





# Posterior Contraction Rates for Bayesian Inverse Problems

Sergios Agapiou

Mathematics Institute  
University of Warwick, UK

Stochastic Analysis Seminar,  
21st January 2013, University of Oxford

<http://homepages.warwick.ac.uk/~mariba/>

-  S. Agapiou, S. Larsson and A. M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Under revision Stoch. Proc. Appl., <http://arxiv.org/abs/1203.5753>
-  Y. Pokern, A. M. Stuart and J. H. Van Zanten, *Posterior consistency via precision operators for nonparametric drift estimation in SDEs*, Accepted Stoch. Proc. Appl., <http://arxiv.org/abs/1202.0976>
-  S. Agapiou, A. M. Stuart and Y. X. Zhang, *An abstract framework for the study of posterior contraction in Bayesian inverse problems*, in preparation.
-  S. Agapiou, A. M. Stuart and Y. X. Zhang, *Bayesian posterior contraction rates for linear severely ill-posed inverse problems*, <http://arxiv.org/abs/1210.1563>

# Outline

- 1 Introduction
- 2 Linear Inverse Problem
- 3 Nonparametric Drift Estimation in SDEs
- 4 Abstract Theory - Conclusions

# Outline

- 1 Introduction
- 2 Linear Inverse Problem
- 3 Nonparametric Drift Estimation in SDEs
- 4 Abstract Theory - Conclusions

# Bayesian Inverse Problems

- $(X, \langle \cdot, \cdot \rangle, \|\cdot\|)$  separable Hilbert space.
- Probabilistic approach to problem of recovering  $u$  from noisy, indirect observations,  $y$ .
- *Likelihood*: distribution of  $y|u$ .
- *Prior*:  $u \sim \mu_0$ , encoding prior beliefs.
- *Posterior*:  $u|y \sim \mu^y$ , object of interest.
- Link: *Bayes' theorem*

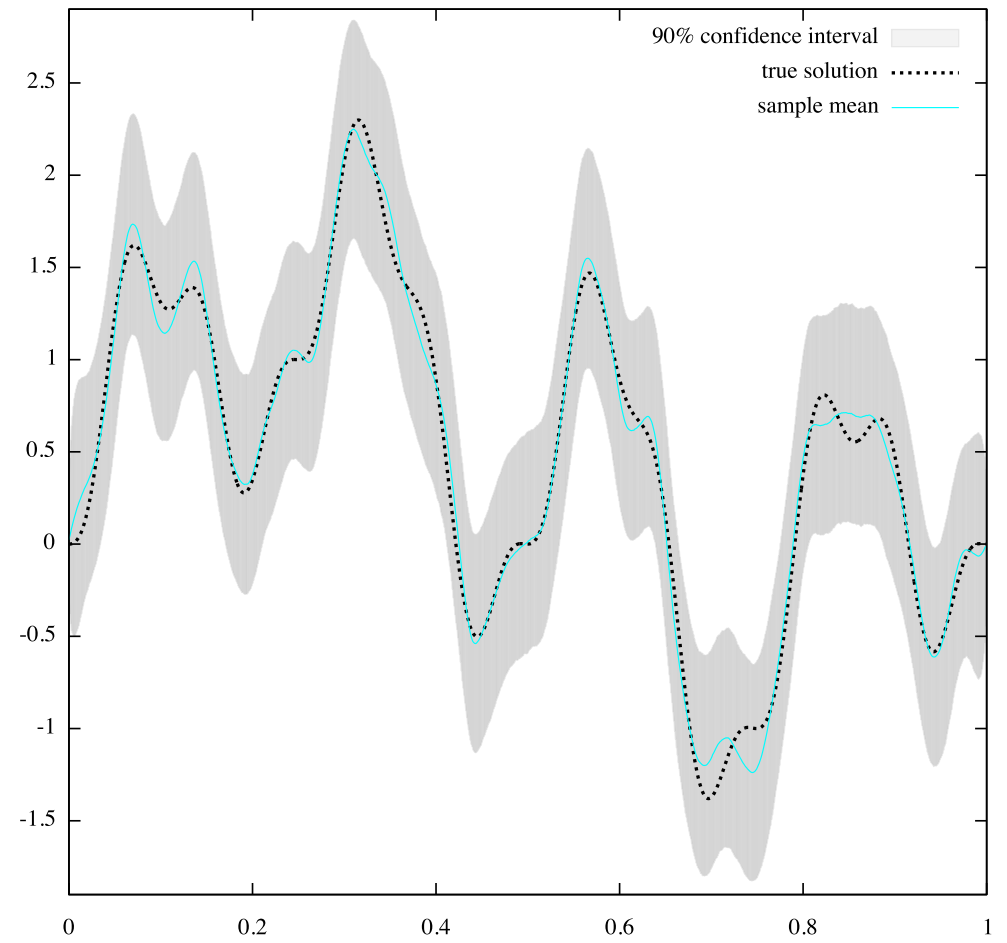
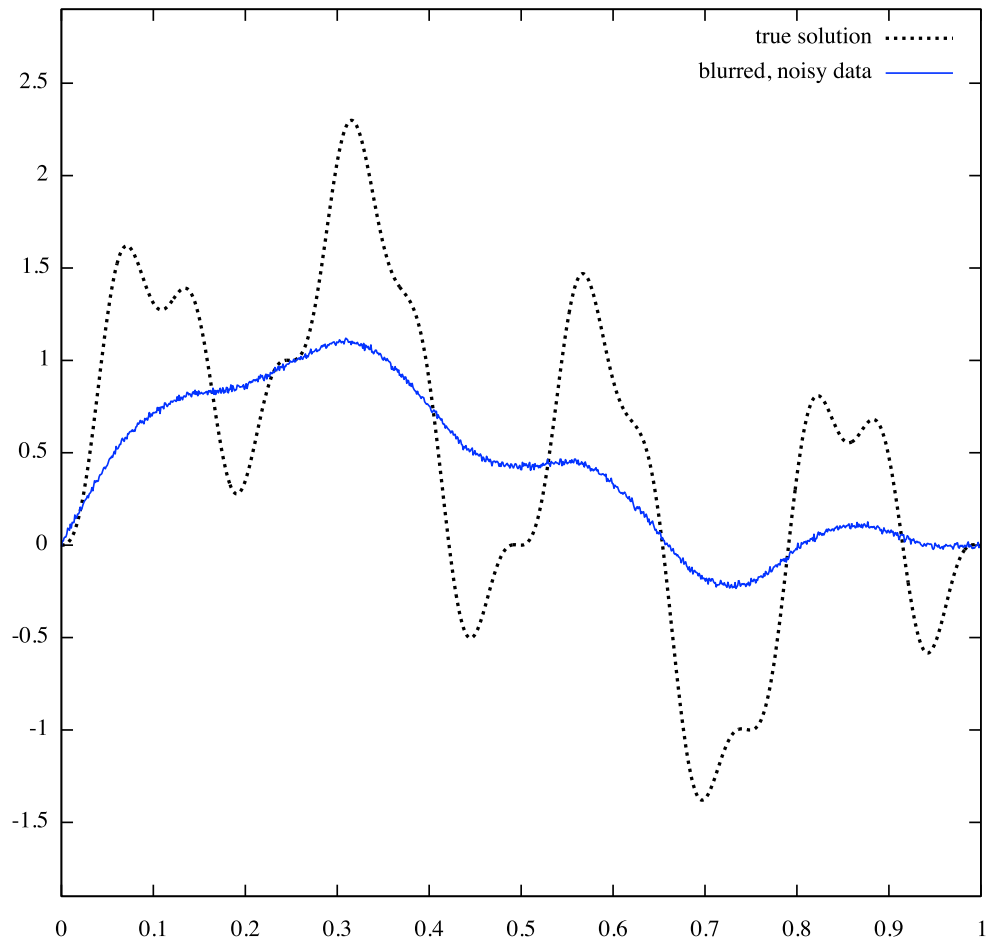
$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mathbb{P}(u)$$

