Nonparametric Drift Estimation for Stochastic Differential Equations

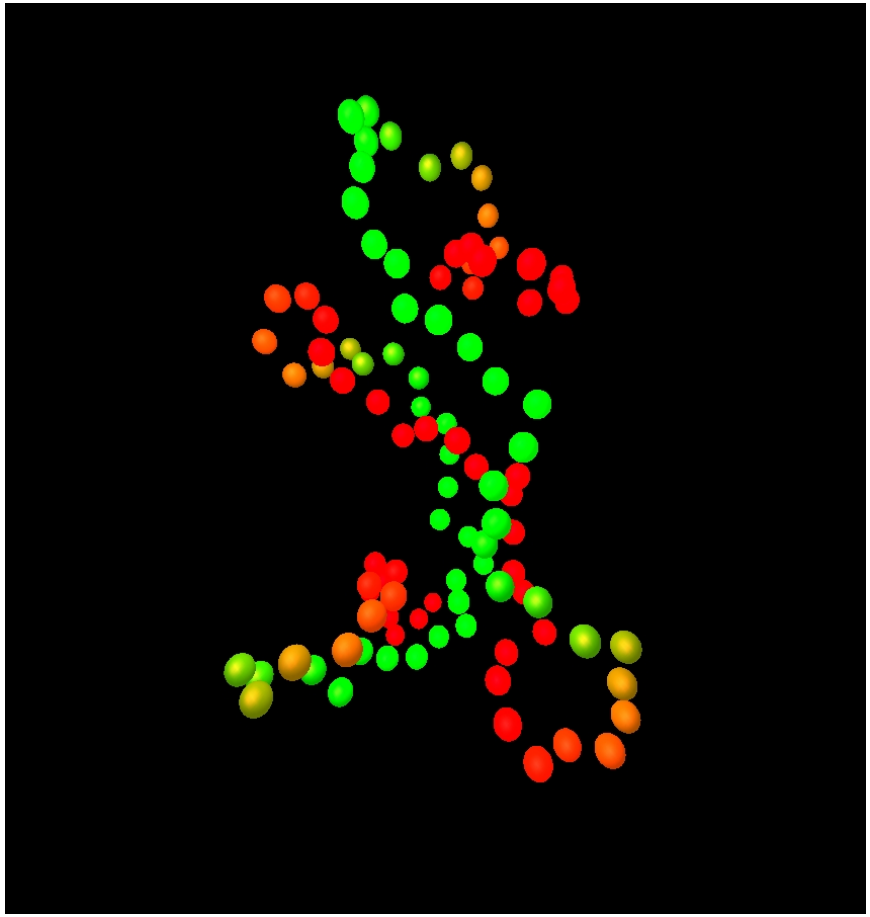
Yvo Pokern

EPSRC Symposium on MCMC and Related Methods
Joint work with O. Papaspiliopoulos, G. Roberts and A. Stuart

Outline

Fitting SDEs to Molecular Dynamics

MD data



$$X(n\Delta t) \in \mathbb{R}^N$$

- Multiple Timescales
- High frequency data
- High dimension, only few dimensions of chemical interest
- Diffusion good description at some timescales only.

Programme

Start from SDE

$$dx = b(x)dt + dB, \quad x(0) = x_0$$

and high-frequency discrete time observations x_i.

- Write down likelihood for $b(\cdot)$ on function space H.
- Modify likelihood to make the local time L an (almost) sufficient statistic.
- Specify prior on function space H, compute posterior. Make Bayesian framework rigorous.
- Application: Toy example from Molecular Dynamics

SDE properties - Girsanov

$$dx = b(x)dt + dB$$

- Generates measure $\mathbb{P}$ on path space $\mathcal{C}([0, T], [0, 2\pi))$.
- $\mathbb{P}$ is absolutely continuous w.r.t. $\mathbb{Q}$ generated by Brownian Motion.
- The Radon-Nikodym derivative is

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp(I[b])$$

$I[b]$ viewed as functional of the drift:

$$I[b] = -\frac{1}{2}\int_0^T \left(b^2(x)dt - 2b(x)dx\right)$$

SDE properties - Girsanov

$$dx = b(x)dt + dB$$

- Generates measure $\mathbb{P}$ on path space $\mathcal{C}([0, T], [0, 2\pi))$.
- $\mathbb{P}$ is absolutely continuous w.r.t. $\mathbb{Q}$ generated by Brownian Motion.
- The Radon-Nikodym derivative is

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp(I[b])$$

$I[b]$ viewed as functional of the drift:

$$I[b] = -\frac{1}{2}\int_0^T \left(b^2(x)dt - 2b(x)dx\right)$$

Local Time

Local time is the empirical occupation density:

$$\frac{1}{T}\int_0^T f(x_t)dt = \int_{\mathbb{R}} f(a)\frac{L(a)}{T}da$$

- L is continuous but not differentiable.
- In the limit $T \to \infty$ it becomes smooth:

$$\lim_{T\to\infty}\frac{1}{T}L = \varrho^{\infty}$$

The Likelihood Functional

$$l[b] = \log \frac{d\mathbb{P}}{d\mathbb{Q}} = -\frac{1}{2}\int_0^T \left(b^2(x)dt - 2b(x)dx\right)$$

- Start from the Girsanov change of measure.
- Apply the Ito formula for $V(x)$ to rewrite the stochastic integral as boundary terms plus correction.
- Replace Integrals along the trajectory by integral against local time $L(a)da$.

The Likelihood Functional

$$I[b] = -\frac{1}{2}\int_0^T \left(b^2(x)dt - 2b(x)dx\right)$$

- Start from the Girsanov change of measure.
- Apply the Ito formula for $V(x)$ to rewrite the stochastic integral as boundary terms plus correction.
- Replace Integrals along the trajectory by integral against local time $L(a)da$.

The Likelihood Functional

$$I[b] = \frac{1}{2}\left(V(X_T) - V(X_0)\right) + \frac{W}{2}\left(V(2\pi) - V(0)\right) - \frac{1}{2}\int_0^T \left(b^2(x) + 2b'(x)\right) dt.$$

- Start from the Girsanov change of measure.
- Apply the Ito formula for $V(x)$ to rewrite the stochastic integral as boundary terms plus correction.
- Replace Integrals along the trajectory by integral against local time $L(a)da$.

The Likelihood Functional

$$I[b] = \frac{1}{2}\left(V(X_T) - V(X_0)\right) + \frac{W}{2}\left(V(2\pi) - V(0)\right) - \frac{1}{2}\int_0^{2\pi}\left(b^2(a) + 2b'(a)\right) L(a)da.$$

- Start from the Girsanov change of measure.
- Apply the Ito formula for $V(x)$ to rewrite the stochastic integral as boundary terms plus correction.
- Replace Integrals along the trajectory by integral against local time $L(a)da$.

If local time was smooth

$$
\begin{aligned}
l[b] =& \text{Boundary} - \frac{1}{2}\int \left(b^2 + 2b'\right) L da \\
=& \text{Boundary} - \frac{1}{2}\int b^2 L - 2bL' da \\
\leq& \text{Boundary} - \|L\|_\infty \|b\|_{L^2}^2 - \|b\|_{L^2}^2 - \|L'\|_{L^2}^2
\end{aligned}
$$

- the log-likelihood $l[b]$ is bounded above on $b \in L^2(0, 2\pi)$.
- Taking the functional derivative yields the MLE

$$\widehat{b} = \frac{L'}{2L}$$

If local time was smooth

$$
\begin{aligned}
l[b] =& \text{Boundary} - \frac{1}{2}\int \left(b^2 + 2b'\right) L da \\
=& \text{Boundary} - \frac{1}{2}\int b^2 L - 2bL' da \\
\leq& \text{Boundary} - \|L\|_\infty \|b\|_{L^2}^2 - \|b\|_{L^2}^2 - \|L'\|_{L^2}^2
\end{aligned}
$$

