Bayesian Adaptive Smoothing Spline Using Stochastic Differential Equations (SDEs)

Yu Ryan Yue

Baruch College, The City University of New York

Finn Lindgren (Norwegian University of Science and Technology)

Daniel Simpson (University of Helsinki)

Håvard Rue (Norwegian University of Science and Technology)

LGM 2012, Trondheim, Norway *June 1, 2012*

Outline

- Background and motivation
- SDEs for adaptive smoothing
- Modeling of local smoothing function
- MCMC and INLA algorithms
- Simulated example
- Discussion

Background

Classic smoothing spline

Background

Classic smoothing spline

- Nonparametric regression:

$$y_i = f(t_i) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \tau^{-1}\right)$$

for $i = 1, \ldots, n$, where $t_i \in \mathcal{T} \subset \mathbf{R}^1$ and f is unknown but smooth.

Background

Classic smoothing spline

- Nonparametric regression:

$$y_i = f(t_i) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \tau^{-1}\right)$$

for $i = 1, \ldots, n$, where $t_i \in \mathcal{T} \subset \mathbf{R}^1$ and f is unknown but smooth.

- $\widehat{f}$ is smoothing spline (e.g. Wahba, 1990) if

$$\widehat{f} = \arg\min_f \left[\sum_{i=1}^n \left(y_i - f(t_i) \right)^2 + \lambda \int \left(f^{(p)}(t) \right)^2 dt \right]$$

  - $f^{(p)}(t)$: pth derivative of $f(t)$
  - $p = 1, 2$: linear, cubic smoothing spline
  - λ: smoothing parameter

Background

Classic smoothing spline

- Nonparametric regression:

$$y_i = f(t_i) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \tau^{-1}\right)$$

for $i = 1, \ldots, n$, where $t_i \in \mathcal{T} \subset \mathbf{R}^1$ and f is unknown but smooth.

- $\widehat{f}$ is smoothing spline (e.g. Wahba, 1990) if

$$\widehat{f} = \arg\min_f \left[\sum_{i=1}^{n} \left(y_i - f(t_i) \right)^2 + \lambda \int \left(f^{(p)}(t) \right)^2 dt \right]$$

  - $f^{(p)}(t)$: pth derivative of $f(t)$
  - $p = 1, 2$: linear, cubic smoothing spline
  - λ: smoothing parameter

- $\widehat{f}$ can be derived in a reproducing kernel Hilbert space.

Background

Classic smoothing spline

- Nonparametric regression:

$$y_i = f(t_i) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \tau^{-1}\right)$$

for $i = 1, \ldots, n$, where $t_i \in \mathcal{T} \subset \mathbf{R}^1$ and f is unknown but smooth.

- $\widehat{f}$ is smoothing spline (e.g. Wahba, 1990) if

$$\widehat{f} = \arg\min_f \left[\sum_{i=1}^{n} \left(y_i - f(t_i) \right)^2 + \lambda \int \left(f^{(p)}(t) \right)^2 dt \right]$$

  - $f^{(p)}(t)$: pth derivative of $f(t)$
  - $p = 1, 2$: linear, cubic smoothing spline
  - λ: smoothing parameter

- $\widehat{f}$ can be derived in a reproducing kernel Hilbert space.
- Choose λ by e.g., CV or GCV.

Background

Stochastic differential equation

Background

Stochastic differential equation

- Kimeldorf & Wahba (1970) and Wahba (1978), consider

$$d^p f(t)/dt^p = dW(t)/dt, \quad (1)$$

  - $W(t)$ is Wiener process with mean 0 and variance t.
  - $dW(t)/dt$ is referred to as "white noise".

Background

Stochastic differential equation

- Kimeldorf & Wahba (1970) and Wahba (1978), consider

$$d^p f(t)/dt^p = dW(t)/dt, \tag{1}$$

  - $W(t)$ is Wiener process with mean 0 and variance t.
  - $dW(t)/dt$ is referred to as "white noise".

- The exact solution of (1) is

$$f(t) = \beta_0 + \beta_1 t + \cdots + \beta_{p-1} t^{p-1} + Z(t)/\sqrt{\lambda\tau},$$

Background

Stochastic differential equation

- Kimeldorf & Wahba (1970) and Wahba (1978), consider

$$d^p f(t)/dt^p = dW(t)/dt, \quad (1)$$

  - $W(t)$ is Wiener process with mean 0 and variance t.
  - $dW(t)/dt$ is referred to as "white noise".

- The exact solution of (1) is

$$f(t) = \beta_0 + \beta_1 t + \cdots + \beta_{p-1} t^{p-1} + Z(t)/\sqrt{\lambda\tau},$$

  - $\beta_0, \beta_1, \ldots, \beta_{p-1} \sim N(0, \sigma^2)$ as $\sigma^2 \to \infty$.

Background

Stochastic differential equation

- Kimeldorf & Wahba (1970) and Wahba (1978), consider

$$d^p f(t)/dt^p = dW(t)/dt, \tag{1}$$

  - $W(t)$ is Wiener process with mean 0 and variance t.
  - $dW(t)/dt$ is referred to as "white noise".

- The exact solution of (1) is

$$f(t) = \beta_0 + \beta_1 t + \cdots + \beta_{p-1} t^{p-1} + Z(t)/\sqrt{\lambda\tau},$$

  - $\beta_0, \beta_1, \ldots, \beta_{p-1} \sim N(0, \sigma^2)$ as $\sigma^2 \to \infty$.
  - $Z(t)$ is a Gaussian process with mean 0 and covariance

$$\mathbf{\Sigma}(s,t) = \int_0^1 \frac{(s-u)_+^{p-1}}{(p-1)!} \frac{(t-u)_+^{p-1}}{(p-1)!} du.$$

Background

Stochastic differential equation

- Kimeldorf & Wahba (1970) and Wahba (1978), consider

$$d^p f(t)/dt^p = dW(t)/dt, \tag{1}$$

  - $W(t)$ is Wiener process with mean 0 and variance t.
  - $dW(t)/dt$ is referred to as "white noise".

- The exact solution of (1) is

$$f(t) = \beta_0 + \beta_1 t + \cdots + \beta_{p-1} t^{p-1} + Z(t)/\sqrt{\lambda\tau},$$

  - $\beta_0, \beta_1, \ldots, \beta_{p-1} \sim N(0, \sigma^2)$ as $\sigma^2 \to \infty$.
  - $Z(t)$ is a Gaussian process with mean 0 and covariance

$$\mathbf{\Sigma}(s,t) = \int_0^1 \frac{(s-u)_+^{p-1}}{(p-1)!} \frac{(t-u)_+^{p-1}}{(p-1)!} du.$$

  - $\mathbf{\Sigma}$ is completely dense.

