Langevin equations for landmark image registration with uncertainty

Stephen Marsland
(Joint work with Tony Shardlow, University of Bath)

Massey University, New Zealand

Image Registration with Landmarks

- Define a set of landmarks (automatically or manually) on two or more images (reference $\mathbf{q}_i^R$ and template $\mathbf{q}_i^T$)
- Find deformation (warp) ϕ such that $\phi(\mathbf{q}_i^R) = \mathbf{q}_i^T$
- Apply to rest of image

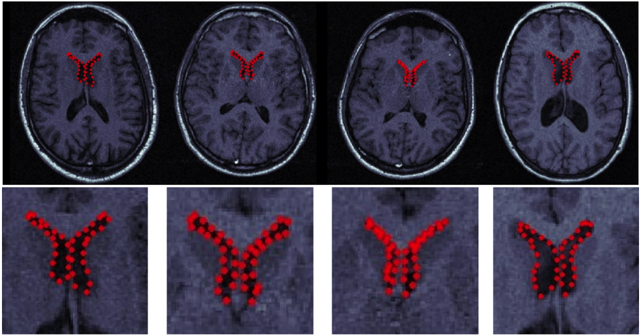


Image Registration with Landmarks

- Many choices of ϕ, usually desire a diffeomorphism
- Can define a velocity field $v(\mathbf{q}, t)$ that takes the landmarks from $\mathbf{q}_i^R$ at $t = 0$ to $\mathbf{q}_i^T$ at $t = 1$
- Expand $v(\mathbf{q}, t) = \sum_{i=1}^{n} \mathbf{p}_i(t) G(\mathbf{q}, \mathbf{q}_j(t))$

Image Registration with Landmarks

- Leads to reduced optimisation problem

$$\min \int_0^1 \sum_{i=1}^{n} \mathbf{p}_i(t)\mathbf{p}_j(t) G(\mathbf{q}_i(t), \mathbf{q}_j(t)) dt$$

with velocity constraints

$$\frac{d\mathbf{q}_i}{dt} = \sum_{i=1}^{n} \mathbf{p}_i(t) G(\mathbf{q}_i(t), \mathbf{q}_j(t))$$

and landmark constraints

$$\mathbf{q}_i(0) = \mathbf{q}_i^R \text{ and } \mathbf{q}_i(1) = \mathbf{q}_i^T.$$

- Usually take G to be Gaussian: $G(\mathbf{q}_i, \mathbf{q}_j) = \exp(-\|\mathbf{q}_i - \mathbf{q}_j\|^2 / l^2)$
- Finite dimensional in space

Optimal Control Formulation

In Hamiltonian formulation we are solving a boundary-value problem:

$$\begin{aligned}\frac{d\mathbf{q}}{dt} &= \nabla_p H(\mathbf{q}, \mathbf{p}), \\ \frac{d\mathbf{p}}{dt} &= -\nabla_q H(\mathbf{q}, \mathbf{p})\end{aligned}$$

for

$$H = \frac{1}{2}\sum_{i,j} \mathbf{p}_i^T \mathbf{p}_j G(\mathbf{q}_i, \mathbf{q}_j),$$

with boundary conditions

$$\mathbf{q}_i(0) = \text{reference}, \mathbf{q}_j(1) = \text{target}.$$

Defines a shooting method, need appropriate time-stepping method (Euler) and initial momenta $\mathbf{p}(0)$.

Uncertainty in Landmarks

- Two major sources of uncertainty
  - landmark positions
  - velocity field far from landmarks
- How can you incorporate this uncertainty?