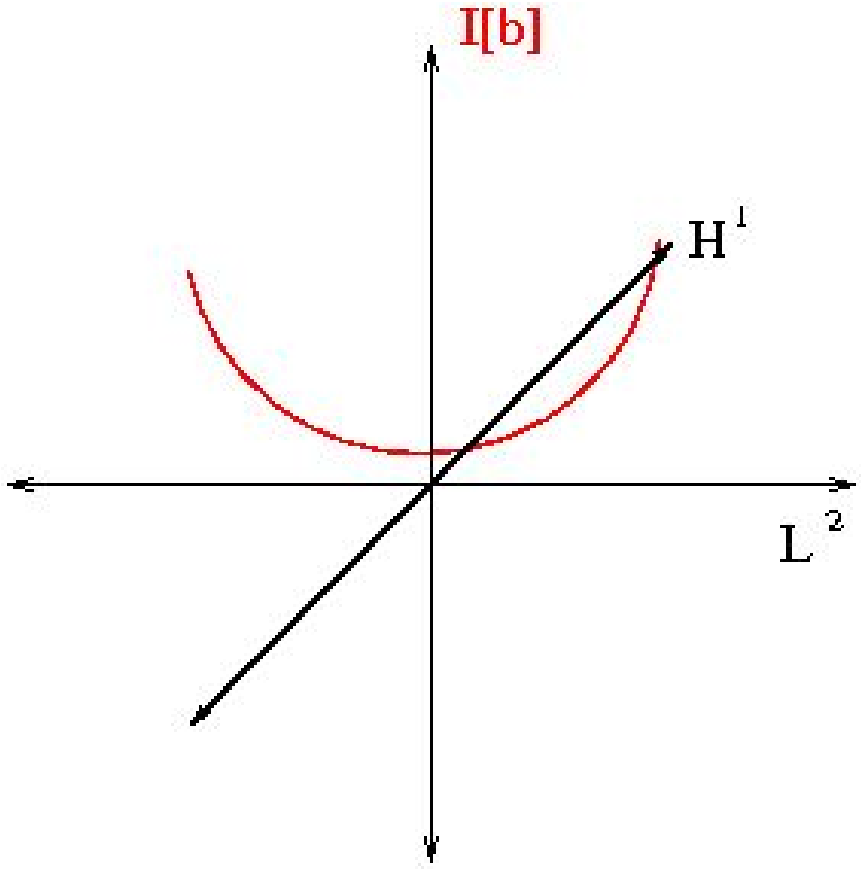
- the log-likelihood $l[b]$ is bounded above on $b \in L^2(0, 2\pi)$.
- Taking the functional derivative yields the MLE

$$
\widehat{b} = \frac{L'}{2L}
$$

Infinite Dimensional Trouble

$$-I[b] = \int_0^{2\pi} |b(a)|^2 L(a) + b'(a) L(a) da$$

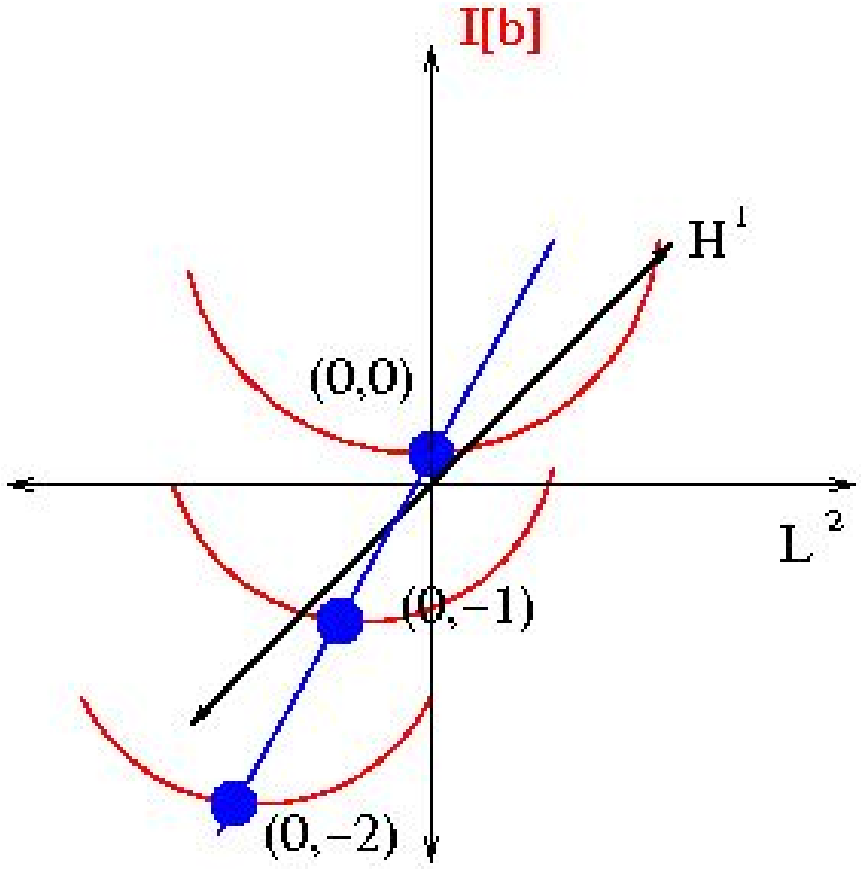
- For smooth L the functional is positive definit.
- Quadratic positive definit functionals are bounded below.
- **BUT** L is not differentiable!



Infinite Dimensional Trouble

$$-I[b] = \int_0^{2\pi} \underbrace{|b(a)|^2 L(a)}_{\text{Term A}} + \underbrace{b'(a)L(a)}_{\text{Term B}} \, da$$

- Quadratic in the L^2-direction (Term A)
- Linear in the Derivative-direction (Term B)
- Hence cylindrical paraboloid - not bounded below!



Three Options

$$\text{Boundary} - \frac{1}{2}\int_0^{2\pi} \left(|b(a)|^2 + b'(a)\right) L(a)da.$$

1. Assume a parametric form $b(x, \theta)$
2. Introduce a regularised version of $L(a)da$.
3. Introduce a prior measure on drift functions $b(\cdot)$ and perform Bayesian estimation.

Gaussian Prior on drift functions

Specify a prior Gaussian measure for zero-mean drift functions by

- Its Mean:

$$b_0 \in H^2_{\text{per}}([0, 2\pi])$$

- Its Precision (operator): $\mathcal{C}_0 = \Delta^2$ on $[0, 2\pi]$ periodic, mean-zero.

Formally:

$$db \propto \exp\left(-\int_0^{2\pi} |\Delta b(a) - \Delta b_0(a)|^2 \, da\right) d\lambda$$

Finding the Posterior - formally

Multiply prior "density" by likelihood:

$$\mu \propto \exp\left(-\int_0^{2\pi} |\Delta b(a)|^2 \, da\right) \cdot \exp\left(\int_0^{2\pi} |b(a)|^2 L(a) + b'(a) L(a) da\right)$$

Complete the square to find that the posterior is Gaussian with

- Mean

$$\left(\Delta^2 + L\right) \widehat{b} = \frac{1}{2} L' + \tilde{\chi} x_0, x_T + W$$

- Posterior Covariance

$$\left(\Delta^2 + L\right)^{-1}$$

Finding the Posterior - formally

Multiply prior "density" by likelihood:

$$\mu \propto \exp\left(-\int_0^{2\pi} |\Delta b(a)|^2 \, da\right) \cdot \exp\left(\int_0^{2\pi} |b(a)|^2 L(a) + b'(a) L(a) da\right)$$

Complete the square to find that the posterior is Gaussian with

- Mean

$$\left(\Delta^2 + L\right) \widehat{b} = \frac{1}{2} L' + \tilde{\chi} x_0, x_T + W$$

- Posterior Covariance

$$\left(\Delta^2 + L\right)^{-1}$$

Posterior Mean is the solution of a PDE

Theorem

Let $L \in \mathcal{C}([0, 2\pi])$ be continuous and periodic and not identically zero. Then the PDE for the posterior mean

$$\Delta^2 u + Lu = \frac{1}{2}L' + W + \tilde{\chi}_{x_0, x_T} \tag{1}$$

has a unique weak solution $u \in H^2_{\mathrm{per}}([0, 2\pi])$.