# Frequentist Posterior Consistency

- Here prior  $\mu_0 = \mathcal{N}(0, \tau^2 \mathcal{C}_0)$ . Assume posterior is Gaussian  $\mu^y = \mathcal{N}(m, \mathcal{C})$ .
- Consider data  $y^\dagger$  produced from an underlying truth  $u^\dagger$ .
- **Posterior Consistency**: as more data become available, can we recover the truth?
- $y^\dagger = y^\dagger(\epsilon)$  and  $\mu^y = \mu^y(\epsilon)$ ,  $\epsilon \rightarrow 0$  models improvement in the data (and model).

AIM: Show  $\mu^{y^\dagger}(\epsilon) \rightarrow \delta_{u^\dagger}$ , as  $\epsilon \rightarrow 0$  for appropriate choice  $\tau = \tau(\epsilon)$ .

# Frequentist Posterior Consistency



# Gaussian Measures in Separable Hilbert Spaces

- $\mathcal{N}(m, \mathcal{C})$  in  $X$ 
  - mean  $m \in X$ ;
  - covariance  $\mathcal{C}$  selfadjoint trace class linear operator in  $X$ , eigenpairs  $\{\phi_k, \lambda_k\}_{k \in \mathbb{N}}$ .

- Karhunen-Loeve expansion:  $u \sim \mathcal{N}(m, \mathcal{C})$

$$u = m + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k \phi_k,$$

$\{\xi_k\}_{k \in \mathbb{N}}$  i.i.d.  $\mathcal{N}(0, 1)$  in  $\mathbb{R}$ .

- Decay of  $\lambda_k$  determines regularity of  $u$ .



# Outline

- 1 Introduction
- 2 Linear Inverse Problem**
- 3 Nonparametric Drift Estimation in SDEs
- 4 Abstract Theory - Conclusions

# Linear Inverse Problem

- $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset X \rightarrow X$  selfadjoint positive definite linear operator with bounded inverse.
- Find  $u$  from noisy observation of  $\mathcal{A}^{-1}u, y$ .

- Model:

$$y = \mathcal{A}^{-1}u + \frac{1}{\sqrt{n}}\xi,$$

- Interested in *small noise limit*,  $n \rightarrow \infty$ .
- Naive approach  $u \approx \mathcal{A}y$ ,  $\xi$  is rough, need to **regularize**.

# Linear Inverse Problem - Deterministic Approach

- **Tikhonov-Phillips Regularization:**  $u$  approximated by minimizer of

$$J(u) := \frac{1}{2} \left\| \mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u) \right\|^2 + \frac{\lambda}{2} \left\| \mathcal{C}_0^{-\frac{1}{2}}u \right\|^2,$$

$\mathcal{C}_i : \mathcal{X} \rightarrow \mathcal{X}$ , self-adjoint, possibly compact, positive definite linear operators.

- $\lambda$  **regularization parameter**, function of noise level  $n^{-\frac{1}{2}}$  going to zero as  $n \rightarrow \infty$  to recover unknown  $u$ .

# Linear Inverse Problem - Bayesian Approach

Assume  $\xi \sim \mathcal{N}(0, \mathcal{C}_1)$ ,  $\mathcal{C}_1 : X \rightarrow X$  selfadjoint positive definite.

- *Likelihood*:  $y|u \sim \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$ .
- *Prior*:  $u \sim \mathcal{N}(0, \tau^2\mathcal{C}_0)$ ,  $\mathcal{C}_0 : X \rightarrow X$  selfadjoint positive definite trace class.
- *Posterior* distribution on  $u|y, \mu^y$ .

# Linear Inverse Problem - Assumptions

- *Hilbert Scale*  $(X^s)_{s \in \mathbb{R}}$ , for  $X^s = \mathcal{D}(\mathcal{C}_0^{-\frac{s}{2}})$  with  $\langle u, v \rangle_s = \langle \mathcal{C}_0^{-\frac{s}{2}} u, \mathcal{C}_0^{-\frac{s}{2}} v \rangle$ .

- Assumptions

$$\exists s_0 \in [0, 1) \text{ s.t. } \text{Tr}(\mathcal{C}_0^s) < \infty \quad \forall s > s_0;$$

$$\mathcal{C}_1 \simeq \mathcal{C}_0^\beta, \quad \beta \geq 0;$$

$$\mathcal{A}^{-1} \simeq \mathcal{C}_0^\ell, \quad \ell \geq 0.$$

- $\Delta := 1 + 2\ell - \beta > 2s_0$ . For simplicity  $2\ell - \beta \geq 0$ .

# Linear Inverse Problem - Posterior Identification

Theorem (A., Larsson, Stuart 12)

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi_n(u; y))$$

where

$$\Phi_n(u; y) = \frac{n}{2} \left\| \mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u) \right\|^2 - \frac{n}{2} \underbrace{\left\| \mathcal{C}_1^{-\frac{1}{2}}y \right\|^2}_{\text{}}.$$

In particular,  $\mu^y = \mathcal{N}(m, \frac{1}{n}\mathcal{B}_\lambda^{-1})$

$$\mathcal{B}_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}$$

$$\mathcal{B}_\lambda m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y,$$

$\lambda = \frac{1}{n\tau^2}$  regularization parameter.

(Lax Milgram)

# Linear Inverse Problem - Posterior Consistency

- Consider

$$y^\dagger = \mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\xi,$$

$u^\dagger \in X^\gamma$  fixed true solution,  $\gamma$  a priori known.

- Posterior  $\mu^{y=y^\dagger} = \mathcal{N}(m^\dagger, \frac{1}{n}\mathcal{B}_\lambda^{-1})$ .