Background

Bayesian smoothing spline

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$
- Hierarchical model:

$$\begin{aligned} \mathbf{y} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{z} + \varepsilon, \quad \varepsilon \sim N\left(\mathbf{0}, \tau^{-1}\mathbf{I}\right) \\ \boldsymbol{\beta} &\propto 1, \quad \mathbf{z} \sim N\left(\mathbf{0}, \delta^{-1}\boldsymbol{\Sigma}\right), \quad \delta = \lambda\tau \end{aligned}$$

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$
- Hierarchical model:

$$\begin{aligned}\mathbf{y} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \tau^{-1}\mathbf{I}\right) \\ \boldsymbol{\beta} &\propto 1, \quad \mathbf{z} \sim N\left(\mathbf{0}, \delta^{-1}\boldsymbol{\Sigma}\right), \quad \delta = \lambda\tau\end{aligned}$$

- Negative log posterior:

$$\frac{\tau}{2}\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right)'\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right) + \frac{\delta}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$$

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$
- Hierarchical model:

$$\begin{aligned} \mathbf{y} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \tau^{-1}\mathbf{I}\right) \\ \boldsymbol{\beta} &\propto 1, \quad \mathbf{z} \sim N\left(\mathbf{0}, \delta^{-1}\boldsymbol{\Sigma}\right), \quad \delta = \lambda\tau \end{aligned}$$

- Negative log posterior:

$$\frac{\tau}{2}\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right)'\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right) + \frac{\delta}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$$

- Posterior mean is Bayesian estimator of smoothing spline.

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$
- Hierarchical model:

$$\begin{aligned} \mathbf{y} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{z} + \varepsilon, \quad \varepsilon \sim N\left(\mathbf{0}, \tau^{-1}\mathbf{I}\right) \\ \boldsymbol{\beta} &\propto 1, \quad \mathbf{z} \sim N\left(\mathbf{0}, \delta^{-1}\boldsymbol{\Sigma}\right), \quad \delta = \lambda\tau \end{aligned}$$

- Negative log posterior:

$$\frac{\tau}{2}\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right)'\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right) + \frac{\delta}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$$

- Posterior mean is Bayesian estimator of smoothing spline.
- For fully Bayesian inference, need ‘sensible’ priors on τ and δ (Speckman & Sun, 2003; Yue et al., 2012).

Background

Bayesian smoothing spline

- Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\mathbf{z} = (Z(t_1), \ldots, Z(t_n))'$
- Hierarchical model:

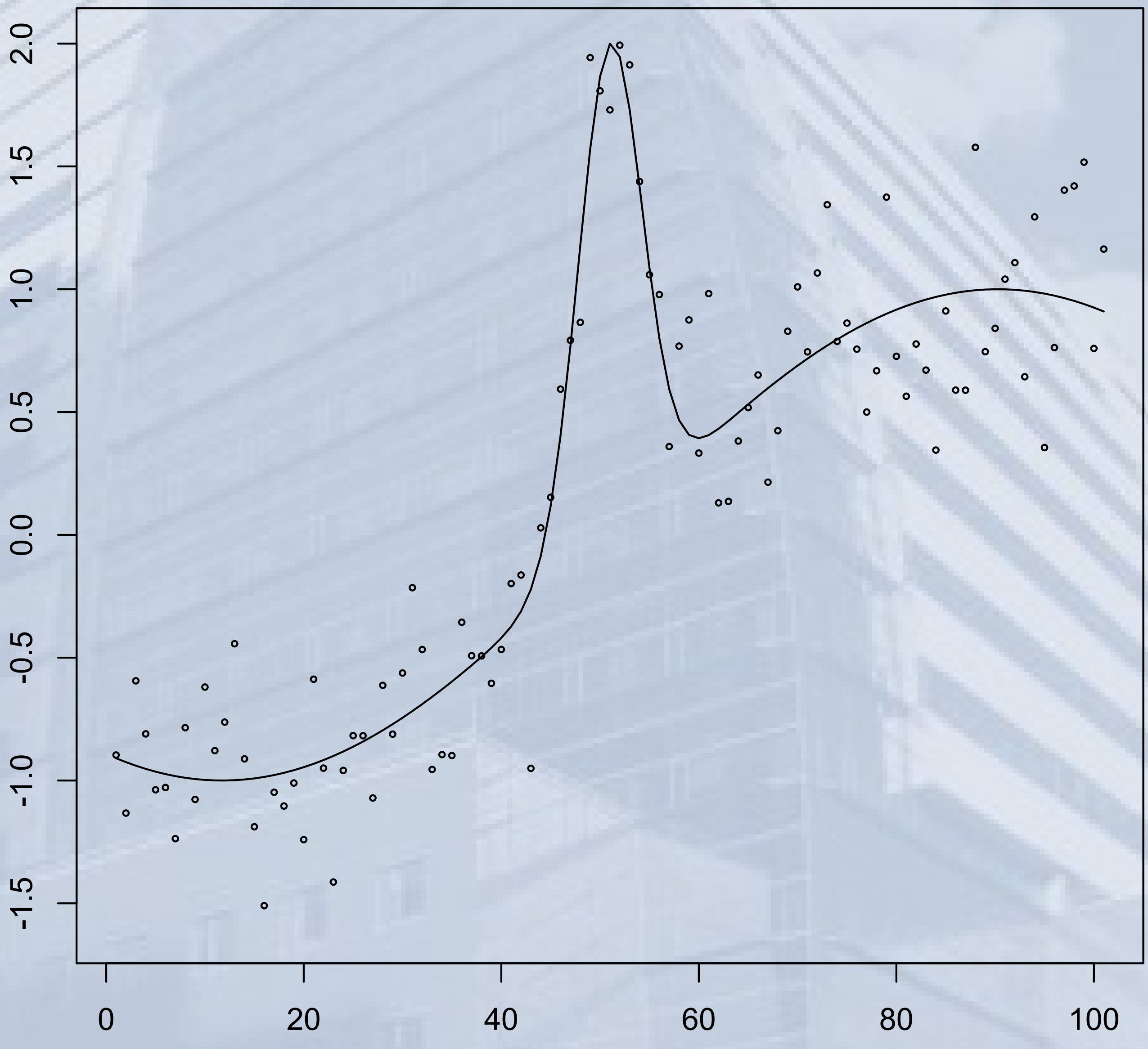
$$\begin{aligned} \mathbf{y} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{z} + \varepsilon, \quad \varepsilon \sim N\left(\mathbf{0}, \tau^{-1}\mathbf{I}\right) \\ \boldsymbol{\beta} &\propto 1, \quad \mathbf{z} \sim N\left(\mathbf{0}, \delta^{-1}\boldsymbol{\Sigma}\right), \quad \delta = \lambda\tau \end{aligned}$$

- Negative log posterior:

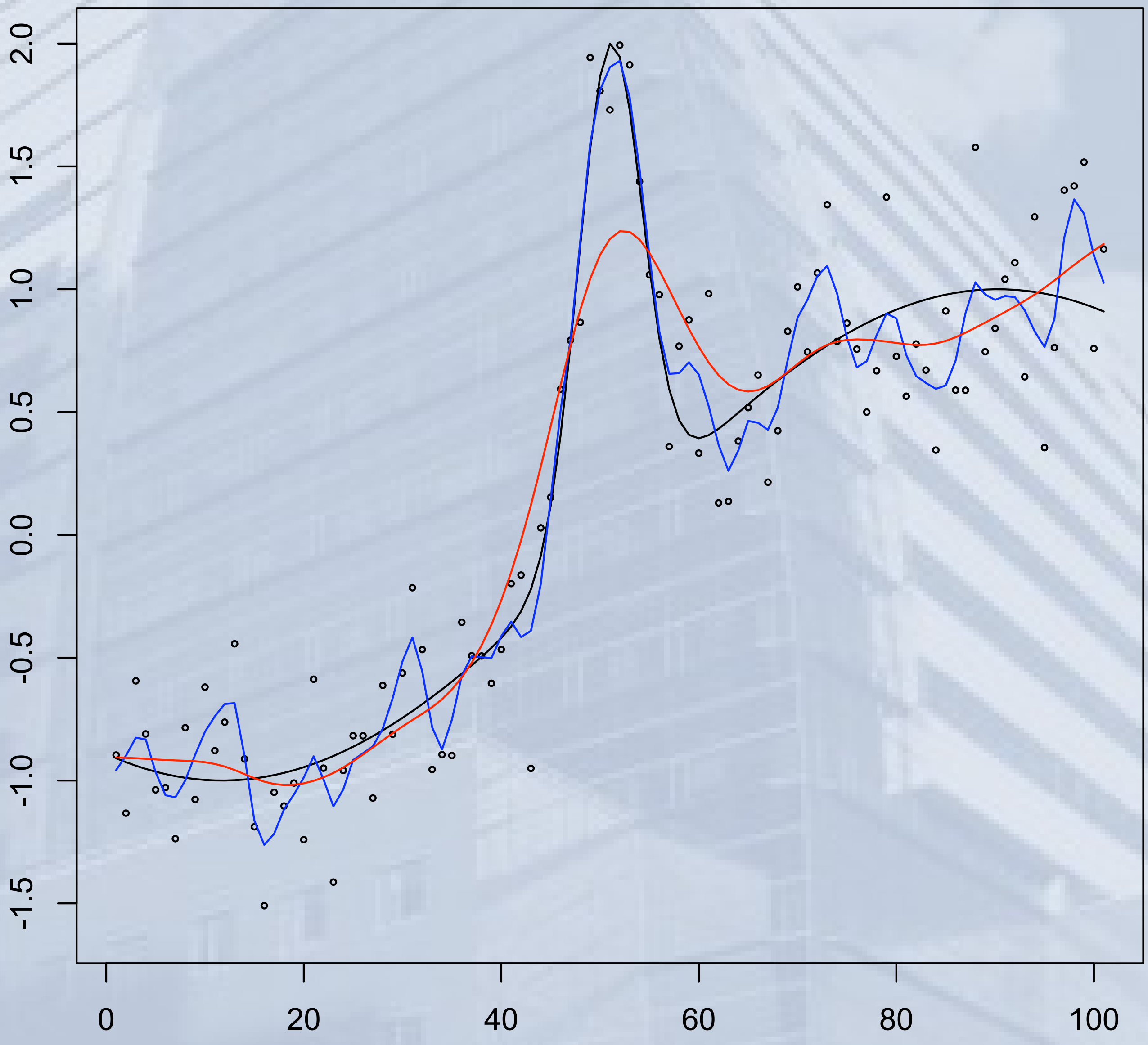
$$\frac{\tau}{2}\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right)'\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{z}\right) + \frac{\delta}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$$

- Posterior mean is Bayesian estimator of smoothing spline.
- For fully Bayesian inference, need ‘sensible’ priors on τ and δ (Speckman & Sun, 2003; Yue et al., 2012).
- Straightforward MCMC, but inefficient for big n.

Inhomogeneous function



Problem: single smoothing parameter



Brief review on existing work

Brief review on existing work

Smoothing splines

- Hybrid adaptive splines (Luo & Wahba, 1997)
- Local generalized cross-validation (Cummins et al., 2001)
- Spatially adaptive smoothing splines (Pintore et al., 2006)
- Bayesian discretized adaptive smoothing spine (Yue et al., 2012)

Brief review on existing work

Smoothing splines

- Hybrid adaptive splines (Luo & Wahba, 1997)
- Local generalized cross-validation (Cummins et al., 2001)
- Spatially adaptive smoothing splines (Pintore et al., 2006)
- Bayesian discretized adaptive smoothing spine (Yue et al., 2012)

Regression splines

- Adaptively select the number and locations of knots (e.g., Friedman, 1991; Smith & Kohn, 1996; Di Matteo et al, 2001)

Brief review on existing work

Smoothing splines

- Hybrid adaptive splines (Luo & Wahba, 1997)
- Local generalized cross-validation (Cummins et al., 2001)
- Spatially adaptive smoothing splines (Pintore et al., 2006)
- Bayesian discretized adaptive smoothing spine (Yue et al., 2012)

Regression splines

- Adaptively select the number and locations of knots (e.g., Friedman, 1991; Smith & Kohn, 1996; Di Matteo et al, 2001)

P-splines

- Model smoothing parameters with another P-spline (e.g., Lang & Brezger, 2004; Baladandayuthapani et al., 2005; Crainiceanu et al., 2007; Krivobokova et al., 2008; Scheipl & Kneib, 2009)

Sparse solution of SDE

Sparse solution of SDE

- Consider

$$d^2f(t)/dt^2 = dW(t)/dt$$

Sparse solution of SDE

- Consider

$$d^2 f(t)/dt^2 = dW(t)/dt$$

- Define inner product $\langle f, g \rangle = \int f(t)g(t)dt$

Sparse solution of SDE

- Consider

$$d^2 f(t)/dt^2 = dW(t)/dt$$

- Define inner product $\langle f, g \rangle = \int f(t)g(t)dt$
- We seek a weak solution that satisfies:

$$\left\langle \phi, d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \phi, dW/dt \right\rangle \quad (2)$$

  - $\phi(t)$: any sensible test function.

Sparse solution of SDE

- Consider

$$d^2 f(t)/dt^2 = dW(t)/dt$$

- Define inner product $\langle f, g \rangle = \int f(t)g(t)dt$
- We seek a weak solution that satisfies:

$$\left\langle \phi, d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \phi, dW/dt \right\rangle \qquad (2)$$

  – $\phi(t)$: any sensible test function.
- Let $t_1 < t_2 < \cdots < t_n$ be fixed points.

Sparse solution of SDE

- Consider

$$d^2 f(t)/dt^2 = dW(t)/dt$$

- Define inner product $\langle f, g \rangle = \int f(t)g(t)dt$
- We seek a weak solution that satisfies:

$$\left\langle \phi, d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \phi, dW/dt \right\rangle \tag{2}$$

  – $\phi(t)$: any sensible test function.
- Let $t_1 < t_2 < \cdots < t_n$ be fixed points.
- Approximate f as