Uncertainty in Landmarks

Shape Model

Langevin Equation

- Formulation is based on conservation of energy (the Hamiltonian)
- Could enclose system in a heat bath
  - Langevin equation
  - constant temperature in equilibrium
- Classically:
  - system of SDEs for potential energy V:

$$
\begin{aligned}
d\mathbf{q}_i &= \mathbf{p}_i dt \\
d\mathbf{p}_i &= [-\lambda \mathbf{p}_i - \nabla_{\mathbf{q}_i} V]dt + \sigma d\mathbf{W}_i(t),
\end{aligned}
$$

    where $\mathbf{W}_i(t)$ is *iid* $\mathbb{R}^2$ Brownian motion.
  - This has invariant measure $\exp(-\beta H)$ where $H = \frac{1}{2}\sum_i \|\mathbf{p}_i\|^2 + V(q)$ and $\beta = 2\lambda/\sigma^2$ is the inverse temperature

Langevin Equation

- Formulation is based on conservation of energy (the Hamiltonian)
- Could enclose system in a heat bath
  - Langevin equation
  - constant temperature in equilibrium
- Classically:
  - system of SDEs for potential energy V:

$$
\begin{aligned}
d\mathbf{q}_i &= \mathbf{p}_i dt \\
d\mathbf{p}_i &= [-\lambda \mathbf{p}_i - \nabla_{\mathbf{q}_i} V] dt + \sigma d\mathbf{W}_i(t),
\end{aligned}
$$

    where $\mathbf{W}_i(t)$ is *iid* $\mathbb{R}^2$ Brownian motion.
  - This has invariant measure $\exp(-\beta H)$ where $H = \frac{1}{2}\sum_i \|\mathbf{p}_i\|^2 + V(q)$ and $\beta = 2\lambda/\sigma^2$ is the inverse temperature

Noisy Landmark Registration

- The generalised version is:

$$\begin{aligned} d\mathbf{q}_i &= \nabla_{\mathbf{p}_i} H dt \\ d\mathbf{p}_i &= [-\lambda \nabla_{\mathbf{p}_i} H - \nabla_{\mathbf{q}_i} H] dt + \sigma d\mathbf{W}_i(t). \end{aligned}$$

- Gibbs distribution still invariant
- For large β, system is concentrated at small H
- So solve this system subject to

$$\mathbf{q}_i(0) = \mathbf{q}_i^R + \text{noise}, \mathbf{q}_i(1) = \mathbf{q}_i^T + \text{noise}$$

  and $\mathbf{p}_i(0) \sim \exp(-\beta H(\mathbf{q}_i^R, \cdot))$. Noise modelled as zero mean, *iid* Gaussian
- For H separable this was studied by Hairer et al., 2011

Noisy Landmark Registration

- The generalised version is:

$$\begin{aligned} d\mathbf{q}_i &= \nabla_{\mathbf{p}_i} H dt \\ d\mathbf{p}_i &= [-\lambda \nabla_{\mathbf{p}_i} H - \nabla_{\mathbf{q}_i} H] dt + \sigma d\mathbf{W}_i(t). \end{aligned}$$

- Gibbs distribution still invariant
- For large β, system is concentrated at small H
- So solve this system subject to

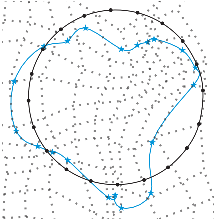
$$\mathbf{q}_i(0) = \mathbf{q}_i^R + \text{noise}, \mathbf{q}_i(1) = \mathbf{q}_i^T + \text{noise}$$

  and $\mathbf{p}_i(0) \sim \exp(-\beta H(\mathbf{q}_i^R, \cdot))$. Noise modelled as zero mean, *iid* Gaussian

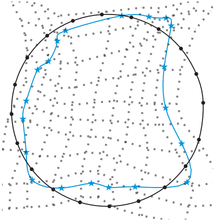
- For H separable this was studied by Hairer et al., 2011