- AIM: Find optimal **rate** = **rate**( $\gamma$ ), such that, as  $n \rightarrow \infty$ ,

$$\mathbb{E} \left\| m^\dagger - u^\dagger \right\|^2 + \text{Tr} \left( \frac{1}{n} \mathcal{B}_\lambda^{-1} \right) = \mathcal{O}(n^{-\text{rate}}).$$

# Linear Inverse Problem - Error Equation

- Mean

$$\mathcal{B}_\lambda m^\dagger = \underbrace{\mathcal{A}^{-1} \mathcal{C}_1^{-1} \mathcal{A}^{-1} u^\dagger}_{\text{}} + \frac{1}{\sqrt{n}} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi.$$

- Truth

$$\mathcal{B}_\lambda u^\dagger = \underbrace{\mathcal{A}^{-1} \mathcal{C}_1^{-1} \mathcal{A}^{-1} u^\dagger}_{\text{}} + \lambda \mathcal{C}_0^{-1} u^\dagger.$$

- Error  $e = m^\dagger - u^\dagger$

$$e = \mathcal{B}_\lambda^{-1} \left( \frac{1}{\sqrt{n}} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi - \lambda \mathcal{C}_0^{-1} u^\dagger \right).$$



# Linear Inverse Problem - Error Analysis

## Lemma (A., Larsson, Stuart 12)

For  $\eta = (1 - \theta)(\beta - 2\ell) + \theta$ ,  $\theta \in [0, 1]$

$$\|\mathcal{B}_\lambda^{-1}\|_{\mathcal{L}(X^{-\eta}, X)} = \mathcal{O}(\lambda^{-c}),$$

$c = c(\theta, \beta - 2\ell) \in (0, 1)$ , increasing in  $\theta$ .

- For  $\eta_i = (1 - \theta_i)(\beta - 2\ell) + \theta_i$ , where  $\theta_i \in [0, 1]$  sufficiently large,

$$\mathbb{E} \left\| \frac{1}{\sqrt{n}} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi \right\|_{-\eta_1} = \mathcal{O}(n^{-\frac{1}{2}})$$

and

$$\|\lambda \mathcal{C}_0^{-1} u^\dagger\|_{-\eta_2} = \mathcal{O}(\lambda).$$

$$\mathbb{E} \|e\| = \mathcal{O}(\lambda^{-c_1} n^{-\frac{1}{2}}) + \mathcal{O}(\lambda^{1-c_2}).$$

# Linear Inverse Problem - Convergence

- Choose  $\lambda = \lambda(n)$  (hence  $\tau = \tau(n)$ ) optimally.
- $\text{Tr}(\frac{1}{n}\mathcal{B}_\lambda^{-1})$  bounded by noise term.

## Theorem (A., Larsson, Stuart 12)

Assume  $u^\dagger \in X^\gamma$ ,  $\gamma \geq 1$ . For appropriate choice of  $\lambda = \lambda(n) \rightarrow 0$  we have as  $n \rightarrow \infty$

$$\text{rate} = \begin{cases} \frac{\gamma}{2(\Delta + \gamma - 1 + s_0 + \delta)}, & \text{if } \gamma \in [1, \Delta + 1] \\ \frac{\Delta + 1}{2(2\Delta + s_0 + \delta)}, & \text{if } \gamma > \Delta + 1, \end{cases}$$

$\delta > 0$  arbitrarily small.