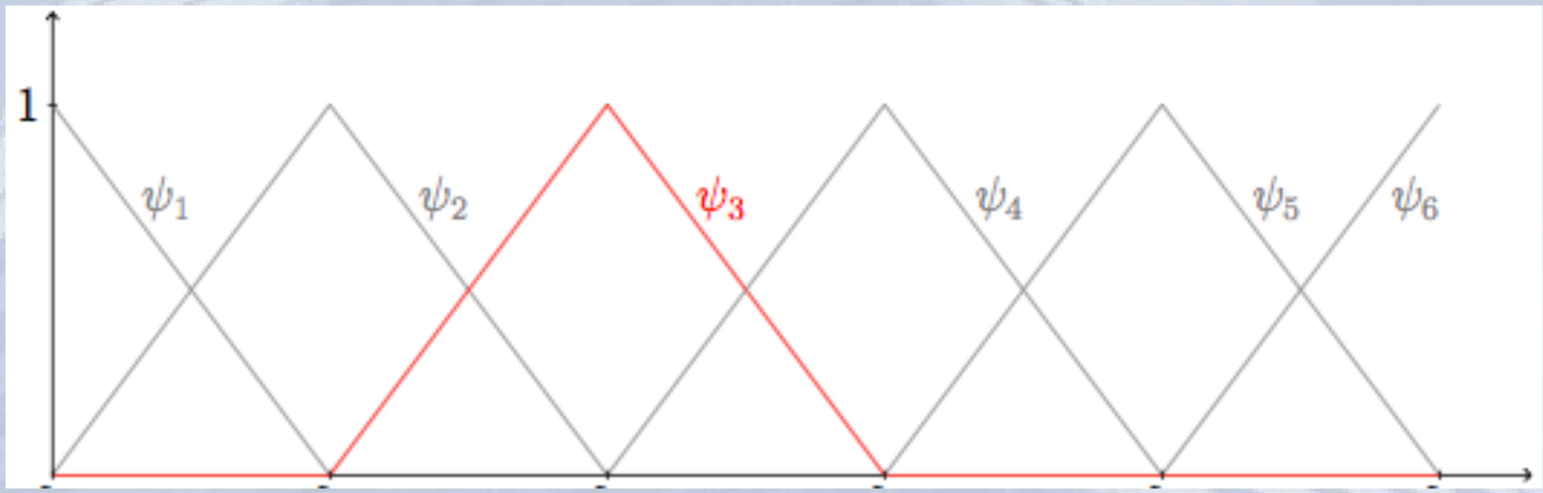
$$f(t) \approx \sum_{i=1}^{n} \psi_i(t) w_i$$

  – ψ_i: basis functions.

  – w_i: random weights.

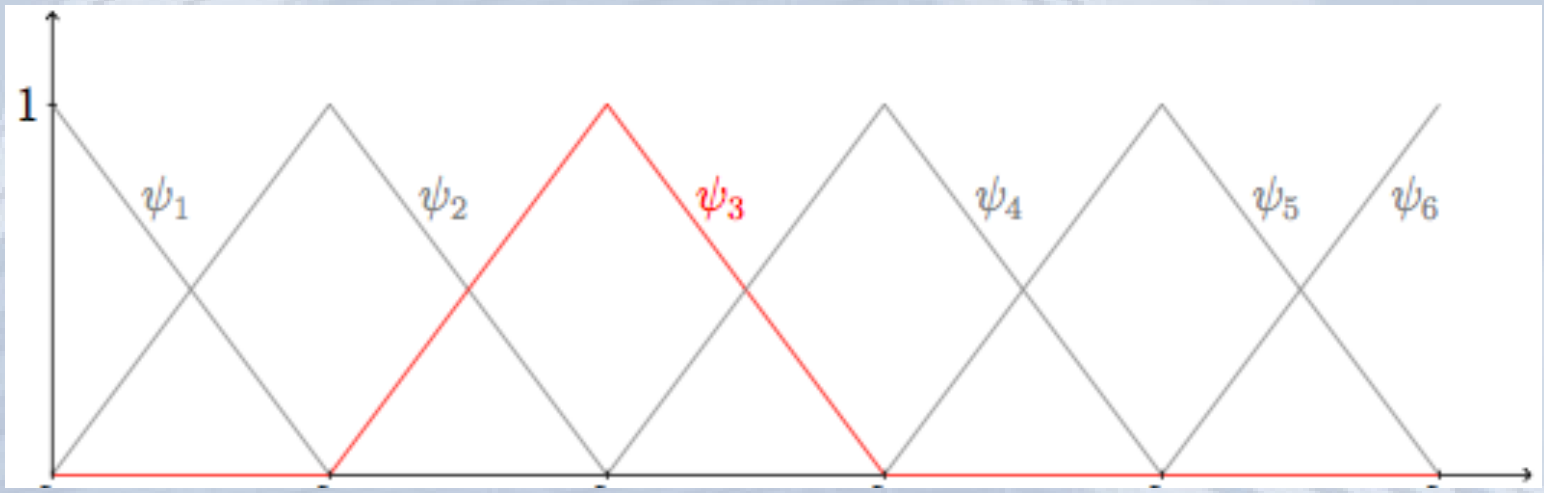
Sparse solution of SDE

Piecewise linear basis



Sparse solution of SDE

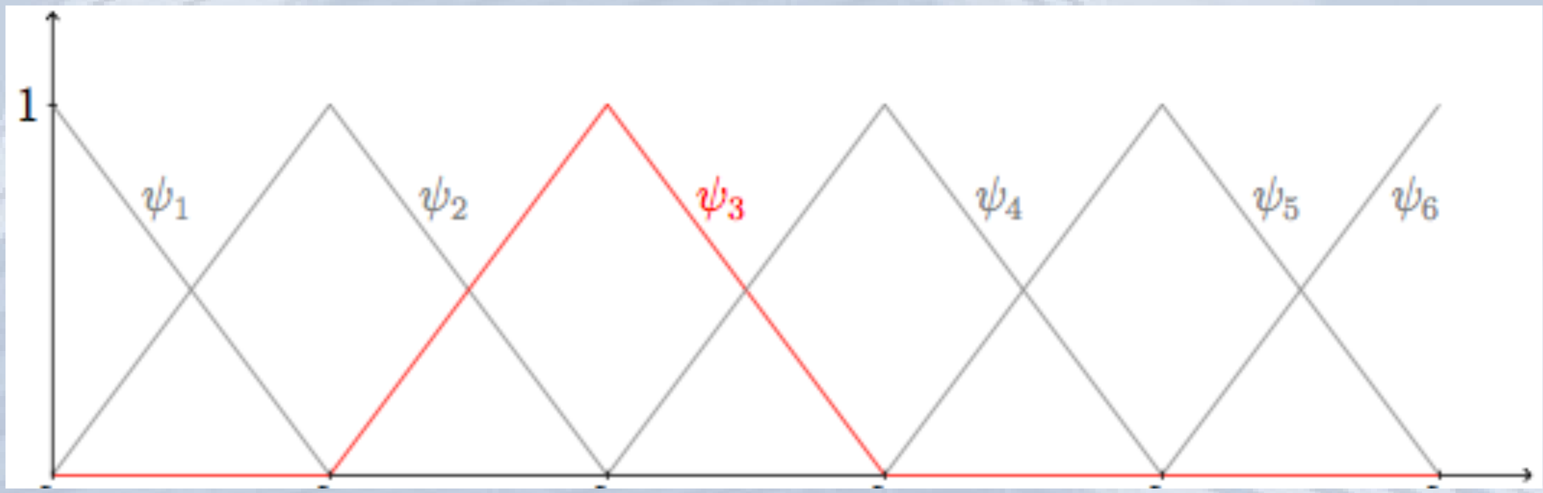
Piecewise linear basis



- The explicit form of the solution.