Robustness of the Posterior Mean

Theorem

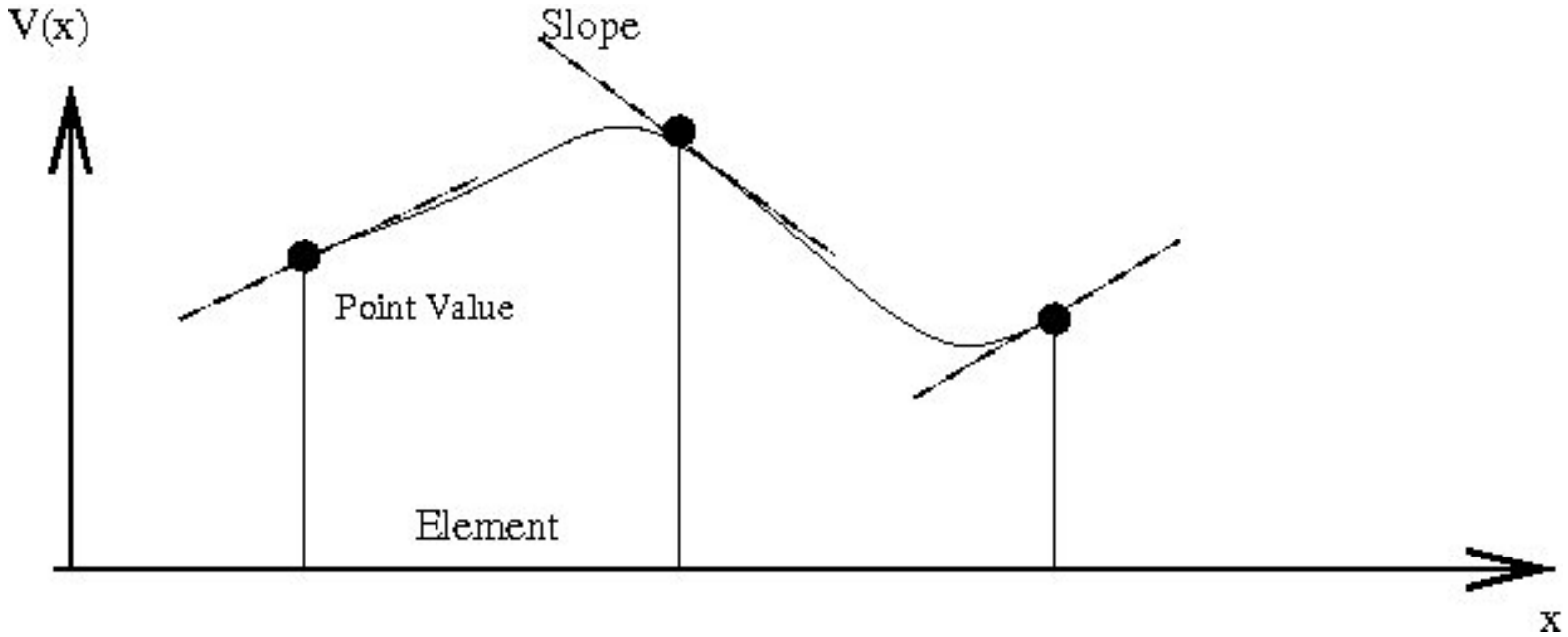
There exists a constant $C(W, \|L\|_\infty) > 0$ such that for all admissible perturbed local times $\tilde{L}$ the deviation of the perturbed posterior mean $\tilde{u}$ from the unpterturbed posterior mean u is bounded in the H^2-norm:

$$\|\tilde{u} - u\|_{H^2} \leq C(W, \|L\|_\infty)\|\tilde{L} - L\|_{L^2} \tag{2}$$

Cleanup

- Observe absolute continuity of posterior and prior measure, Δ^2 and $\Delta^2 + L$ differ only in lower order differential parts.
- Compute the Radon-Nikodym derivative and identify with the likelihood.
- Simple estimate of local time from pointwise estimations combined with Hölder continuity of local time, so that $\|\hat{L} - L\|_{L^2}$ is small.

Numerical Treatment

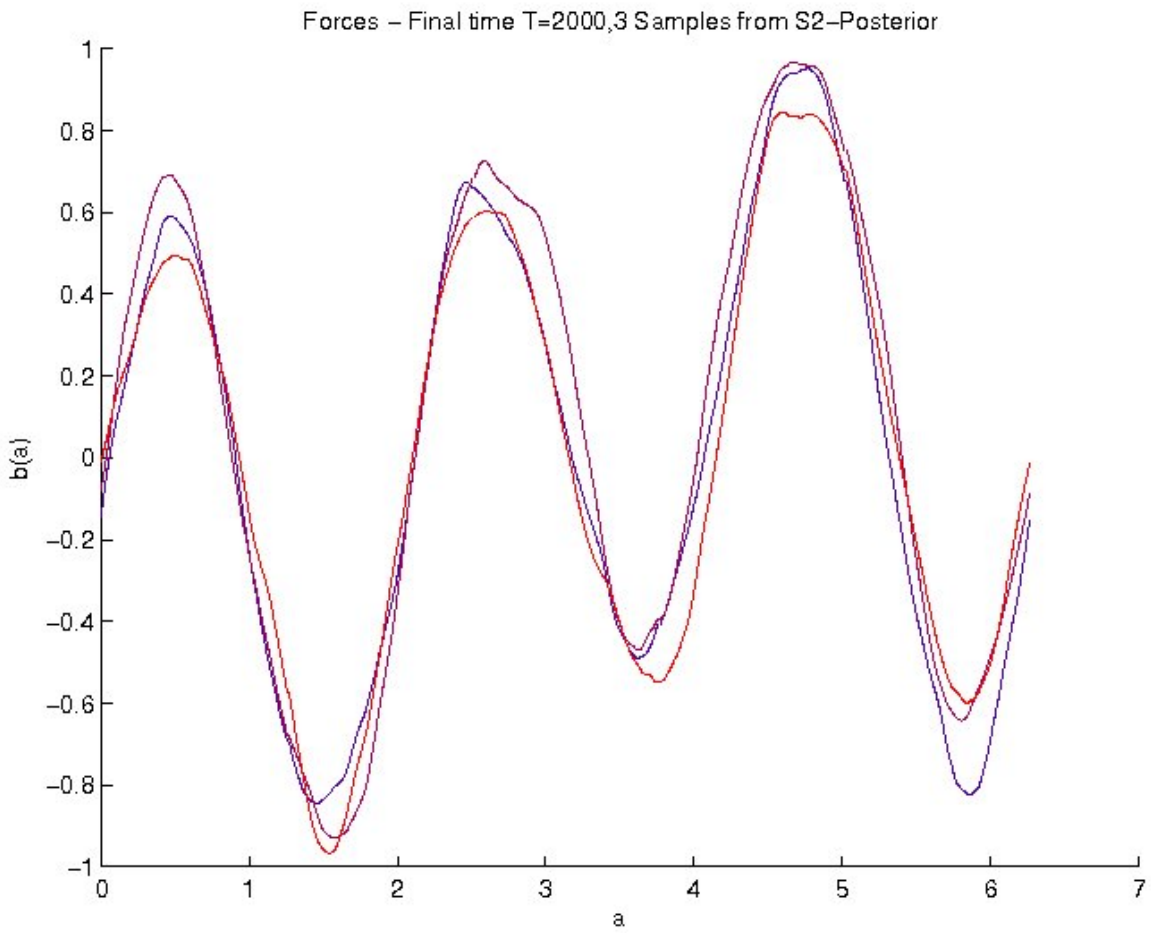


- Fourth order elliptic PDE with non-regular right hand side.
- Use piecewise cubic polynomial base functions on each finite element.