# Linear Inverse Problem - Optimality, Diagonal Case

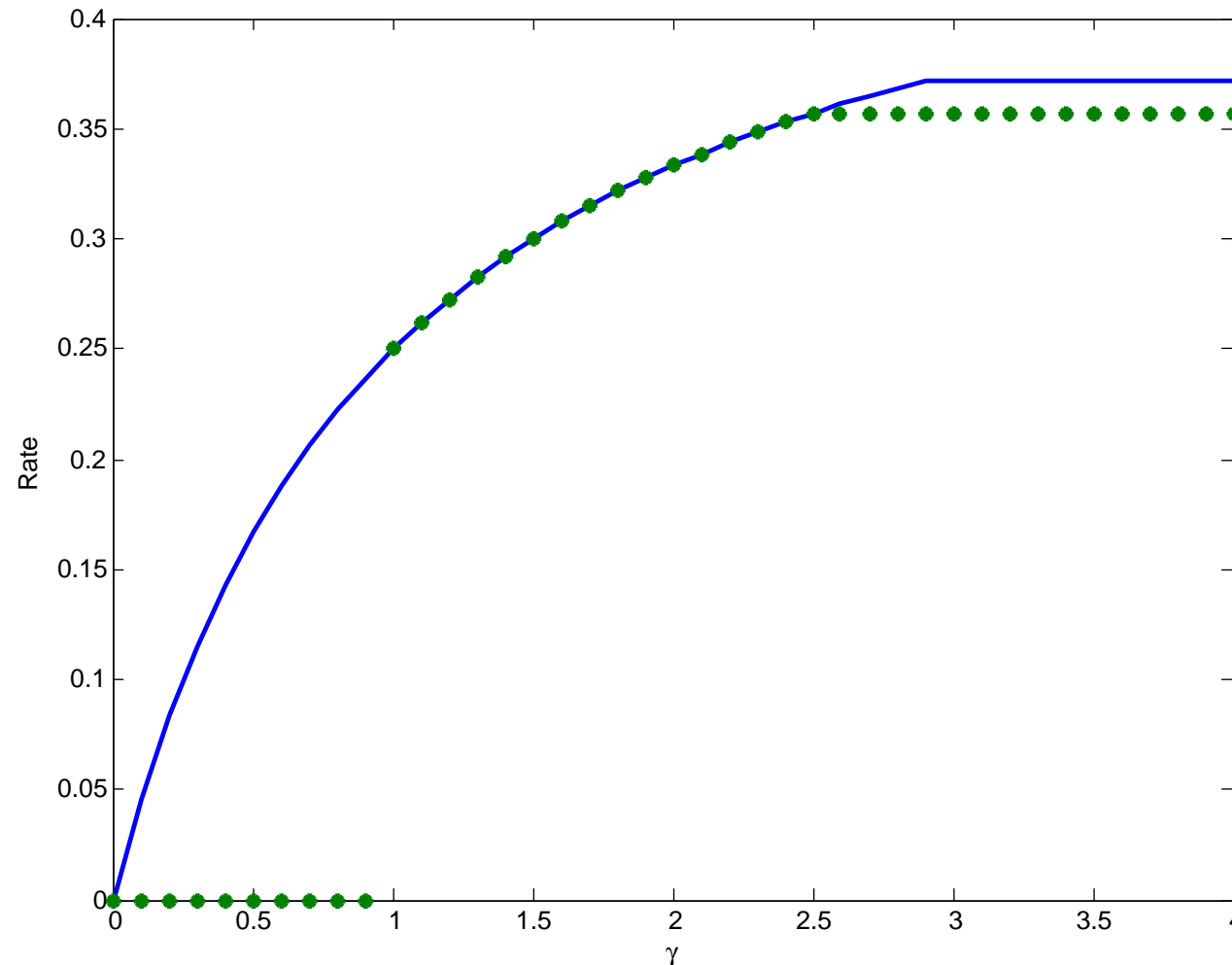
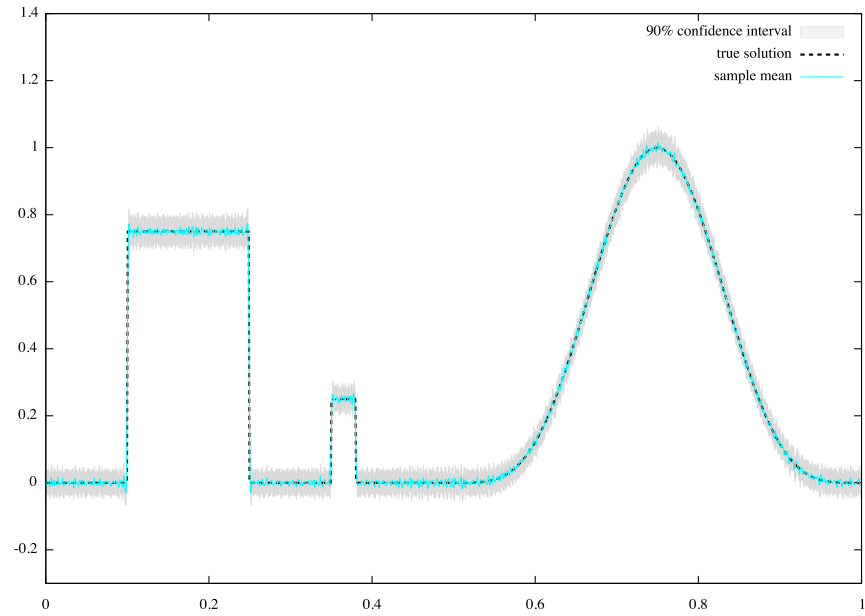
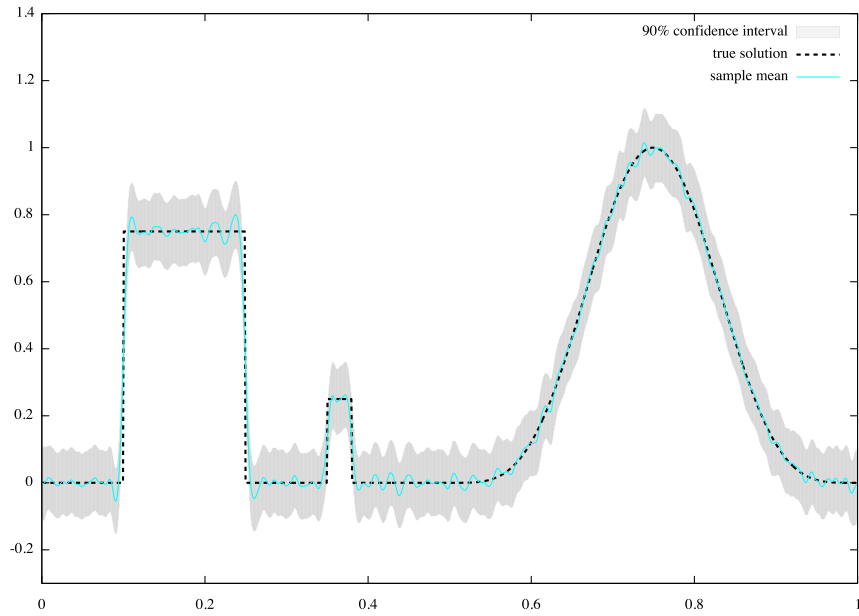
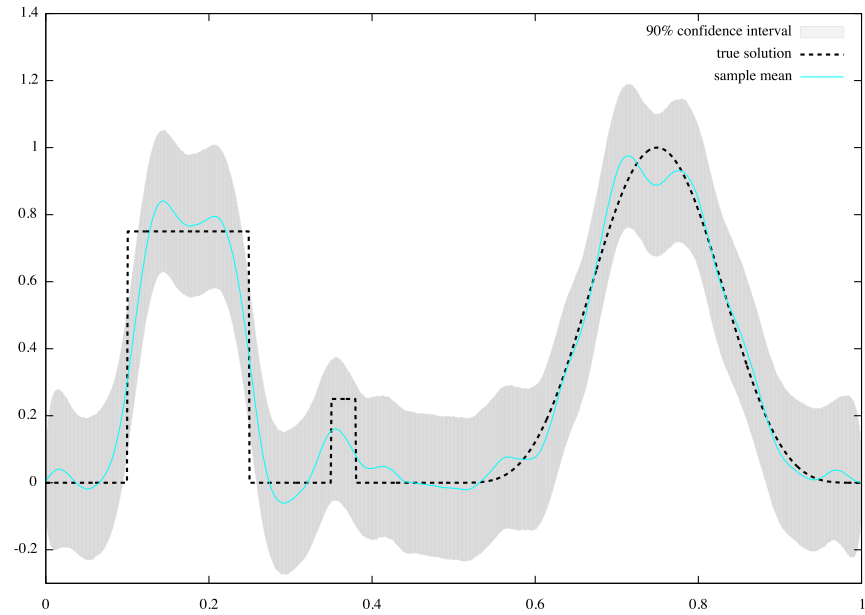
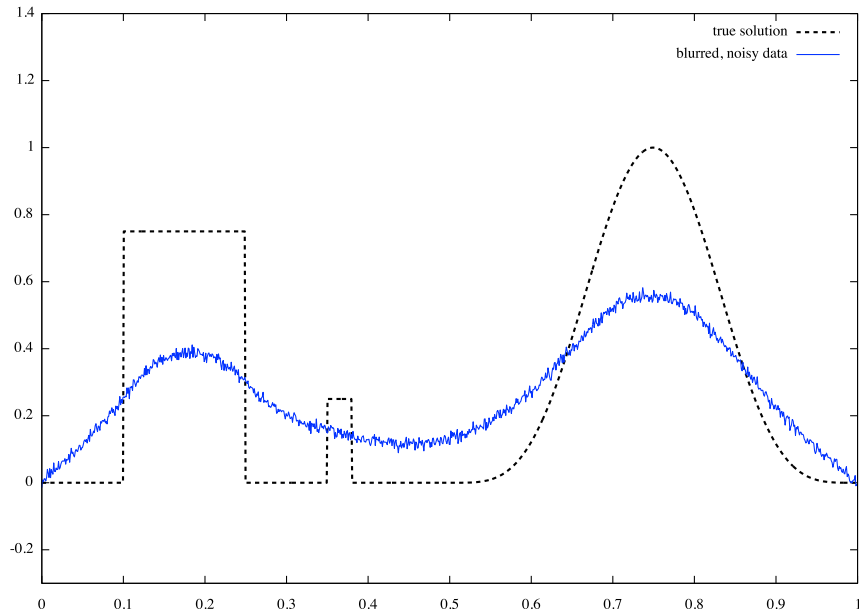


Figure: blue=diagonal analysis and green=our more general analysis for  $C_0^{-1} = -\frac{d^2}{dx^2}$ ,  $\beta = \ell = \frac{1}{2}$ , so that  $s_0 = \frac{1}{2}$ ,  $\Delta = \frac{3}{2}$  and  $\text{rate} = \frac{\gamma}{2+2\gamma+\delta}$ ,  $\gamma \in [1, 2.5]$ .

# Linear Inverse Problem - Posterior Consistency



# Outline

- 1 Introduction
- 2 Linear Inverse Problem
- 3 Nonparametric Drift Estimation in SDEs**
- 4 Abstract Theory - Conclusions

# Nonparametric Drift Estimation in SDEs

- Consider the SDE

$$dY_t = u(Y_t)dt + dW_t, \quad Y_0 = 0,$$

where  $u \in \dot{C}^1(\mathbb{T})$ ,  $\mathbb{T} = [0, 1)$  periodic.