Sparse solution of SDE

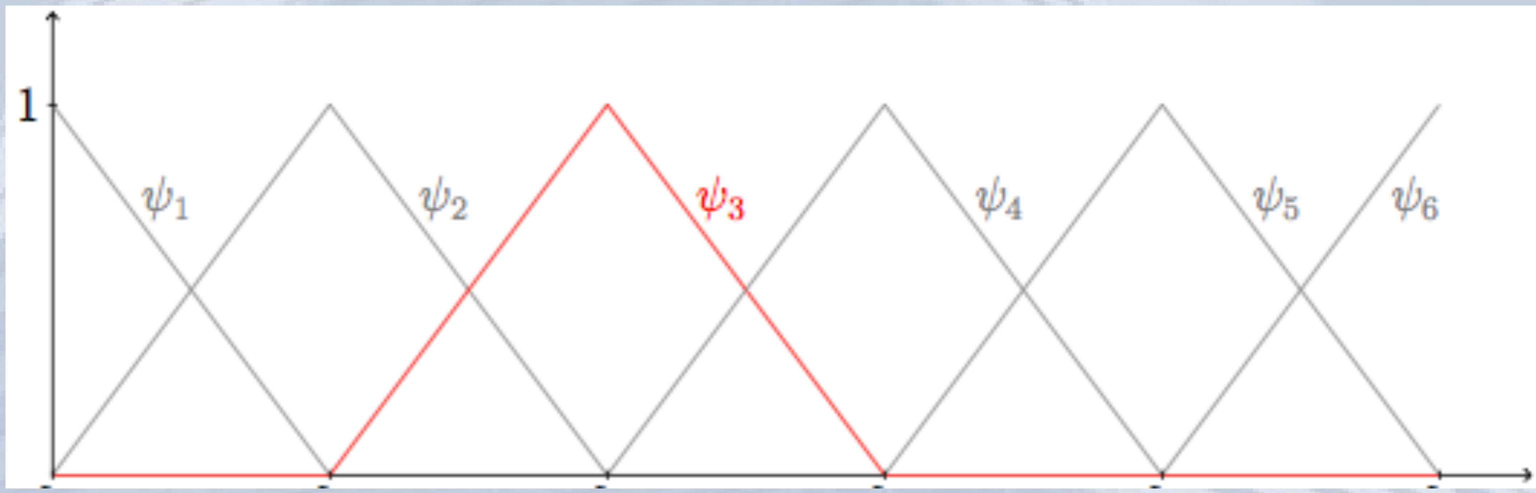
Piecewise linear basis



- The explicit form of the solution.
- Little gain with more complicated basis functions.

Sparse solution of SDE

Piecewise linear basis



- The explicit form of the solution.
- Little gain with more complicated basis functions.
- Nice theoretical results are available.

Sparse Solution of SDE

Galerkin finite element method (Lindgren & Rue, 2008)

Sparse Solution of SDE

Galerkin finite element method (Lindgren & Rue, 2008)

- Choose a finite set $\{\phi_j(t)\}_{j=1}^{n}$ and let $\phi_j(t) = \psi_j(t)$.

Sparse Solution of SDE

Galerkin finite element method (Lindgren & Rue, 2008)

- Choose a finite set $\{\phi_j(t)\}_{j=1}^{n}$ and let $\phi_j(t) = \psi_j(t)$.
- Weak SDE (2) becomes a linear system of equations

$$\sum_{i=1}^{n} w_i \left\langle \psi_j, d^2\psi_i/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{3}$$

Sparse Solution of SDE

Galerkin finite element method (Lindgren & Rue, 2008)

- Choose a finite set $\{\phi_j(t)\}_{j=1}^{n}$ and let $\phi_j(t) = \psi_j(t)$.
- Weak SDE (2) becomes a linear system of equations

$$\sum_{i=1}^{n} w_i \left\langle \psi_j, d^2\psi_i/dt^2 \right\rangle \overset{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{3}$$

- LHS of (3) = $\mathbf{Hw}$, where

$$\mathbf{H}[i, i-1] = \frac{1}{h_{i-1}}, \; \mathbf{H}[i,i] = -\left(\frac{1}{h_{i-1}} + \frac{1}{h_i}\right), \; \mathbf{H}[i, i+1] = \frac{1}{h_i}.$$

Sparse Solution of SDE

Galerkin finite element method (Lindgren & Rue, 2008)

- Choose a finite set $\{\phi_j(t)\}_{j=1}^{n}$ and let $\phi_j(t) = \psi_j(t)$.
- Weak SDE (2) becomes a linear system of equations

$$\sum_{i=1}^{n} w_i \left\langle \psi_j, d^2\psi_i/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{3}$$

- LHS of (3) = $\mathbf{Hw}$, where

$$\mathbf{H}[i, i-1] = \frac{1}{h_{i-1}},\ \mathbf{H}[i,i] = -\left(\frac{1}{h_{i-1}} + \frac{1}{h_i}\right),\ \mathbf{H}[i, i+1] = \frac{1}{h_i}.$$

- RHS of (3) = $N(0, \mathbf{B}^{-1})$ with $\mathbf{B}[i,j] = \langle \psi_i, \psi_j \rangle$.

Sparse Solution of SDE

Sparse Solution of SDE

- Then, **w** has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.
- The null space of $\mathbf{Q}$ is spanned by $\mathbf{T}$.

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.
- The null space of $\mathbf{Q}$ is spanned by $\mathbf{T}$.
- $\mathbf{w}$ is an intrinsic GMRF (Rue & Held, 2005).

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.
- The null space of $\mathbf{Q}$ is spanned by $\mathbf{T}$.
- $\mathbf{w}$ is an intrinsic GMRF (Rue & Held, 2005).
- Estimated function $\hat{\mathbf{f}} = \mathbf{\Psi}\hat{\mathbf{w}}$ is the best finite approximation to continuous process (Lindgren & Rue, 2008).

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.
- The null space of $\mathbf{Q}$ is spanned by $\mathbf{T}$.
- $\mathbf{w}$ is an intrinsic GMRF (Rue & Held, 2005).
- Estimated function $\hat{\mathbf{f}} = \mathbf{\Psi}\hat{\mathbf{w}}$ is the best finite approximation to continuous process (Lindgren & Rue, 2008).
- Do not need to play with knots.

Sparse Solution of SDE

- Then, $\mathbf{w}$ has density

$$[\mathbf{w} \mid \delta] \propto \delta^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{w}'\mathbf{Q}\mathbf{w}\right), \quad \mathbf{Q} = \mathbf{H}'\mathbf{B}^{-1}\mathbf{H}.$$

- To make $\mathbf{Q}$ sparse, replace $\mathbf{B}$ with $\tilde{\mathbf{B}}$ where $\tilde{\mathbf{B}}[i,i] = \langle \psi_i, 1 \rangle$.
- The null space of $\mathbf{Q}$ is spanned by $\mathbf{T}$.
- $\mathbf{w}$ is an intrinsic GMRF (Rue & Held, 2005).
- Estimated function $\hat{\mathbf{f}} = \mathbf{\Psi}\hat{\mathbf{w}}$ is the best finite approximation to continuous process (Lindgren & Rue, 2008).
- Do not need to play with knots.
- Easy extension to adaptive smoothing.