$$b(a) = \sum_{e=1}^{K} \sum_{f=1}^{4} B_{e,f} \phi_{e,f}(a)$$

Numerics: Samples from Posterior

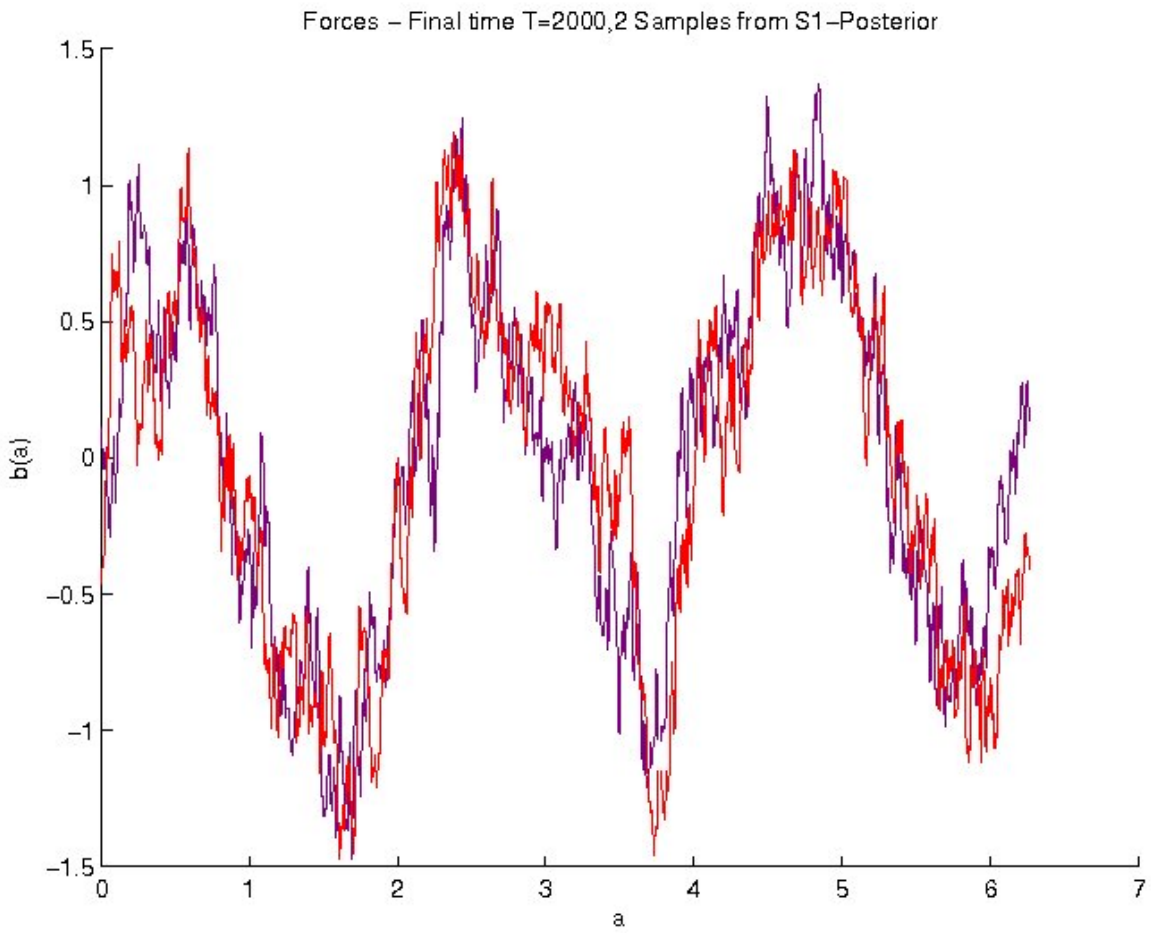
$$dx = -sin(x) + 3cos^2(x)sin(x)dt + dB$$