- Find drift  $u$  from observation  $Y = \{Y_t\}_{t \in [0, T]}$ .
- Interested in the *long observation time limit*,  $T \rightarrow \infty$ .

# Nonparametric Drift Estimation in SDEs - Intuition

- Work with periodized version of  $Y$ , also denoted  $Y$ .
- For  $t \rightarrow \infty$ ,  $Y_t$  reaches equilibrium in form of *invariant distribution* with density

$$\rho(x) = c \exp \left( 2 \int_0^x u(y) dy \right), \quad x \in [0, 1].$$

- Can recover  $u$  from  $\rho$

$$u = \frac{\rho'}{2\rho}.$$

# Nonparametric Drift Estimation in SDEs - Local Time

- *Local time*  $L_T(x; Y)$  measures time that  $Y$  spends around  $x$

$$L_T(x; Y) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^T 1\{Y_s \in (x - \varepsilon, x + \varepsilon)\} ds.$$

- Would like to write

$$"u \approx \frac{L'_T}{2L_T} "$$

but  $L_T$  not differentiable.



# Nonparametric Drift Estimation in SDEs - Intuition

## Lemma (Van Zanten 12)

For every  $\alpha < 1/2$

$$\left\| \frac{1}{T} L_T(\cdot; Y) - \rho \right\|_{H^\alpha} = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right),$$

as  $T \rightarrow \infty$ .

- i.e.  $u \approx \frac{L'_T(\cdot; Y)}{2L_T(\cdot; Y)}$  but in weak spaces, [regularize](#).

# Drift Estimation in SDEs - Bayesian Approach

- *Likelihood*: law of  $Y|u$  (Girsanov)

$$\frac{d\mathbb{P}_u^T}{d\mathbb{P}_0^T}(Y) = \exp(-\Phi_T(u; Y)),$$

$$\begin{aligned}\Phi_T(u; Y) &= \frac{1}{2} \int_0^1 (L_T(x; Y)(u^2(x) + u'(x)) - 2\chi_T(x; Y)u(x)) dx \\ &= \frac{1}{2} \int_0^1 \left( L_T(x; Y)u^2(x) - 2(\chi_T(x; Y) + \frac{L'_T(x; Y)}{2})u(x) \right) dx\end{aligned}$$

# Drift Estimation in SDEs - Bayesian Approach

- *Prior:*  $u \sim \mathcal{N}(0, \tau^2 \mathcal{C}_0)$ ,

$$\mathcal{C}_0^{-1} = \left( -\frac{d^2}{dx^2} \right)^p + I,$$

$p \in 2, 3, \dots; p \geq 2$  secures  $u \in \dot{C}^1(\mathbb{T})$  a.s. (Gaussianity).

- Joint distribution of  $Y$  and  $u$  in general non-Gaussian.
- *Posterior:*  $u|Y \sim \mu^Y$  Gaussian.

# Drift Estimation in SDEs - Posterior Identification

Theorem (Pokern, Stuart, Van Zanten 12)

$$\frac{d\mu^Y}{d\mu_0}(u) = \frac{1}{Z(Y)} \exp(-\Phi_T(u; Y)).$$

In particular,  $\mu^Y = \mathcal{N}(m, \frac{1}{T}\mathcal{B}_\lambda^{-1})$

$$\mathcal{B}_\lambda = \mathcal{B}_\lambda(Y, T) = \frac{L_T(\cdot; Y)}{T} I + \lambda C_0^{-1}$$

$$\mathcal{B}_\lambda m = \frac{1}{2} \frac{L_T(\cdot; Y)'}{T} + \frac{\chi_T(\cdot; Y)}{T}$$

$\lambda = \frac{1}{\tau^2 T}$  regularization parameter.

(Lax Milgram)

# Drift Estimation in SDEs - Posterior Consistency

- Consider  $Y^\dagger = \{Y_t^\dagger\}_{t \in [0, T]}$  where

$$dY_t^\dagger = u^\dagger(Y_t^\dagger)dt + dW_t$$

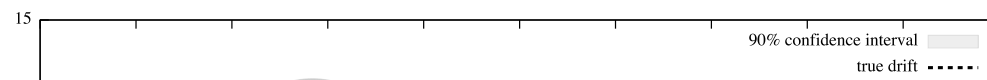
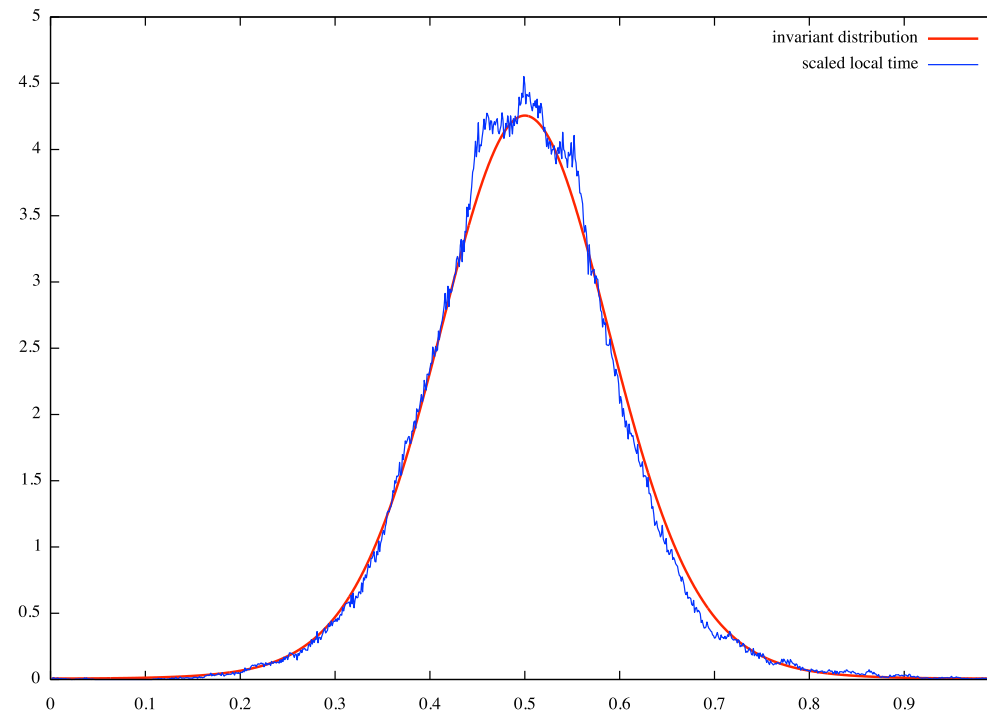
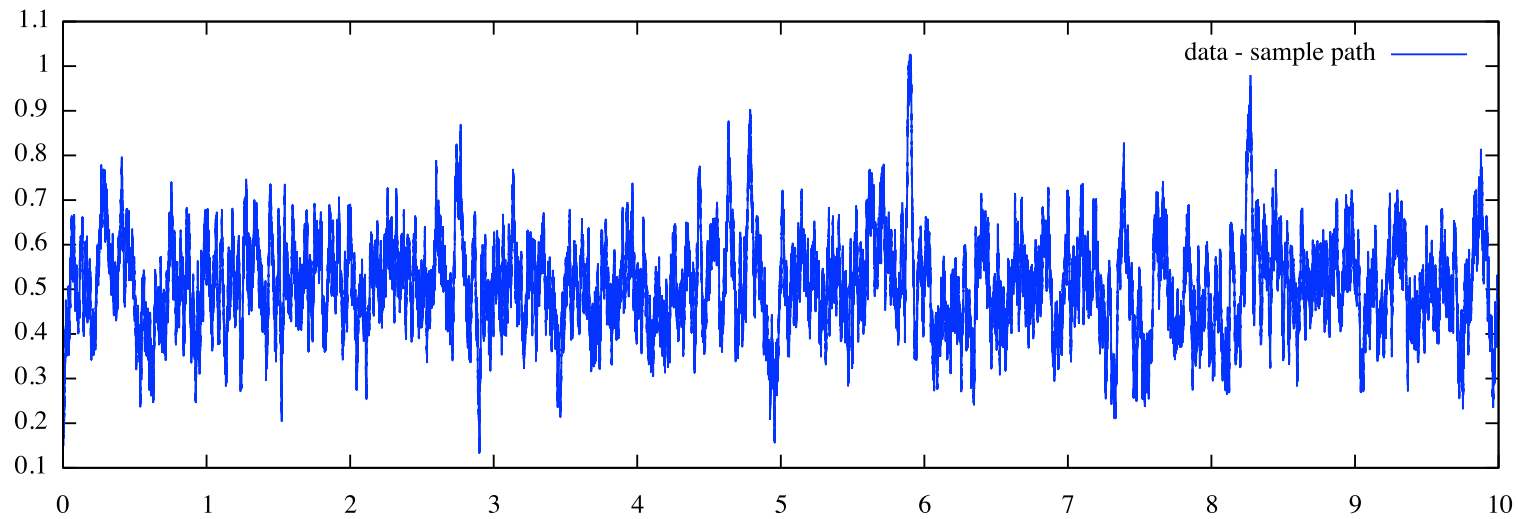
$u^\dagger \in H^\gamma$  fixed true drift,  $\gamma$  a priori known.