Inverse Temperature Effect

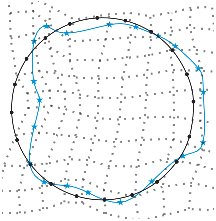
$\beta = 10$



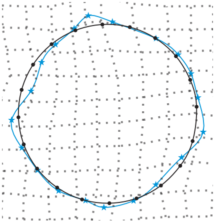
$\beta = 20$



$\beta = 40$

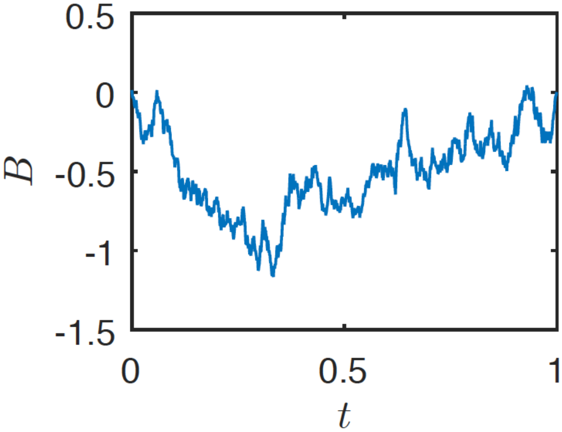


$\beta = 80$



Brownian Bridges

- B is a Brownian Bridge if its distribution is $W(t)$ conditioned by $W(T) = b$
- No noise on boundary conditions
- Example of a Brownian bridge conditioned on $W(1) = 0$



Brownian Bridges

- Brownian motion is a Gaussian random field, defined by mean and covariance
- Can condition this on observations, get a second Gaussian random field with well-known formulae for mean and covariance
- Covariance does not depend on b

Diffusion Bridges

- Bridges very useful for parameter estimation for SDEs with discrete data
- Diffusion process $X(t)$ with SDE

$$dX = f(X)dt + \sigma(X)dW(t), X(0) = a, t \in [0, T]$$

- Bridge process satisfies a modified SDE

$$dX = \tilde{f}(X, t)dt + \sigma(X)dW(t), t \in [0, T]$$

  where $\tilde{f}(x, t) = f(x) + [\sigma(x)\sigma^T(x)]\nabla \log p(t, x; T, b)$, where p is the transition density
- This is the Doob h-transform
- But hard to find p
- Change of measure relative to a linear diffusion bridge. Girsanov formula gives the density and hence a formula for rejection sampling

Diffusion Bridges

- Bridges very useful for parameter estimation for SDEs with discrete data
- Diffusion process $X(t)$ with SDE

$$dX = f(X)dt + \sigma(X)dW(t), X(0) = a, t \in [0, T]$$

- Bridge process satisfies a modified SDE

$$dX = \tilde{f}(X, t)dt + \sigma(X)dW(t), t \in [0, T]$$

where $\tilde{f}(x, t) = f(x) + [\sigma(x)\sigma^T(x)]\nabla \log p(t, x; T, b)$, where p is the transition density

- This is the Doob h-transform
- But hard to find p
- Change of measure relative to a linear diffusion bridge. Girsanov formula gives the density and hence a formula for rejection sampling

Diffusion Bridges

- Bridges very useful for parameter estimation for SDEs with discrete data
- Diffusion process $X(t)$ with SDE

$$dX = f(X)dt + \sigma(X)dW(t), X(0) = a, t \in [0, T]$$

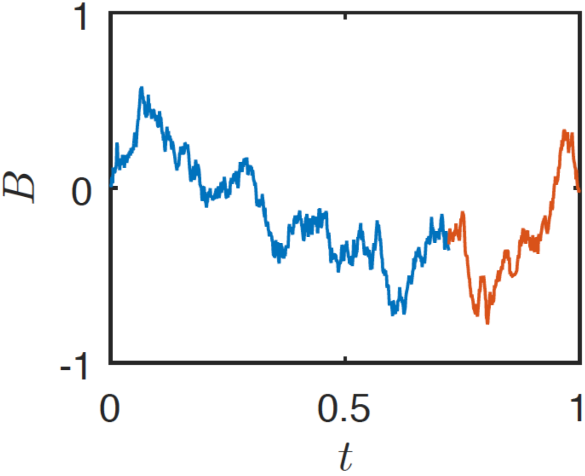
- Bridge process satisfies a modified SDE

$$dX = \tilde{f}(X, t)dt + \sigma(X)dW(t), t \in [0, T]$$

  where $\tilde{f}(x, t) = f(x) + [\sigma(x)\sigma^T(x)]\nabla \log p(t, x; T, b)$, where p is the transition density
- This is the Doob h-transform
- But hard to find p
- Change of measure relative to a linear diffusion bridge. Girsanov formula gives the density and hence a formula for rejection sampling