Second order prior covariance: $\mathcal{A} = \Delta^{-2}$

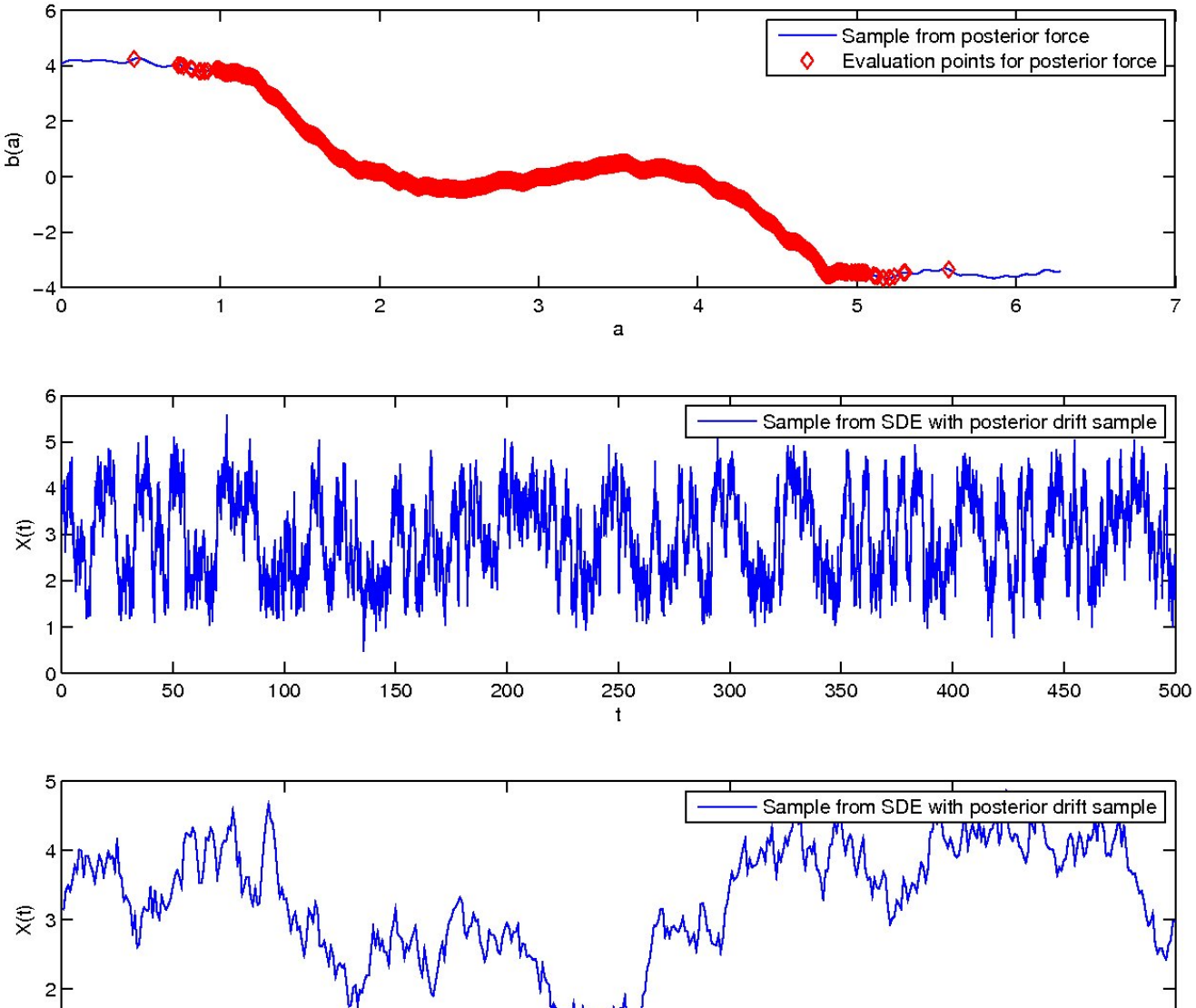
Numerics: Samples from Posterior

$$dx = -sin(x) + 3cos^2(x)sin(x)dt + dB$$



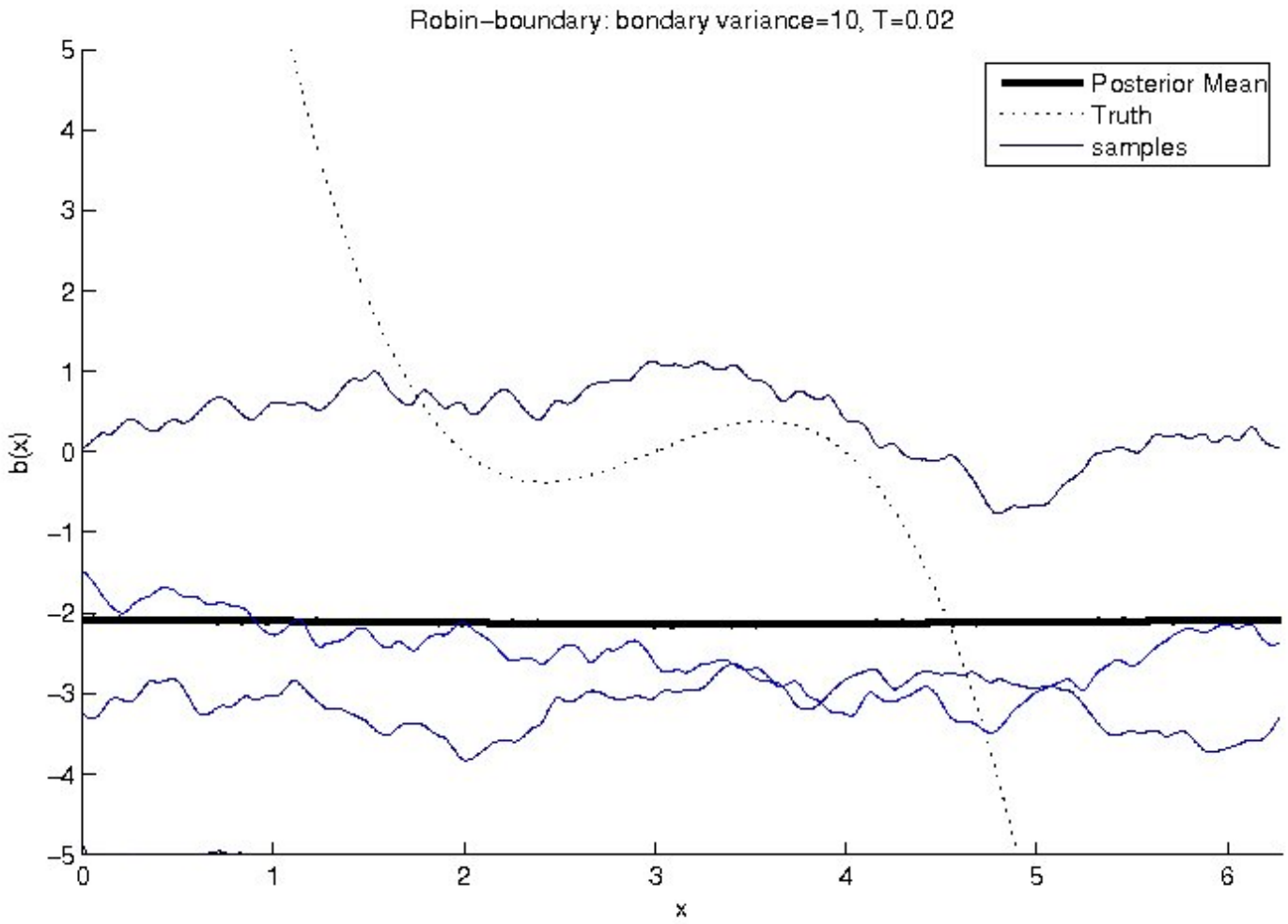
First order prior covariance: $\mathcal{A} = \Delta^{-1}$