- Posterior  $\mu^{Y=Y^\dagger} = \mathcal{N}(m^\dagger, \frac{1}{T}\mathcal{B}_\lambda^{-1})$ , where  $\mathcal{B}_\lambda = \mathcal{B}_\lambda(Y^\dagger)$ .

- AIM: Find optimal **rate** = **rate**( $\gamma$ ), such that, as  $T \rightarrow \infty$ ,

$$\|m^\dagger - u^\dagger\|^2 + \text{Tr}\left(\frac{1}{T}\mathcal{B}_\lambda^{-1}\right) = \mathcal{O}_{\mathbb{P}}(T^{-\text{rate}}).$$

# Nonparametric Drift Estimation in SDEs - Example



# Drift Estimation in SDEs - Error Equation

- Error  $e = m^\dagger - u^\dagger$

$$e = \mathcal{B}_\lambda^{-1} \left( \frac{1}{2} \left( \frac{L_T(\cdot; Y^\dagger)}{T} - \rho^\dagger \right)' - \lambda \mathcal{C}_0^{-1} u^\dagger - \left( \frac{L_T(\cdot; Y^\dagger)}{T} - \rho^\dagger \right) u^\dagger + \frac{\chi_T(\cdot; Y^\dagger)}{T} \right).$$

- As  $\lambda \rightarrow 0$ ,  $\|\mathcal{B}_\lambda^{-1}\|_{\mathcal{L}(H^{-\theta}, H)} = \mathcal{O}_{\mathbb{P}}(\lambda^{-\frac{\theta}{2}})$ ,  $\theta \in [0, 1]$ .
- As  $\lambda \rightarrow 0$  and  $T \rightarrow \infty$ , terms in parenthesis go to zero in different sufficiently weak spaces.
- Choose  $\lambda = \lambda(T)$  (hence  $\tau = \tau(T)$ ) optimally.

# Drift Estimation in SDEs - Convergence

Theorem (Pokern, Stuart, Van Zanten 12; slightly improved by A., Stuart, Zhang)

Assume  $u^\dagger \in H^\gamma$ ,  $\gamma \geq p$ . For appropriate choice  $\lambda = \lambda(T) \rightarrow 0$  we have as  $T \rightarrow \infty$

$$\text{rate} = \begin{cases} \frac{\gamma}{1+2\gamma+\delta}, & \text{if } \gamma \in [p, 2p] \\ \frac{2p}{1+4p+\delta}, & \text{if } \gamma > 2p, \end{cases}$$

$\delta > 0$  arbitrarily small.