Adaptive SDE v1

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

where $\lambda(t)$ is adaptive smoothing function.

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

where $\lambda(t)$ is adaptive smoothing function.

- A small $\lambda(t)$ allows big $f''(t)$ for more roughness, while a large value diminishes the derivative to increase smoothness.

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

where $\lambda(t)$ is adaptive smoothing function.

- A small $\lambda(t)$ allows big $f''(t)$ for more roughness, while a large value diminishes the derivative to increase smoothness.
- In the same spirit with Pintore et al. (2006):

$$\widehat{f}_\lambda = \text{arg}_f \min \sum_{i=1}^{n} \Big(y_i - f(t_i) \Big)^2 + \int \Big[\lambda(t) f''(t) \Big]^2 dt. \tag{5}$$

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

  where $\lambda(t)$ is adaptive smoothing function.

- A small $\lambda(t)$ allows big $f''(t)$ for more roughness, while a large value diminishes the derivative to increase smoothness.

- In the same spirit with Pintore et al. (2006):

$$\widehat{f}_\lambda = \arg_f \min \sum_{i=1}^{n} \left(y_i - f(t_i) \right)^2 + \int \left[\lambda(t) f''(t) \right]^2 dt. \tag{5}$$

  - Closed-form solution by reproducing kernels in Hilbert space.

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

where $\lambda(t)$ is adaptive smoothing function.

- A small $\lambda(t)$ allows big $f''(t)$ for more roughness, while a large value diminishes the derivative to increase smoothness.
- In the same spirit with Pintore et al. (2006):

$$\hat{f}_\lambda = \arg_f \min \sum_{i=1}^{n} \Big(y_i - f(t_i) \Big)^2 + \int \Big[\lambda(t) f''(t) \Big]^2 dt. \tag{5}$$

  – Closed-form solution by reproducing kernels in Hilbert space.

  – Piecewise-constant model for $\lambda(t)$.

Adaptive SDE v1

- Formulation I

$$\lambda(t) d^2 f(t)/dt^2 = dW(t)/dt, \tag{4}$$

where $\lambda(t)$ is adaptive smoothing function.

- A small $\lambda(t)$ allows big $f''(t)$ for more roughness, while a large value diminishes the derivative to increase smoothness.
- In the same spirit with Pintore et al. (2006):

$$\hat{f}_\lambda = \arg_f \min \sum_{i=1}^{n} \Big(y_i - f(t_i)\Big)^2 + \int \Big[\lambda(t) f''(t)\Big]^2 dt. \tag{5}$$

  - Closed-form solution by reproducing kernels in Hilbert space.
  - Piecewise-constant model for $\lambda(t)$.
  - Computationally intensive...

Adaptive SDE v1

Adaptive SDE v1

- Weak formulation of (4)

$$\left\langle \psi_j, \lambda d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{6}$$

Adaptive SDE v1

- Weak formulation of (4)

$$\left\langle \psi_j, \lambda d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{6}$$

- It can be proved that (6) yields

$$\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$$

  - $\mathbf{\Lambda} = \text{diag}(\lambda(t_1), \ldots, \lambda(t_n))$
  - $\mathbf{H}$ and $\tilde{\mathbf{B}}$ are the same as before.

Adaptive SDE v1

- Weak formulation of (4)

$$\left\langle \psi_j, \lambda d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{6}$$

- It can be proved that (6) yields

$$\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$$

  - $\mathbf{\Lambda} = \text{diag}(\lambda(t_1), \ldots, \lambda(t_n))$
  - $\mathbf{H}$ and $\tilde{\mathbf{B}}$ are the same as before.

- So, $\mathbf{w}$ is a GMRF with

$$\mathbf{Q}_\lambda = \mathbf{H}'\mathbf{\Lambda}\tilde{\mathbf{B}}^{-1}\mathbf{\Lambda H}.$$

Adaptive SDE v1

- Weak formulation of (4)

$$\left\langle \psi_j, \lambda d^2 f/dt^2 \right\rangle \stackrel{d}{=} \left\langle \psi_j, dW/dt \right\rangle, \quad j = 1, \ldots, n. \tag{6}$$

- It can be proved that (6) yields

$$\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$$

  - $\mathbf{\Lambda} = \mathrm{diag}(\lambda(t_1), \ldots, \lambda(t_n))$
  - $\mathbf{H}$ and $\tilde{\mathbf{B}}$ are the same as before.

- So, $\mathbf{w}$ is a GMRF with

$$\mathbf{Q}_\lambda = \mathbf{H}'\mathbf{\Lambda}\tilde{\mathbf{B}}^{-1}\mathbf{\Lambda H}.$$

- $\mathbf{Q}_\lambda$ is symmetric and sparse matrix.

Adaptive SDE v2

Adaptive SDE v2

- Formulation II

$$d^2\lambda(t)f(t)/dt^2 = dW(t)/dt. \tag{7}$$

Adaptive SDE v2

- Formulation II

$$d^2\lambda(t)f(t)/dt^2 = dW(t)/dt. \tag{7}$$

- A small $\lambda(t)$ compresses the scale giving quick oscillations, while a high value stretch $f(t)$, decreasing the roughness.

Adaptive SDE v2

- Formulation II

$$d^2\lambda(t)f(t)/dt^2 = dW(t)/dt. \tag{7}$$

- A small $\lambda(t)$ compresses the scale giving quick oscillations, while a high value stretch $f(t)$, decreasing the roughness.
- Letting $\tilde{f}(t) = \lambda(t)f(t)$, the solution to (7) yields

$$\hat{f}_\lambda = \text{arg}_f \min \sum_{i=1}^{n} \left(y_i - \tilde{f}(t_i) \right)^2 + \int \tilde{f}''(t)^2 dt.$$

Adaptive SDE v2

- Formulation II

$$d^2\lambda(t)f(t)/dt^2 = dW(t)/dt. \tag{7}$$

- A small $\lambda(t)$ compresses the scale giving quick oscillations, while a high value stretch $f(t)$, decreasing the roughness.
- Letting $\tilde{f}(t) = \lambda(t)f(t)$, the solution to (7) yields

$$\hat{f}_\lambda = \text{arg}_f \min \sum_{i=1}^{n} \left(y_i - \tilde{f}(t_i) \right)^2 + \int \tilde{f}''(t)^2 dt.$$

- The weak solution of (7)

$$\mathbf{H\Lambda w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$$

Adaptive SDE v2

- Formulation II

$$d^2\lambda(t)f(t)/dt^2 = dW(t)/dt. \tag{7}$$

- A small $\lambda(t)$ compresses the scale giving quick oscillations, while a high value stretch $f(t)$, decreasing the roughness.
- Letting $\tilde{f}(t) = \lambda(t)f(t)$, the solution to (7) yields

$$\hat{f}_\lambda = \text{arg}_f \min \sum_{i=1}^{n} \left(y_i - \tilde{f}(t_i) \right)^2 + \int \tilde{f}''(t)^2 dt.$$

- The weak solution of (7)

$$\mathbf{H\Lambda w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$$

- So, **w** is a GMRF with

$$\mathbf{Q}_\lambda = \mathbf{\Lambda H}'\tilde{\mathbf{B}}^{-1}\mathbf{H\Lambda}.$$

Modeling $\lambda(t)$

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and ‘friendly’ smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).
- Letting $\log(\lambda(t)) = \nu(t)$, consider SDE (Lindgren et al., 2011)

$$\left(\kappa^2 - d^2/dt^2\right)\nu(t) = dW(t)/dt, \quad \kappa > 0. \tag{8}$$

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).
- Letting $\log(\lambda(t)) = \nu(t)$, consider SDE (Lindgren et al., 2011)

$$\left(\kappa^2 - d^2/dt^2\right) \nu(t) = dW(t)/dt, \quad \kappa > 0. \tag{8}$$

- Use full-rank basis expansion: $\nu(t) = \sum_{k=1}^{n} \psi_k(t)\gamma_k$.