Samples from the Posterior are usable



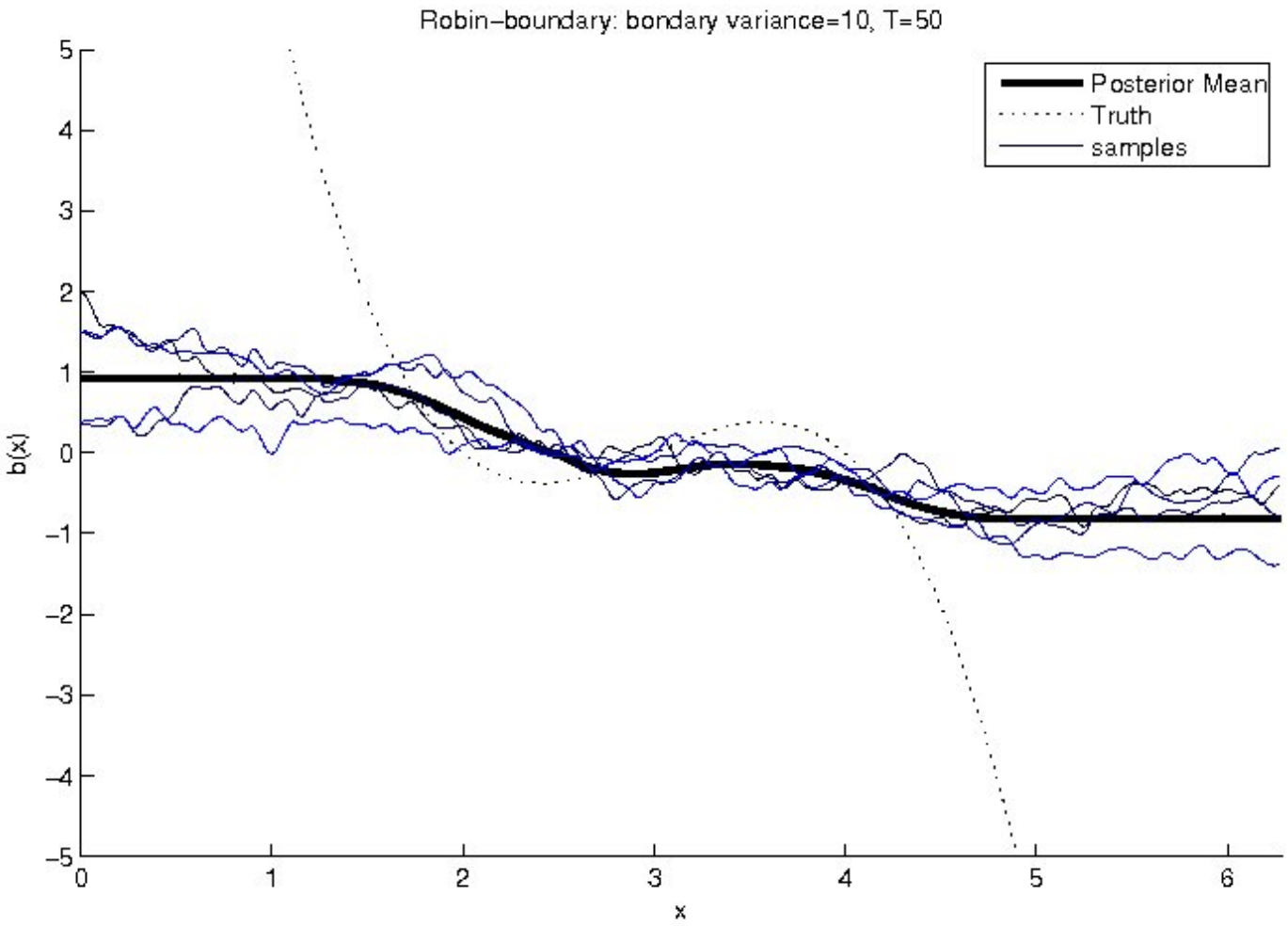
Convergence as $T \to \infty$

Gaussian boundary conditions with second order covariance operator. $T = 0.02$



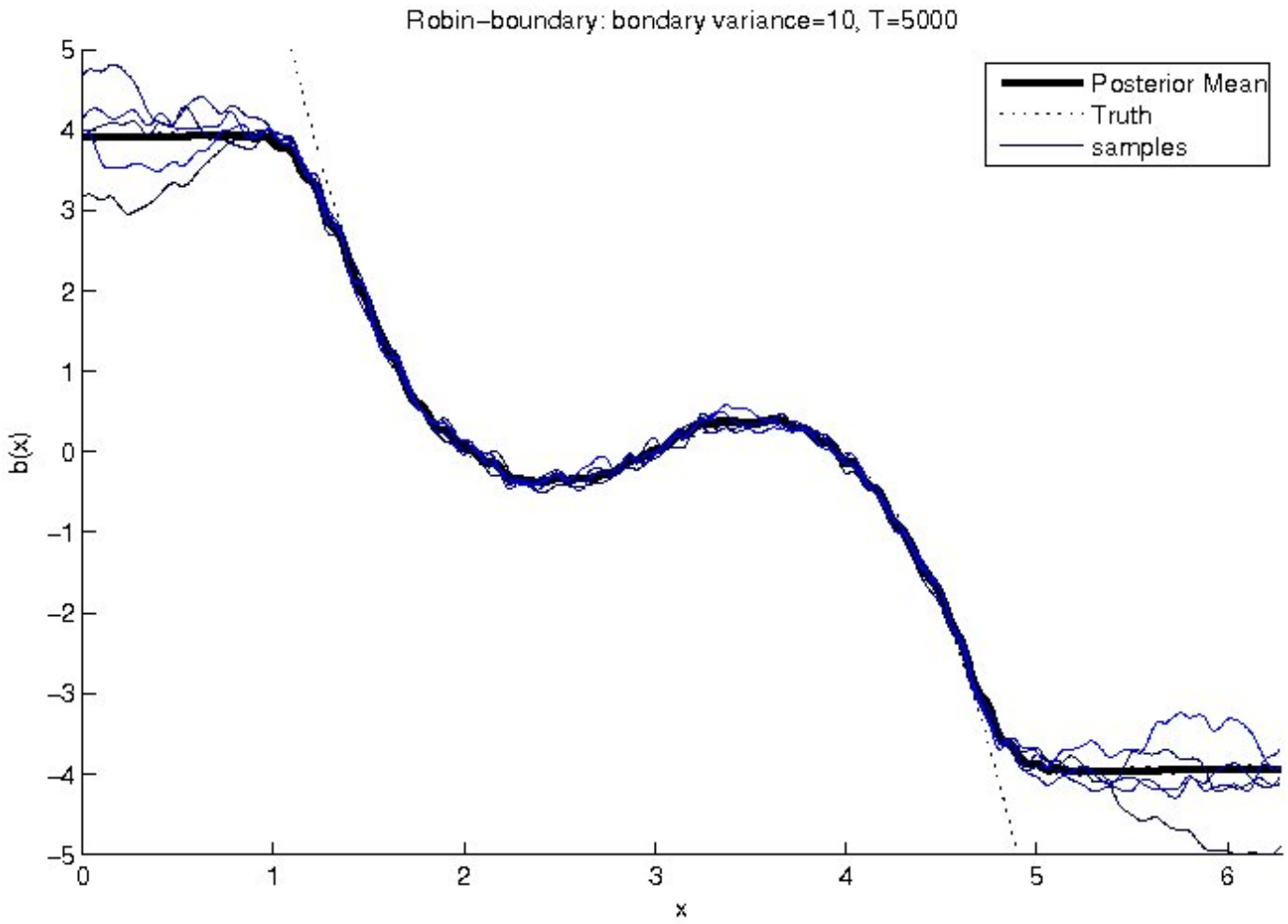
Convergence as $T \to \infty$

Gaussian boundary conditions with second order covariance operator. $T = 50$



Convergence as $T \to \infty$

Gaussian boundary conditions with second order covariance operator. $T = 5000$



Rates of posterior contraction

For $Z_1 = \hat{b}(0.38\pi) - b(0.38\pi)$ and $Z_2 = \int_0^{2\pi} \hat{b}(a)\sin(a)da$.
Questions:

- Do we have a law of large numbers $Z_i \to 0$ as $T \to \infty$?
- Do we get CLT-like convergence? $\mathrm{Var}(Z_i) = \mathcal{O}\left(\frac{1}{T}\right)$

(Numerical) Answers:

- Numerically, $\lim_{T\to\infty} Z_i = 0$ is observed.
- Decay of Variance: Answer depends on i!
  High frequency components of L' can dominate the convergence.

Rates of posterior contraction

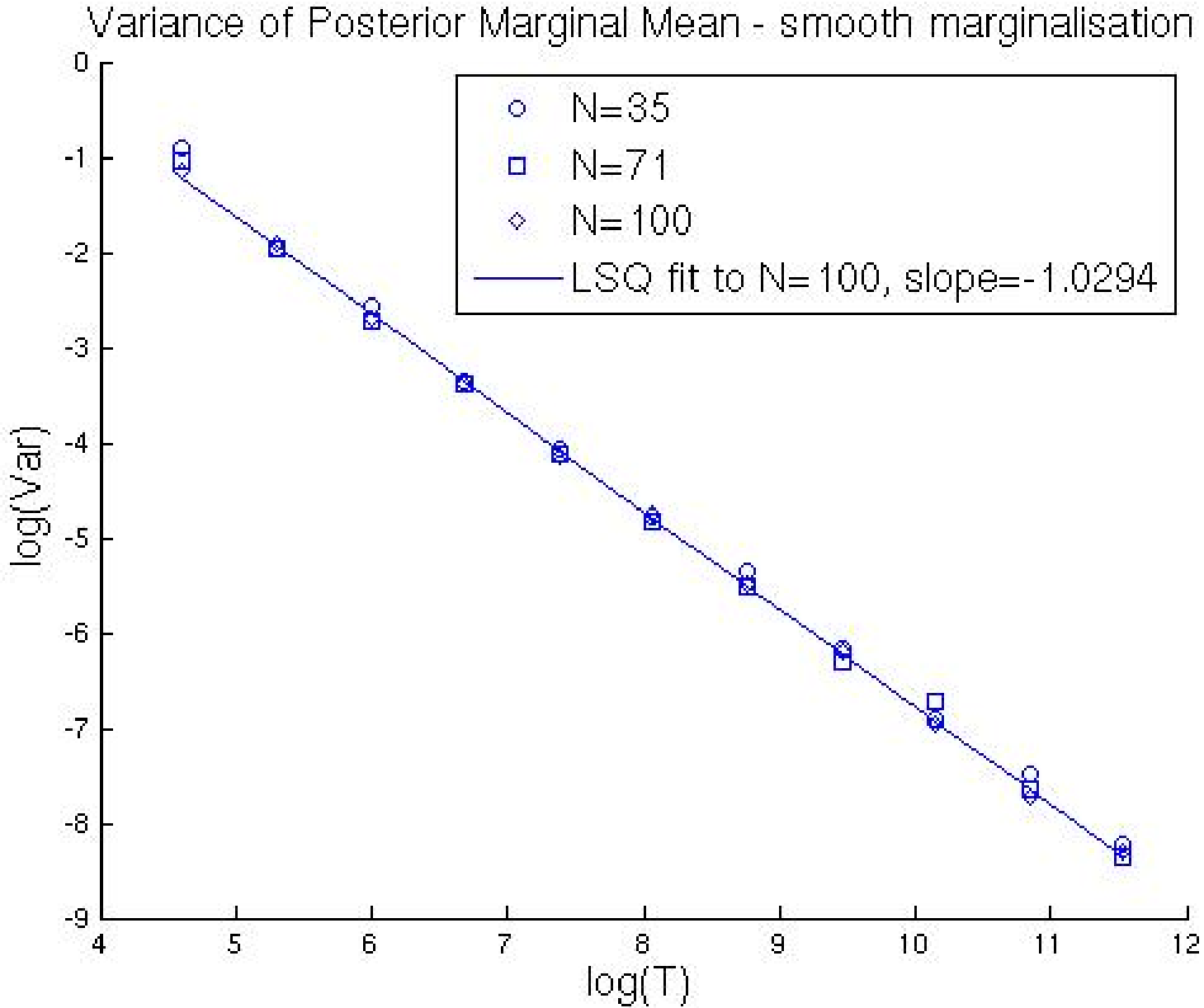
For $Z_1 = \hat{b}(0.38\pi) - b(0.38\pi)$ and $Z_2 = \int_0^{2\pi} \hat{b}(a)\sin(a)da$.
Questions:

- Do we have a law of large numbers $Z_i \to 0$ as $T \to \infty$?
- Do we get CLT-like convergence? $\mathrm{Var}(Z_i) = \mathcal{O}\left(\frac{1}{T}\right)$

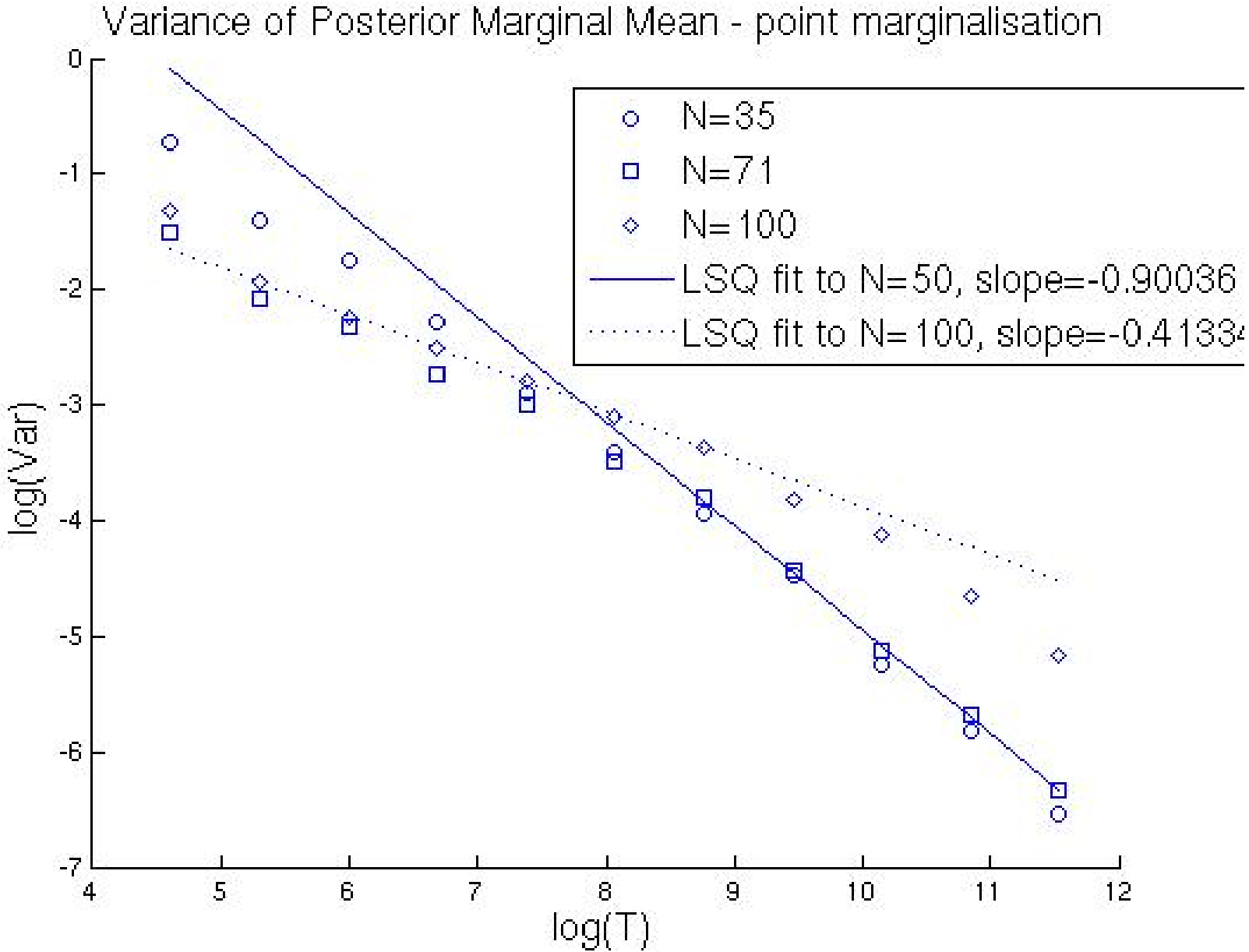
(Numerical) Answers:

- Numerically, $\lim_{T\to\infty} Z_i = 0$ is observed.
- Decay of Variance: Answer depends on i!
  High frequency components of L' can dominate the convergence.

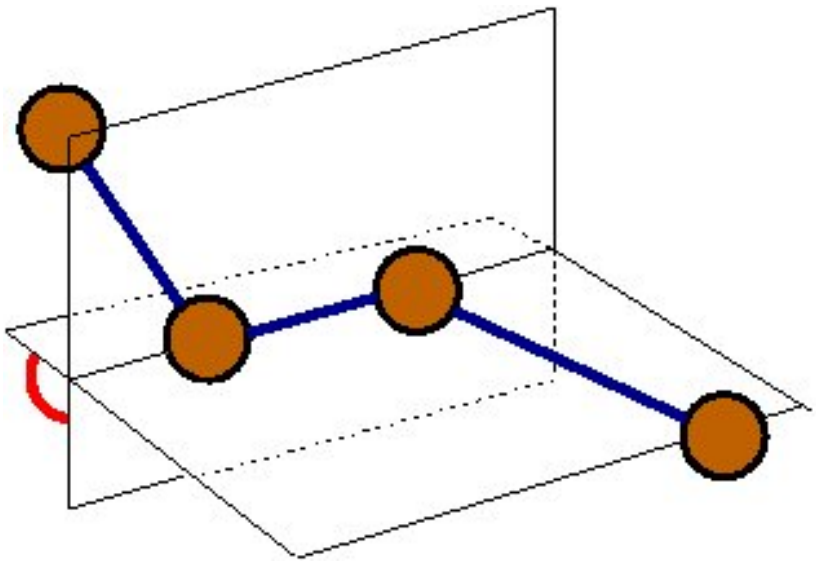
Rate of Posterior Contraction – Smooth Functional