# Drift Estimation in SDEs - Convergence

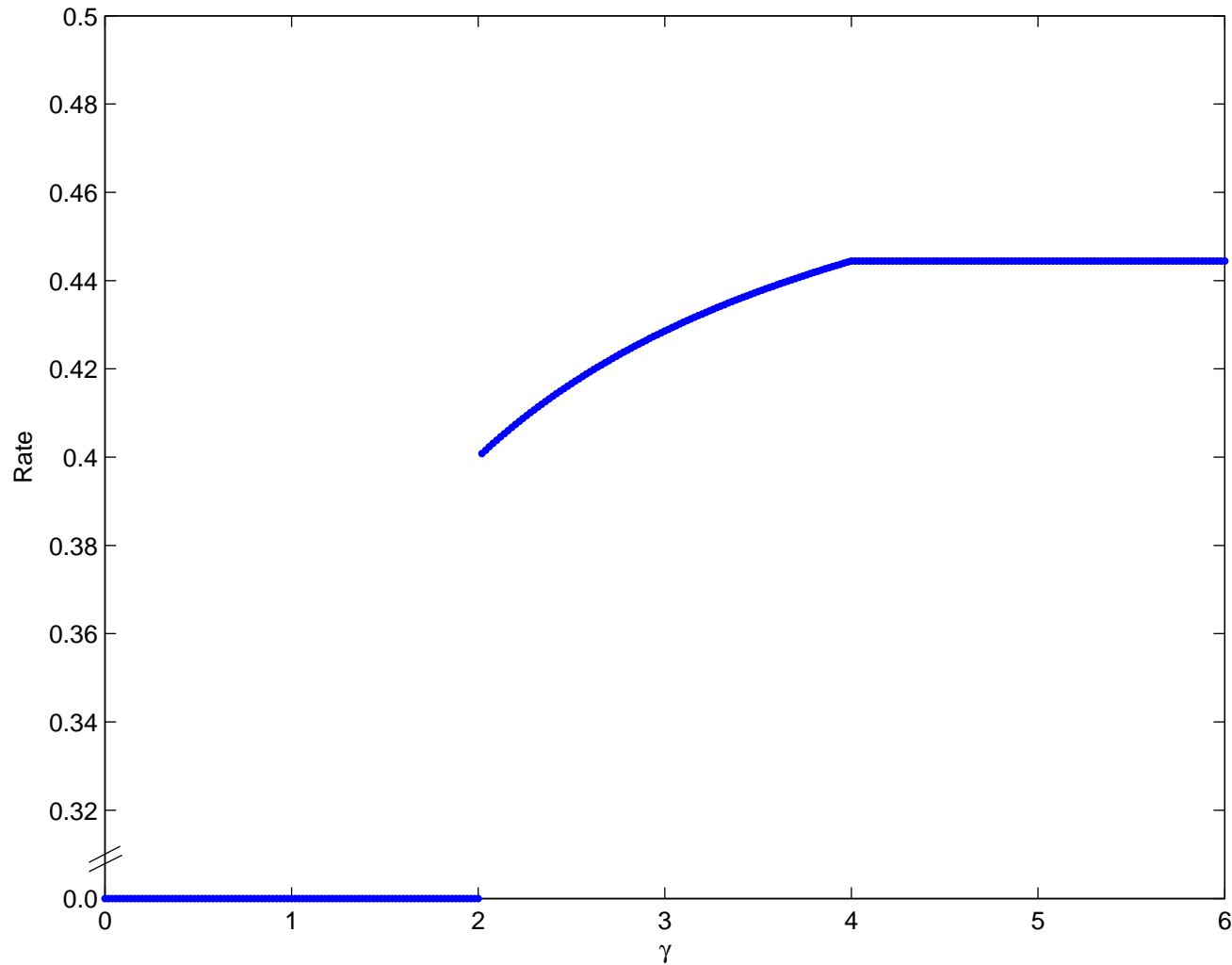


Figure: contraction for  $\mathcal{C}_0^{-1} = \left(-\frac{d^2}{dx^2}\right)^2 + I$  so that **rate** =  $\frac{\gamma}{1+2\gamma+\delta}$ ,  $\gamma \in [2, 4]$ .

# Drift Estimation in SDEs - Molecular Dynamics Example

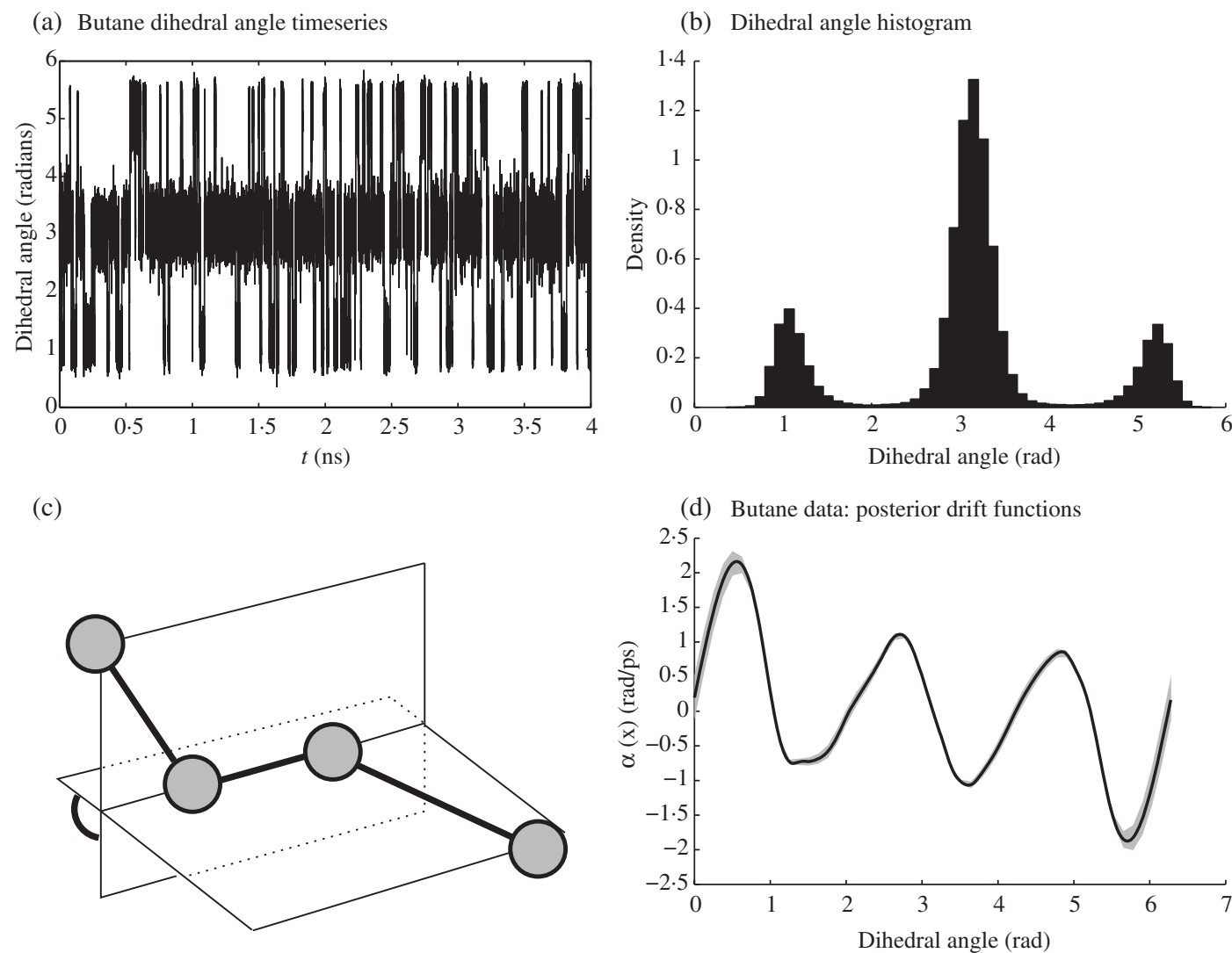


Figure: O. Papasiliopoulos, Y. Pokern, G. O. Roberts, A. M. Stuart (2012), *Nonparametric estimation of diffusions: a differential equations approach*, *Biometrika*.

# Outline

- 1 Introduction
- 2 Linear Inverse Problem
- 3 Nonparametric Drift Estimation in SDEs
- 4 Abstract Theory - Conclusions**

# Abstract Theory

- *Prior:*  $\mu_0 = \mathcal{N}(0, \tau^2 \mathcal{C}_0)$ ,  $\mathcal{C}_0$  selfadjoint positive definite trace class in  $X$ .
- $X^s$  Hilbert scale induced by  $\mathcal{C}_0^{-\frac{1}{2}}$ .
- *Posterior:* assume

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi_\epsilon(u; y))$$

$$\Phi_\epsilon(u; y) = \frac{1}{2\epsilon} \langle u, Q(y; \epsilon)u \rangle - \frac{1}{\epsilon} \langle u, r(y; \epsilon) \rangle$$

- Interested in  $\epsilon \rightarrow 0$  modelling improvement in data.

# Abstract Theory - Posterior Identification

## Theorem

Posterior is Gaussian,  $\mu^y = \mathcal{N}(m, \epsilon \mathcal{B}_\lambda^{-1})$

$$\mathcal{B}_\lambda = \mathcal{B}_\lambda(y, \epsilon) = Q(y; \epsilon) + \lambda \mathcal{C}_0^{-1}$$

$$\mathcal{B}_\lambda m = r(y; \epsilon),$$

$\lambda = \frac{\epsilon}{\tau^2}$  regularization parameter.