Diffusion Bridges

- Simulate forward and backward, and reject samples until they cross
- Approximates the diffusion bridge
- Uses a Metropolis-Hastings iteration



Diffusion Bridges

- Dimensionality problem: paths less likely to cross as number of dimensions increases. Can sometimes find clever coupling of forward and backward Brownian motions.

- If H is separable, can replace H by $\frac{1}{2}\sum_i \|\mathbf{p}_i\|^2$
- Then have a linear SDE
- We are in high dimensions and have a non-separable Hamiltonian ....

Diffusion Bridges

- Dimensionality problem: paths less likely to cross as number of dimensions increases. Can sometimes find clever coupling of forward and backward Brownian motions.

- If H is separable, can replace H by $\frac{1}{2}\sum_i \|\mathbf{p}_i\|^2$
- Then have a linear SDE
- We are in high dimensions and have a non-separable Hamiltonian ....

Diffusion Bridges

- Dimensionality problem: paths less likely to cross as number of dimensions increases. Can sometimes find clever coupling of forward and backward Brownian motions.

- If H is separable, can replace H by $\frac{1}{2}\sum_i \|\mathbf{p}_i\|^2$
- Then have a linear SDE
- We are in high dimensions and have a non-separable Hamiltonian ....

Approximation

- One good thing: Langevin trajectories are close to the Hamiltonian ones since have short time interval and small perturbation
- So let's approximate:

Approximation I: Linearisation

- Linearise about solution of Hamiltonian BVP
- Can discretise resulting linear SDE and compute covariance structure.

1. Solve Hamiltonian Boundary Value Problem. Use mean position of uncertain landmarks
2. Linearise Langevin equation about this solution
3. For initial data at $t = \frac{1}{2}$, propagate forward and backward to find a Gaussian prior distribution, i.e., compute mean and covariance for process $(\mathbf{q}_i(t), \mathbf{p}_i(t))$

- Use explicit Euler
- Condition on boundary data (Euler-Maruyama)

Approximation I: Linearisation

- Linearise about solution of Hamiltonian BVP
- Can discretise resulting linear SDE and compute covariance structure.

1. Solve Hamiltonian Boundary Value Problem. Use mean position of uncertain landmarks
2. Linearise Langevin equation about this solution
3. For initial data at $t = \frac{1}{2}$, propagate forward and backward to find a Gaussian prior distribution, i.e., compute mean and covariance for process $(\mathbf{q}_i(t), \mathbf{p}_i(t))$

- Use explicit Euler
- Condition on boundary data (Euler-Maruyama)

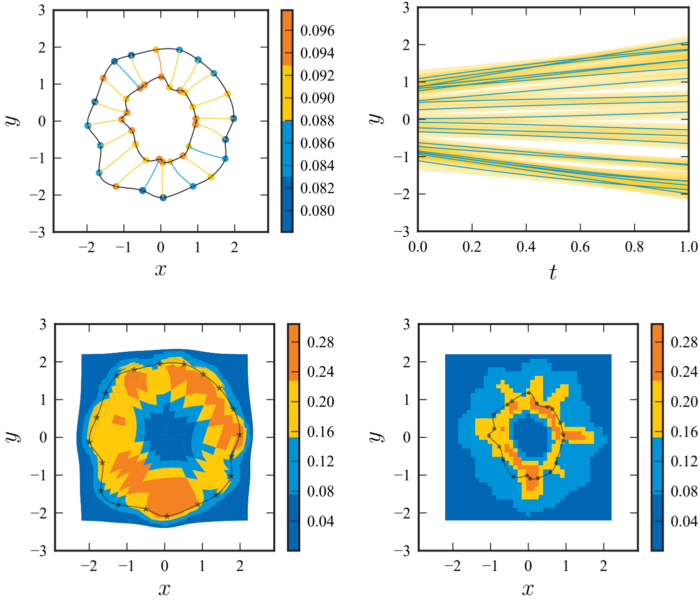
Approximation I: Linearisation

- Linearise about solution of Hamiltonian BVP
- Can discretise resulting linear SDE and compute covariance structure.

