Example: Arnaudon-Holm-Sommer model - positions
Example: Marsland-Shardlow model - momenta

Example: Arnaudon-Holm-Sommer model - momenta
Conditioned SDEs

- Behaviour of SDE is much richer than ODE.
- Can we sample the stochastic landmarks process to match both shapes?
- Reformulation: Can we guide the process from one shape to another?

General problem formulation
Problem description

- Suppose $X$ is a multivariate diffusion process in $\mathbb{R}^d$:
  \[
  dX_t = b(t, X_t) \, dt + \sigma(t, X_t) \, dW_t,
  \]
  where $b \in \mathbb{R}^d$, $\sigma \in \mathbb{R}^{d \times d'}$ and $W$ a $\mathbb{R}^{d'}$-valued Wiener process.
- Informally, for $h$ small and $Z \sim N_{d'}(0, I)$
  \[
  X_{t+h} = X_t + h b(t, X_t) + \sigma(t, X_t) \sqrt{h} Z,
  \]
- Let $t_0 < t_1 < \ldots < t_n$ and assume observations
  \[
  V_i = L_i X_{t_i} + \eta_i \quad i = 0, \ldots, n,
  \]
  with
  - each $L_i$ an $m_i \times d$ matrix with $m_i \leq d$
  - $\{\eta_i\}$ a sequence of independent random variables (independent of $X$) with $\eta_i \sim N(0, \Sigma_i)$. 

Examples for $L_i$

1. Suppose $X$ is two-dimensional.
   - $L_i = I_2$: observe all components.
   - $L_i = \begin{bmatrix} 1 & 0 \end{bmatrix}$: observe only first component.
   - $L_i = \begin{bmatrix} 1 & 1 \end{bmatrix}$: observe the sum of the two components.
2. For $n$ landmarks in $\mathbb{R}^d$, the state space of the diffusion is $\mathbb{R}^{2nd}$.
   Only the positions are observed, not the momenta.
Problem description

**Data:** \( \mathcal{D} = \{V_i, i = 0, \ldots, n\} \).

**Parameter estimation problem:**
- find point estimates for \( \theta \);
- draw from the posterior of \( \theta \), i.e. \( p(\theta | \mathcal{D}) \).

**Continuous-discrete smoothing:** reconstruct the path \( X := (X_t, t \in [0, t_n]) \), conditional on \( \mathcal{D} \).

**Main difficulty:** transition densities of \( X \) are intractable.

**Data-augmentation:** Sample from \( (\theta, X) | \mathcal{D} \) by alternating the steps

1. sample \( \theta | X \);
2. sample \( X | (\theta, \mathcal{D}) \).

Simplifying the problem

Consider the problem with observations

\[
X_0 \quad \text{and} \quad V_T = L_T X_T \quad (+ \text{ noise, if desired}).
\]

Assume \( \theta \) is known.

**Aim:** simulate \( (X_t, t \in [0, T]) \), conditional on \( (X_0, V_T) \)
Guided diffusion processes

SDE for the bridge process: case $L = I$

The conditioned process satisfies the SDE

$$dX_t^* = b(t, X_t^*) \, dt + a(t, X_t^*) r(t, X_t^*) \, dt + \sigma(t, X_t^*) \, dW_t,$$

with $a = \sigma \sigma'$,

$$r(t, x) = \nabla_x \log \rho(t, x)$$

and

$$\rho(t, x) = p(t, x, T, x_T).$$

Here $x_T = L x_T = v$.

Simplest example: Brownian bridge

$$dX_t^* = \frac{x_T - X_t^*}{T - t} \, dt + \sigma \, dW_t.$$
SDE for the bridge process: case $L \neq 1$

Suppose $X_t \in \mathbb{R}^d$ and rank $L = m < d$.

$$\begin{align*}
\{f_1, \ldots, f_m\} & \quad \text{orthonormal basis for Col}(L') \\
\{f_{m+1}, \ldots, f_d\} & \quad \text{orthonormal basis for Ker}(L)
\end{align*}$$

Suppose

$$x_T = x_T(\xi_1, \ldots, \xi_d) = \sum_{i=1}^{d} \xi_i f_i$$

is such that $Lx_T = v$.

Then (re)define

$$\rho(t, x) = \int_{\mathbb{R}^{d-m}} p(t, x; T, x_T) \, d\xi_{m+1} \cdots \xi_d.$$ 

Guided proposals

Recall the conditioned process satisfies

$$dX_t^* = b(t, X_t^*) \, dt + a(t, X_t^*)r(t, X_t^*) \, dt + \sigma(t, X_t^*) \, dW_t,$$

**Main idea:** replace intractable $p$ by $\tilde{p}$, with $\tilde{p}$ transition densities of

$$d\tilde{X}_t = (\tilde{B}(t)\tilde{X}_t + \tilde{\beta}(t)) \, dt + \tilde{\sigma}(t) \, dW_t.$$

Instead of sampling from the true process $X^*$, sample from the guided proposal that solves

$$dX_t^o = b(t, X_t^o) \, dt + a(t, X_t^o)\tilde{r}(t, X_t^o) \, dt + \sigma(t, X_t^o) \, dW_t.$$
Guided proposals

Key question: Does this work, and if so, under which conditions?