# Abstract Theory - Posterior Identification Sketch Proof

- Radon-Nikodym derivative  $\frac{d\mu^y}{d\mu_0}$  gives **existence of posterior**;
- finite dimensional approximation of posterior in Karhunen-Loeve basis of prior

$$\frac{d\mu^{N,y}}{d\mu_0}(u) = \frac{1}{Z_N(y)} \exp(-\Phi_\epsilon(P^N u; y));$$

approximation is Gaussian and converges weakly to posterior;

- Gaussian family closed under weak convergence, hence **posterior Gaussian**. Justify equations for mean and covariance by showing that mean and covariance of the approximation converge to  $m$  and  $\epsilon \mathcal{B}_\lambda^{-1}$ .

# Abstract Theory - Meta Theorem

- Consider data  $y^\dagger = y^\dagger(\epsilon)$  produced from underlying truth  $u^\dagger \in X^\gamma$ .
- Assume  $r(y^\dagger; \epsilon) \approx \hat{r}$  and  $Q(y^\dagger; \epsilon) \approx \hat{Q}$ , where  $u^\dagger = \hat{Q}^{-1}\hat{r}$ .
- $\mu^{y=y^\dagger} = \mathcal{N}(m^\dagger, \epsilon\mathcal{B}_\lambda^{-1})$ , where  $\mathcal{B}_\lambda = \mathcal{B}_\lambda(y^\dagger)$ .

## Theorem

Find and optimize **rate=rate**( $\gamma$ ) such that as  $\epsilon \rightarrow 0$ ,

$$\|m^\dagger - u^\dagger\|^2 + \text{Tr}(\epsilon\mathcal{B}_\lambda^{-1}) = \mathcal{O}(\epsilon^{\text{rate}}),$$

in some probabilistic sense since  $y^\dagger$  random.

# Abstract Theory - Method Overview

- Error  $e = m^\dagger - u^\dagger$  satisfies (Lax-Milgram in  $X^1$ )

$$e = \mathcal{B}_\lambda^{-1} \left( (\hat{r} - r(y^\dagger; \epsilon)) + (Q(y^\dagger; \epsilon) - \hat{Q})u^\dagger + \lambda \mathcal{C}_0^{-1} u^\dagger \right).$$

- Rate of convergence determined by  $\gamma$  and
  - $\|\hat{r} - r(y^\dagger; \epsilon)\|_{-\alpha} = \mathcal{O}(\epsilon^a)$  as  $\epsilon \rightarrow 0$ ;
  - $\|\hat{Q} - Q(y^\dagger; \epsilon)\|_{op} = \mathcal{O}(\epsilon^b)$  as  $\epsilon \rightarrow 0$ ;
  - $\|\mathcal{B}_\lambda^{-1}\|_{op} = \mathcal{O}(\lambda^{-c})$  as  $\lambda \rightarrow 0$ .
- $\text{Tr}(\epsilon \mathcal{B}_\lambda^{-1})$  often dominated by terms arising in  $\|m^\dagger - u^\dagger\|^2$ .
- Rate optimized by choosing  $\lambda = \lambda(\epsilon)$  (inducing choice  $\tau = \tau(\epsilon)$ ).



# Abstract Theory - Method Overview

- Assume

$$\langle Q(y; \epsilon)e, e \rangle \geq c \|e\|_\sigma, \quad \sigma \leq 0,$$

possibly probabilistic statement.

## Lemma

For  $\eta = (1 - \theta)\sigma + \theta$ ,  $\theta \in [0, 1]$

$$\|\mathcal{B}_\lambda^{-1}\|_{\mathcal{L}(X^{-\eta}, X)} \leq \lambda^{-\frac{\theta}{2} + \frac{\sigma}{2(1-\sigma)}}$$



# Abstract Theory - PDE Method Sketch Proof

- Solve  $\mathcal{B}_\lambda e = w$  weakly in  $X^1$
- Let  $w \in X^{-\eta}$ , unique weak solution  $e$

$$\langle \mathcal{B}_\lambda e, e \rangle = \langle w, e \rangle$$

- By assumption, interpolation, Cauchy with  $\delta$ , Young's inequality

$$\begin{aligned} \|e\|_\sigma^2 + \lambda \|e\|_1^2 &\leq \|w\|_{-\eta} \|e\|_\eta \\ &\leq \lambda^{-\frac{\theta}{2}} \|w\|_{-\eta} \|e\|_\sigma^{1-\theta} (\lambda^{\frac{1}{2}} \|e\|_1)^\theta \\ &\leq \frac{c}{2\delta} (\lambda^{-\theta} \|w\|_{-\eta}^2) + \frac{c\delta}{2} \left( \|e\|_\sigma^{2(1-\theta)} (\lambda^{\frac{1}{2}} \|e\|_1)^{2\theta} \right) \\ &\leq \frac{c}{2\delta} (\lambda^{-\theta} \|w\|_{-\eta}^2) + \frac{c\delta}{2} \left( (1-\theta) \|e\|_\sigma^2 + \theta \lambda \|e\|_1^2 \right) \end{aligned}$$

# Abstract Theory - PDE Method Sketch Proof

- Choose  $\delta > 0$  small

$$\|\mathbf{e}\|_{\sigma}^2 \leq c\lambda^{-\theta} \|\mathbf{w}\|_{-\eta}^2$$

$$\|\mathbf{e}\|_1^2 \leq c\lambda^{-\theta-1} \|\mathbf{w}\|_{-\eta}^2.$$

- Interpolation gives

$$\|\mathbf{e}\|^2 \leq c\lambda^{-\theta + \frac{\sigma}{1-\sigma}} \|\mathbf{w}\|_{-\eta}^2$$

or

$$\|\mathcal{B}_{\lambda}^{-1}\mathbf{w}\| \leq c\lambda^{-\frac{\theta}{2} + \frac{\sigma}{2(1-\sigma)}} \|\mathbf{w}\|_{-\eta}.$$



# Conclusions - Further Work





We have

- formulated two inverse problems in Bayesian framework with Gaussian priors;
- shown posterior is Gaussian, worked with precision operators;
- examined posterior consistency;
- highlighted common structure and proposed abstract theory.

Future work

- more examples;
- data driven choice of hyperparameters;
- ...

<http://homepages.warwick.ac.uk/~mariba/>

-  S. Agapiou, S. Larsson and A. M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Under revision Stoch. Proc. Appl., <http://arxiv.org/abs/1203.5753>
-  Y. Pokern, A. M. Stuart and J. H. Van Zanten, *Posterior consistency via precision operators for nonparametric drift estimation in SDEs*, Accepted Stoch. Proc. Appl., <http://arxiv.org/abs/1202.0976>
-  S. Agapiou, A. M. Stuart and Y. X. Zhang, *An abstract framework for the study of posterior contraction in Bayesian inverse problems*, in preparation.
-  S. Agapiou, A. M. Stuart and Y. X. Zhang, *Bayesian posterior contraction rates for linear severely ill-posed inverse problems*, <http://arxiv.org/abs/1210.1563>