1. Solve Hamiltonian Boundary Value Problem. Use mean position of uncertain landmarks
2. Linearise Langevin equation about this solution
3. For initial data at $t = \frac{1}{2}$, propagate forward and backward to find a Gaussian prior distribution, i.e., compute mean and covariance for process $(\mathbf{q}_i(t), \mathbf{p}_i(t))$

- Use explicit Euler
- Condition on boundary data (Euler-Maruyama)

Results: Linearisation



Approximation II: Operator Splitting

- Write generator $\mathcal{L} = \mathcal{L}_0 + \sigma^2 \mathcal{L}_1$, where

$$\begin{aligned}\mathcal{L}_0 &= H_p \frac{\partial}{\partial q} - H_q \frac{\partial}{\partial p} \text{(Liouville operator)} \\ \mathcal{L}_1 &= -\frac{\beta}{2} H_p \frac{\partial}{\partial p} + \frac{1}{2} \frac{\partial^2}{\partial p^2}\end{aligned}$$

- Both have Gibbs invariant and a BCH formula.

Approximation II: Operator Splitting

- Two choices of Strang splitting:

$$e^{L^*} \approx e^{\sigma^2 L_1^*/2} e^{L_0^*} e^{\sigma^2 L_1^*/2}$$

  - Gibbs distribution on initial data
  - Hamiltonian flow
  - Posterior is not Gaussian, but can compute MAP estimate (Laplace approximation)

$$e^{L^*} \approx e^{L_0^*/2} e^{\sigma^2 L_1^*} e^{L_0^*/2}$$

  - Gibbs distribution on midpoint
  - Hamiltonian flow
  - Parameterised by one set of landmarks and two sets of momenta
  - Can compute averages of sets of landmarks

Approximation II: Operator Splitting

Lead to different semigroups:

$$[\mathbf{p}(0), \mathbf{q}(0)] \underbrace{\longmapsto}_{e^{\sigma^2 L_1^*/2}} [\mathbf{p}(1/2), \mathbf{q}(0)] \underbrace{\longmapsto}_{e^{L_0^*}} [\tilde{\mathbf{p}}(1/2), \mathbf{q}(1)] \underbrace{\longmapsto}_{e^{\sigma^2 L_1^*/2}} [\mathbf{p}(1), \mathbf{q}(1)].$$

$$[\mathbf{p}(0), \mathbf{q}(0)] \underbrace{\longmapsto}_{e^{L_0^*/2}} [(\mathbf{p}(1/2), \mathbf{q}(1/2)] \underbrace{\longmapsto}_{e^{\sigma^2 L_1^*}} [(\tilde{\mathbf{p}}(1/2), \mathbf{q}(1/2)] \underbrace{\longmapsto}_{e^{L_0^*/2}} [\mathbf{p}(1), \mathbf{q}(1)].$$

Time-half steps governed by Ornstein–Uhlenbeck SDE

$$d\vec{p} = -\lambda \nabla_{\mathbf{p}} H(\mathbf{p}, \mathbf{q}_0)\, dt + \sigma\, d\mathbf{W}(t), \qquad \mathbf{p}(0) = \mathbf{p}_0,$$