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).
- Letting $\log(\lambda(t)) = \nu(t)$, consider SDE (Lindgren et al., 2011)

$$\left(\kappa^2 - d^2/dt^2\right)\nu(t) = dW(t)/dt, \quad \kappa > 0. \tag{8}$$

- Use full-rank basis expansion: $\nu(t) = \sum_{k=1}^{n} \psi_k(t)\gamma_k$.
- Letting $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)'$, then

$$\boldsymbol{\gamma} \mid \theta \sim N(0, \theta^{-1}\mathbf{R}^{-1}),$$

where $\mathbf{R} = \kappa^4\tilde{\mathbf{B}} - \kappa^2(\mathbf{H}' + \mathbf{H}) + \mathbf{H}'\tilde{\mathbf{B}}^{-1}\mathbf{H}$.

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).
- Letting $\log(\lambda(t)) = \nu(t)$, consider SDE (Lindgren et al., 2011)

$$\left(\kappa^2 - d^2/dt^2\right)\nu(t) = dW(t)/dt, \quad \kappa > 0. \tag{8}$$

- Use full-rank basis expansion: $\nu(t) = \sum_{k=1}^{n} \psi_k(t)\gamma_k$.
- Letting $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)'$, then

$$\boldsymbol{\gamma} \mid \theta \sim N(0, \theta^{-1}\mathbf{R}^{-1}),$$

  where $\mathbf{R} = \kappa^4\tilde{\mathbf{B}} - \kappa^2(\mathbf{H}' + \mathbf{H}) + \mathbf{H}'\tilde{\mathbf{B}}^{-1}\mathbf{H}$.
- Use κ as a tuning parameter.

Modeling $\lambda(t)$

- $\lambda(t) \geq 0$ is a continuous and 'friendly' smoothing function.
- The prior on $\lambda(t)$ must be proper (Yue et al., 2012).
- Letting $\log(\lambda(t)) = \nu(t)$, consider SDE (Lindgren et al., 2011)

$$\left(\kappa^2 - d^2/dt^2\right)\nu(t) = dW(t)/dt, \quad \kappa > 0. \tag{8}$$

- Use full-rank basis expansion: $\nu(t) = \sum_{k=1}^{n} \psi_k(t)\gamma_k$.
- Letting $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)'$, then

$$\boldsymbol{\gamma} \mid \theta \sim N(0, \theta^{-1}\mathbf{R}^{-1}),$$

where $\mathbf{R} = \kappa^4\tilde{\mathbf{B}} - \kappa^2(\mathbf{H}' + \mathbf{H}) + \mathbf{H}'\tilde{\mathbf{B}}^{-1}\mathbf{H}$.

- Use κ as a tuning parameter.
- Efficient MCMC algorithm for adaptive SDE v1.

Modeling $\lambda(t)$

What if we want to use INLA?

Modeling $\lambda(t)$

What if we want to use INLA?

- Let $\nu(t) = \log(\lambda(t))$ and use low-rank basis

$$\nu(t) = \sum_{k=1}^{m} \psi_k(t)\gamma_k.$$

Modeling $\lambda(t)$

What if we want to use INLA?

- Let $\nu(t) = \log(\lambda(t))$ and use low-rank basis

$$\nu(t) = \sum_{k=1}^{m} \psi_k(t)\gamma_k.$$

- The value of m should be small (≤ 15)

Modeling $\lambda(t)$

What if we want to use INLA?

- Let $\nu(t) = \log(\lambda(t))$ and use low-rank basis

$$\nu(t) = \sum_{k=1}^{m} \psi_k(t)\gamma_k.$$

- The value of m should be small (≤ 15)
- Gaussian prior on $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma} \sim N(0, \mathbf{R}^{-}).$$

Modeling $\lambda(t)$

What if we want to use INLA?

- Let $\nu(t) = \log(\lambda(t))$ and use low-rank basis

$$\nu(t) = \sum_{k=1}^{m} \psi_k(t)\gamma_k.$$

- The value of m should be small (≤ 15)
- Gaussian prior on $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma} \sim N(0, \mathbf{R}^{-}).$$

- Possible models for $\nu(t)$: P-splines, Fourier representation, etc.

MCMC for adaptive SDE v1

MCMC for adaptive SDE v1

- Recall $\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$.

MCMC for adaptive SDE v1

- Recall $\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$.
- $\mathbf{w}$ is a random walk model:

$$[\mathbf{w} \mid \delta, \boldsymbol{\gamma}] \propto \prod_{i=1}^{n} (\delta e^{\gamma_i})^{1/2} \exp\left(-\frac{\delta e^{\gamma_i}}{2} \tilde{w}_i^2\right),$$

where $\tilde{\mathbf{w}} = (0, \tilde{w}_2, \ldots, \tilde{w}_{n-1}, 0)' = \mathbf{H w}$.

MCMC for adaptive SDE v1

- Recall $\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$.
- $\mathbf{w}$ is a random walk model:

$$[\mathbf{w} \mid \delta, \boldsymbol{\gamma}] \propto \prod_{i=1}^{n} (\delta e^{\gamma_i})^{1/2} \exp\left(-\frac{\delta e^{\gamma_i}}{2} \tilde{w}_i^2\right),$$

  where $\tilde{\mathbf{w}} = (0, \tilde{w}_2, \ldots, \tilde{w}_{n-1}, 0)' = \mathbf{H w}$.
- Since γ_i depends on $\tilde{w}_i$ only, we construct an accurate GMRF approximation for the full conditional of $\boldsymbol{\gamma}$

$$\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta) \approx \mathcal{F}(\boldsymbol{\gamma} \mid \mathbf{w}, \delta, \theta) \propto [\mathbf{w} \mid \delta, \boldsymbol{\gamma}][\boldsymbol{\gamma} \mid \theta]$$

  where $\boldsymbol{\gamma}_0$ is the mode of $\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)$.

MCMC for adaptive SDE v1

- Recall $\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$.
- $\mathbf{w}$ is a random walk model:

$$[\mathbf{w} \mid \delta, \boldsymbol{\gamma}] \propto \prod_{i=1}^{n} (\delta e^{\gamma_i})^{1/2} \exp\left(-\frac{\delta e^{\gamma_i}}{2}\tilde{w}_i^2\right),$$

where $\tilde{\mathbf{w}} = (0, \tilde{w}_2, \ldots, \tilde{w}_{n-1}, 0)' = \mathbf{Hw}$.