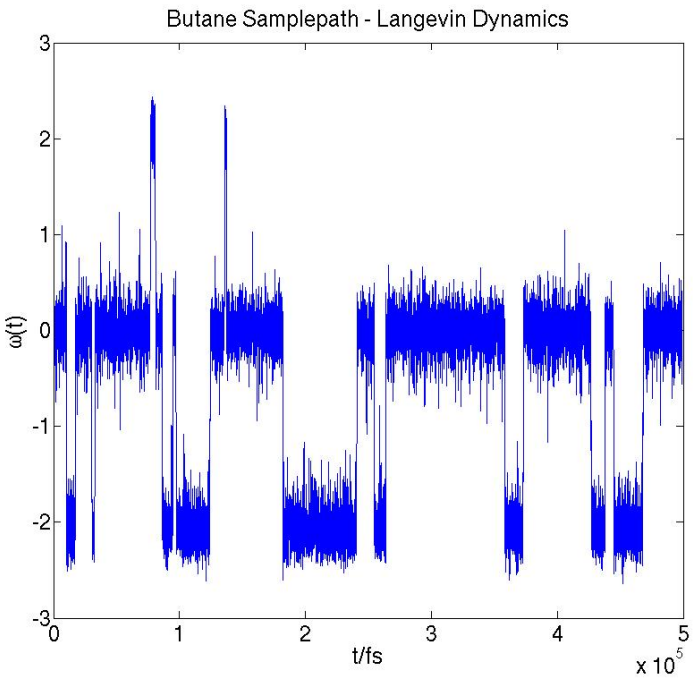


Rate of Posterior Contraction – Point Evaluation



Molecular Dynamics

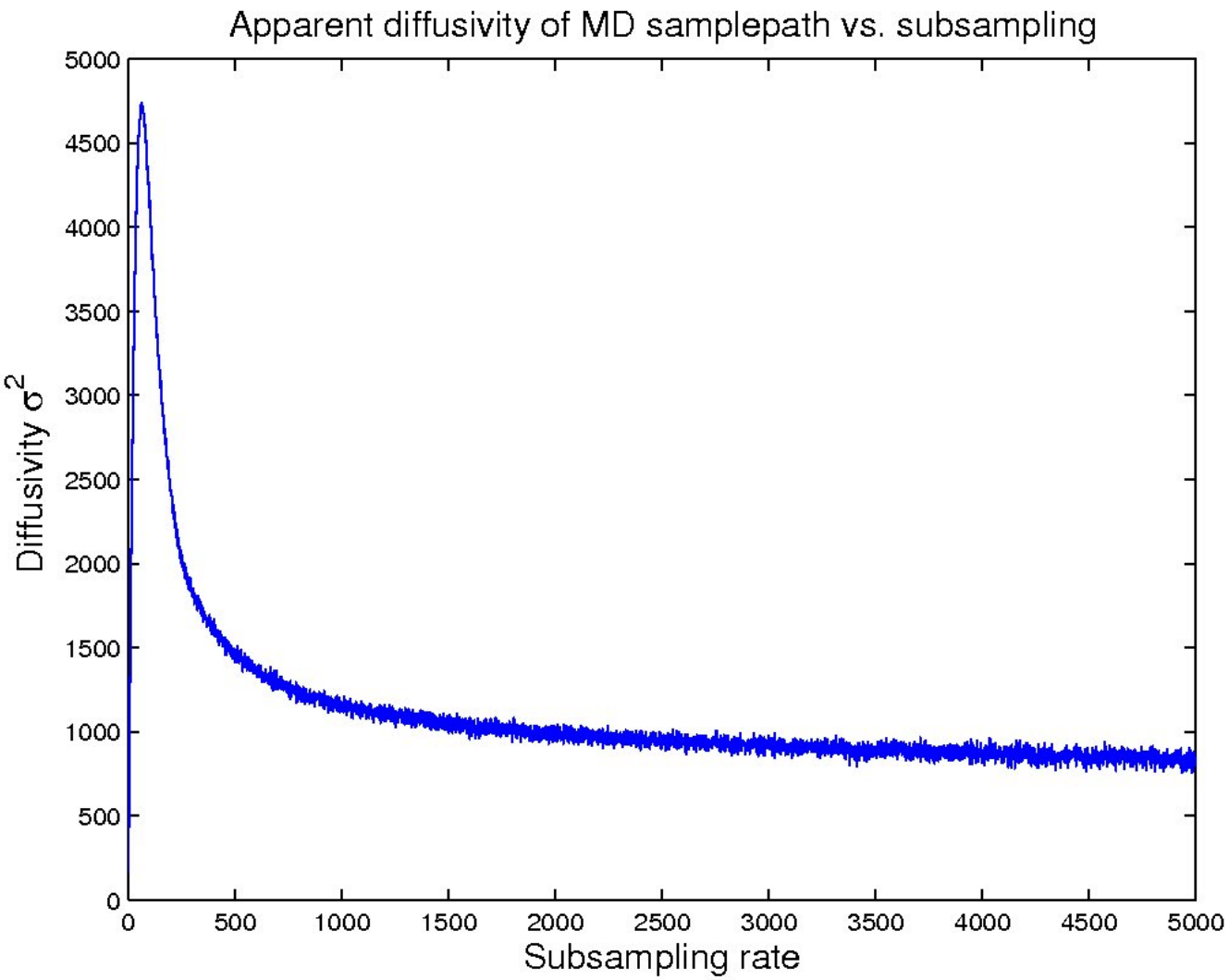




$$
\begin{aligned}
M\ddot{X}(t) &= -\nabla V(X(t)) - \gamma M\dot{X}(t) + \sqrt{2\gamma k_B T M}\dot{B} \\
X &\mapsto \omega(X)
\end{aligned}
$$

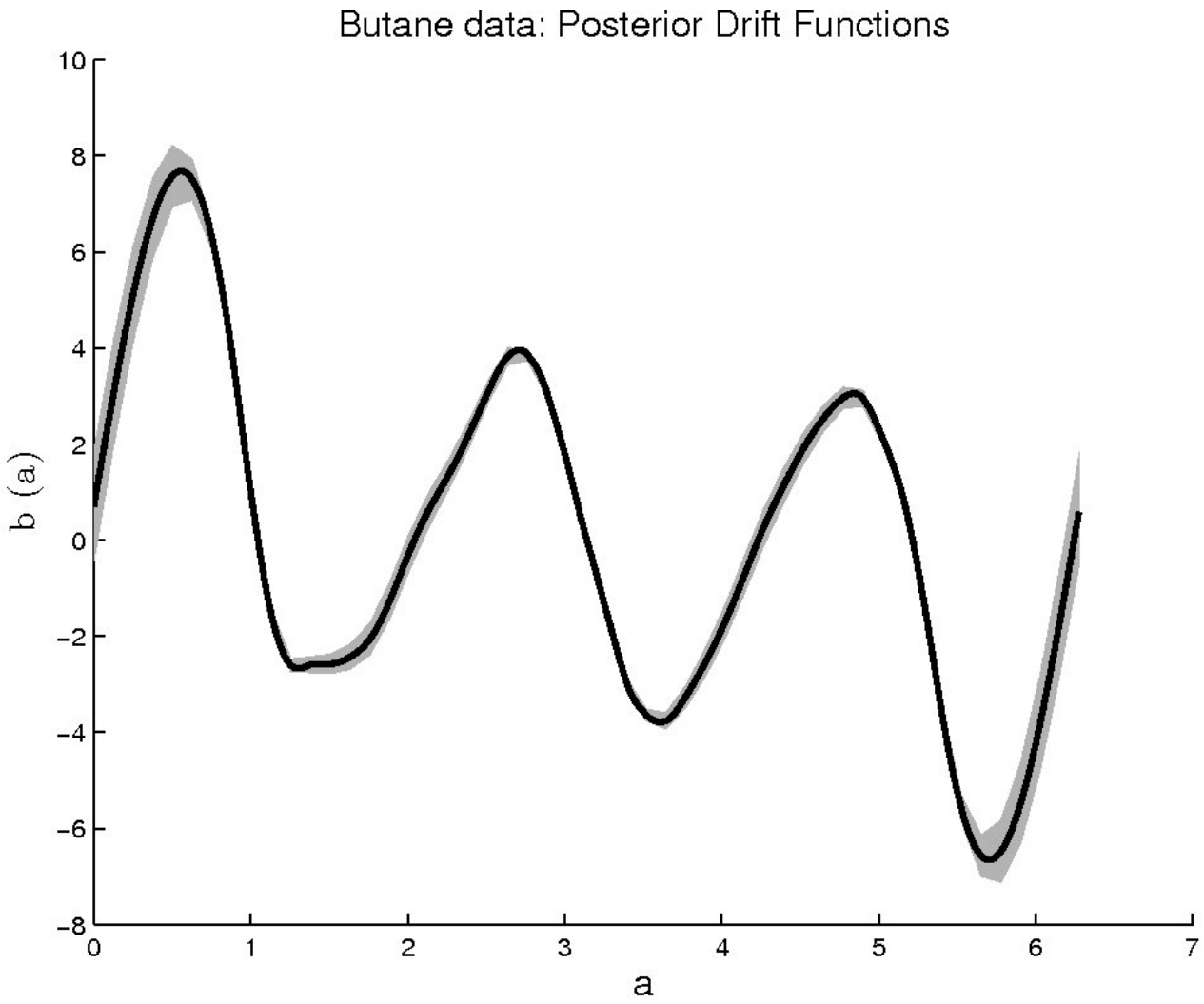
Fitting Result

Whether data looks like a diffusion depends on *timescale*.



Fitting Result

Posterior mean and standard deviation band for $k = 1000$:



Extensions

Numerically, the method also works for

- second order covariance operators
- Gaussian boundary conditions
- the whole real line

Future work:

- Extension to higher dimensions (2,3)
- Low-frequency data by sampling from missing local time.
- Convergence properties for various test functionals (singular limit problem in PDE theory)

Summary

- Nonparametric drift estimation for diffusions on the circle can be performed rigorously for Gaussian prior (conjugate prior).
- Finite element implementation enables error control from discrete time high frequency samples all the way to numerically obtained posterior means.
- Applications are in molecular dynamics and other areas where data with multiple timescales pose problems.