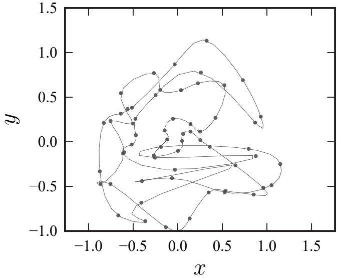
Laplace Approximation

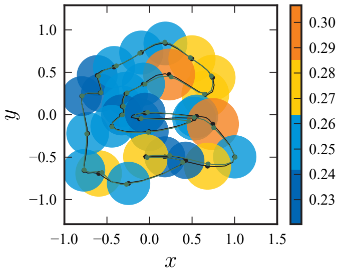
- Approximate posterior covariance max by second-order approximation to likelihood at MAP point
- Unconstrained optimisation by Gauss-Newton

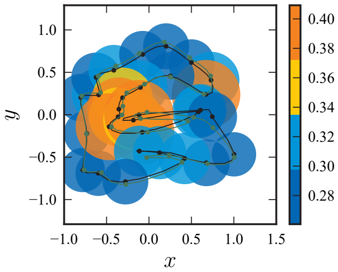
For a change of optimisation function, can construct average over sets of flows (where $(\mathbf{p}^*, \mathbf{q}^*)$ is the average)

$$F(\mathbf{p}^*, \mathbf{q}^*, \mathbf{p}^j) := \beta H(\mathbf{p}^*, \mathbf{q}^*) + \frac{\beta}{4\lambda} \sum_{j=1}^{J} \|\mathbf{p}^j - e^{-\lambda \mathcal{G}(\mathbf{q}^*)} \mathbf{p}^* \|^2 + \frac{1}{2\delta^2} \sum_{j=1}^{J} \|\mathbf{q}^j - S_q(1/2; 0, [\mathbf{p}^j, \mathbf{q}^*])\|^2.$$

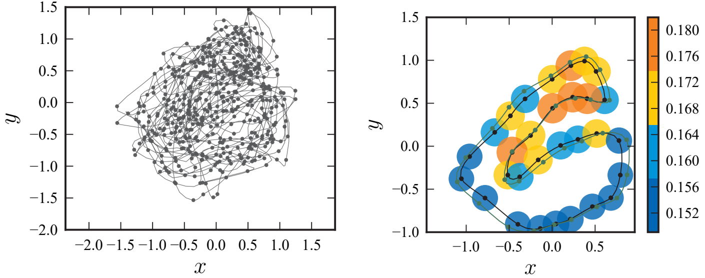
Results: Averaging (2 sample)



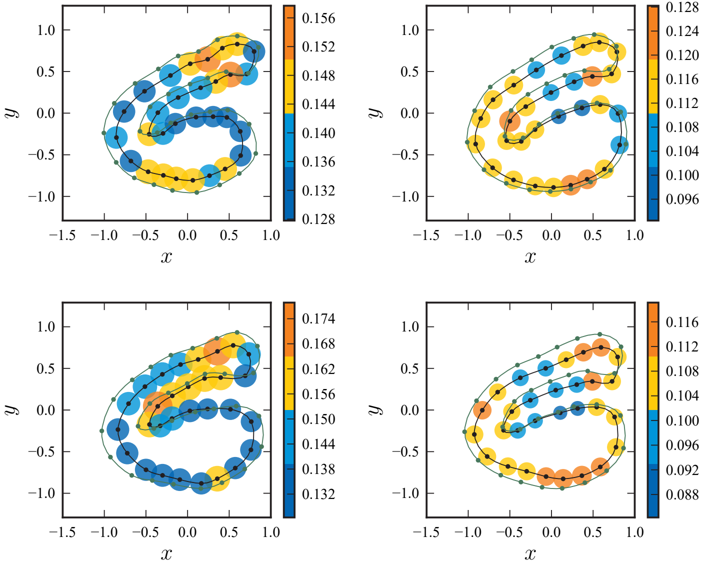




Results: MAP average (16 samples)



Results: MAP average (64 and 256 samples)



Conclusions

- Introduced Langevin equation to define prior distribution on set of diffeomorphisms
- Computationally difficult to sample the diffusion bridge
- Hence, have linearised the equations, and also made an operator splitting
- Second approach enables averaging of multiple landmark sets