Theorem
Under “matching conditions”, we have
\[
\frac{d\mathbb{P}^*}{d\mathbb{P}^0}(X^\circ) = \frac{\tilde{\rho}(0, x_0)}{\rho(0, x_0)} \Psi_T(X^\circ)
\]

Matching conditions

Matching on the diffusivity:
\[
L\alpha(T, X_T)L' = L\tilde{\alpha}(T)L'.
\]

Matching on the drift (only in case of hypo-ellipticity):
\[
Lb(T, X_T) = \tilde{L}b(T, X_T).
\]

Note:
- this has not been proved in full generality yet;
- mainly difficulties in the proof when \(\sigma\) depends on \(x\) and/or hypo-elliptic diffusions.
Let $X^\circ$ to the guided proposal, i.e. $X^\circ$ is the strong solution to

$$dX_t^\circ = (b(t, X_t^\circ) + a(t, X_t^\circ)\tilde{r}(t, X_t^\circ)) \, dt + \sigma(t, X_t^\circ) \, dW_t$$

There is a measurable mapping $g$ such that $X^\circ = g(W)$.

**Algorithm**

Choose $\rho \in [0, 1)$.

1. Draw a Wiener process $Z$ on $[0, T]$, Set $X = g(Z)$.
2. Propose a Wiener process $W$ on $[0, T]$. Set

$$Z^\circ = \rho Z + \sqrt{1 - \rho^2} W$$

and $X^\circ = g(Z^\circ)$.
3. Compute $A = \Psi_T(X^\circ)/\Psi_T(X)$ (where $\Psi_T$ is the likelihood ratio).
   Sample $U \sim \text{Uniform}(0, 1)$. If $U < A$ then set $Z = Z^\circ$ and $X = X^\circ$.
4. Repeat steps (2) and (3).
Computing the guiding term \((1/2)\)

**Lemma**

Assume

\[ d\tilde{X}_t = \tilde{B}(t)\tilde{X}_t \, dt + \tilde{\sigma}(t) \, dW_t \]

and that

\[ \int_t^T \Phi(T, \tau)\tilde{a}(\tau)\Phi(T, \tau) \, d\tau \]

is strictly positive definite for \( t < T \). Then

\[ \tilde{r}(t, x) = L(t)'M(t) (v - L(t)x) , \quad t \in [0, T] , \]

where \( M(t) = \left( \int_t^T L(\tau)\tilde{a}(\tau)L(\tau)' \, d\tau \right)^{-1} \).

Here \( \Phi \) solves

\[ d\Phi(t) = \tilde{B}(t)\Phi(t) \, dt , \quad \Phi(0) = I \]

and

\[ \Phi(t, s) = \Phi(t)\Phi(s)^{-1} , \quad L(t) = L\Phi(T, t) . \]

Computing the guiding term \((2/2)\)

Backwards solve

\[ dL(t) = -L(t)\tilde{B}(t) \, dt , \quad L(T) = L_T \]

Approximate

\[ M(t)^{-1} = \int_t^T L(\tau)\tilde{a}(\tau)L(\tau)' \, d\tau \]

by a numerical quadrature rule (for example trapezoidal rule).
Examples

Example: twice integrated sine

Assume
\[
\begin{align*}
    \frac{dX_t}{dt} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} X_t \ dt + \begin{bmatrix} 0 \\ 0 \\ -6 \sin(2\pi x) \end{bmatrix} dt + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} dW_t, \\
    X_0 &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\end{align*}
\]

Assume either

- \( L = I \), or
- \( L = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \)\(^\prime \).
Case $L = I$

![Graph showing three components over time]

**Figure 1:** Sampled guided diffusion bridges when conditioning on $X_T = \begin{bmatrix} 1/32 & 1/4 & 1 \end{bmatrix}$. $\rho = 0.85$

---

Case $L = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}'$

![Graph showing three components over time]

**Figure 2:** Sampled guided diffusion bridges when conditioning on $LX_T = 1/32$. $\rho = 0.95$
Guided proposals: Marsland-Shardlow model

Guided proposals: Arnaudon-Holm-Sommer model
Concluding remarks

**Main result:** conditional simulation in stochastic landmark deformation models.

- Approach also applies to observations at multiple times $t_0 < t_1 < \cdots < t_n$. *Backward ODEs for computing guiding term resembles those in Kalman smoothing.*
- Obtain full theoretical backup.
- Include parameter estimation.
- Smart MCMC moves on initial momenta and or positions (use of dualnumbers and automatic differentiation).
- Include unknown landmark label matching.
- Applications to other SDE models.
- ...