- Since γ_i depends on $\tilde{w}_i$ only, we construct an accurate GMRF approximation for the full conditional of $\boldsymbol{\gamma}$

$$\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta) \approx \mathcal{F}(\boldsymbol{\gamma} \mid \mathbf{w}, \delta, \theta) \propto [\mathbf{w} \mid \delta, \boldsymbol{\gamma}][\boldsymbol{\gamma} \mid \theta]$$

where $\boldsymbol{\gamma}_0$ is the mode of $\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)$.

- Then, sample proposal $\boldsymbol{\gamma}^*$ and update $\boldsymbol{\gamma}$ with probablity

$$\alpha = \frac{\mathcal{F}(\boldsymbol{\gamma}^* \mid \mathbf{w}, \delta, \theta)\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)}{\mathcal{F}(\boldsymbol{\gamma} \mid \mathbf{w}, \delta, \theta)\mathcal{P}(\boldsymbol{\gamma}^* \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)}.$$

MCMC for adaptive SDE v1

- Recall $\mathbf{\Lambda H w} \sim N(\mathbf{0}, \tilde{\mathbf{B}})$.
- $\mathbf{w}$ is a random walk model:

$$[\mathbf{w} \mid \delta, \boldsymbol{\gamma}] \propto \prod_{i=1}^{n} (\delta e^{\gamma_i})^{1/2} \exp\left(-\frac{\delta e^{\gamma_i}}{2} \tilde{w}_i^2\right),$$

  where $\tilde{\mathbf{w}} = (0, \tilde{w}_2, \ldots, \tilde{w}_{n-1}, 0)' = \mathbf{H w}$.
- Since γ_i depends on $\tilde{w}_i$ only, we construct an accurate GMRF approximation for the full conditional of $\boldsymbol{\gamma}$

$$\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta) \approx \mathcal{F}(\boldsymbol{\gamma} \mid \mathbf{w}, \delta, \theta) \propto [\mathbf{w} \mid \delta, \boldsymbol{\gamma}][\boldsymbol{\gamma} \mid \theta]$$

  where $\boldsymbol{\gamma}_0$ is the mode of $\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)$.
- Then, sample proposal $\boldsymbol{\gamma}^*$ and update $\boldsymbol{\gamma}$ with probablity

$$\alpha = \frac{\mathcal{F}(\boldsymbol{\gamma}^* \mid \mathbf{w}, \delta, \theta)\mathcal{P}(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)}{\mathcal{F}(\boldsymbol{\gamma} \mid \mathbf{w}, \delta, \theta)\mathcal{P}(\boldsymbol{\gamma}^* \mid \boldsymbol{\gamma}_0, \mathbf{w}, \delta, \theta)}.$$

- Other parameters can be updated easily.

INLA for adaptive SDE v2

INLA for adaptive SDE v2

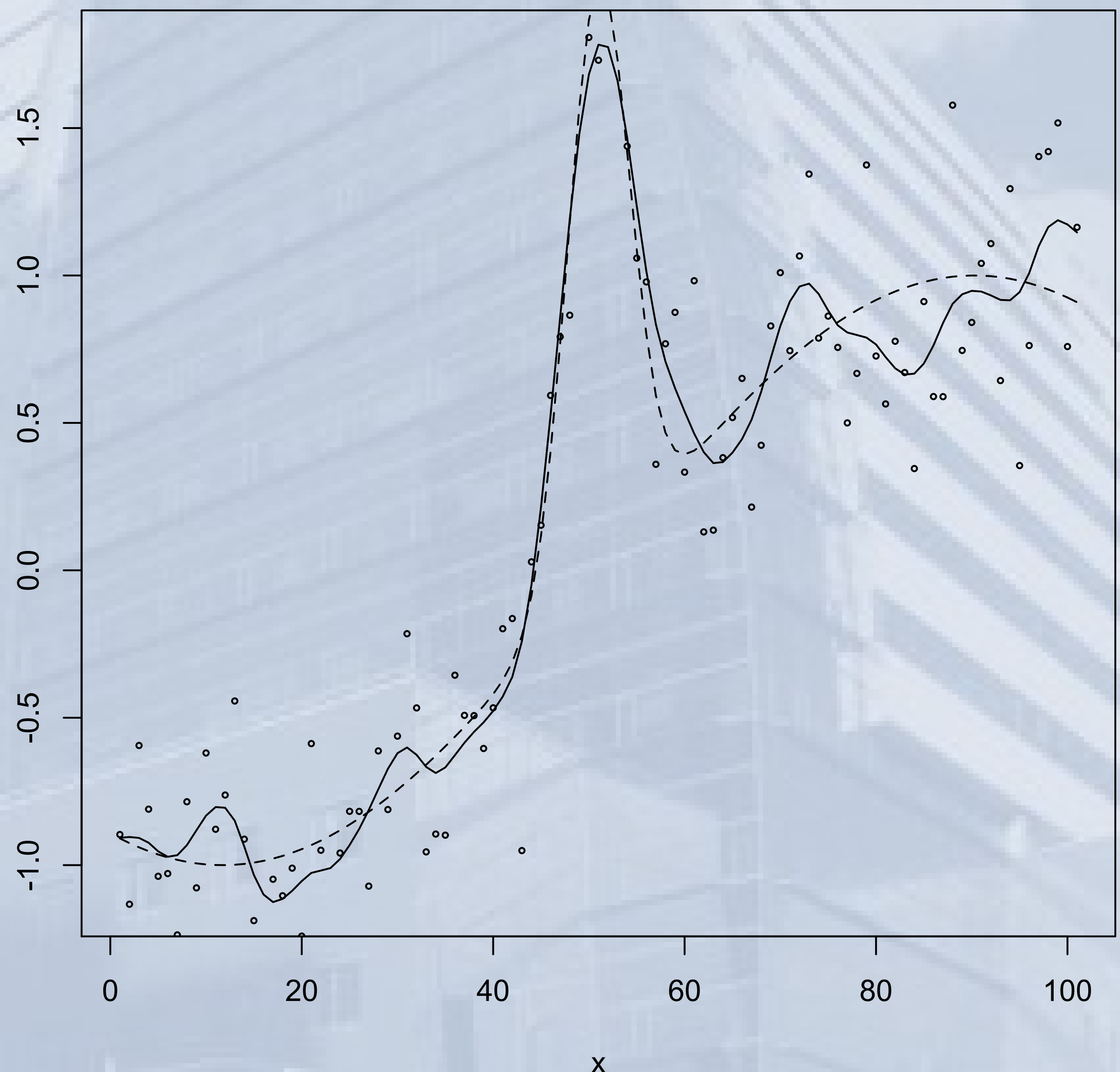
- Caution: expert's use only!

```
spde <- inla.spde2.generic(M0, M1, M2, B0, B1, B2,
         theta.mu, theta.Q, transform="identity", BLC = B0)
data <- list(y=y, x=x)
formula <- y~ -1 + f(x, model=spde)
result <- inla(formula, data=data, family=‘normal’)
fhat <- result$summary.random$x$mean
```

Simulation study

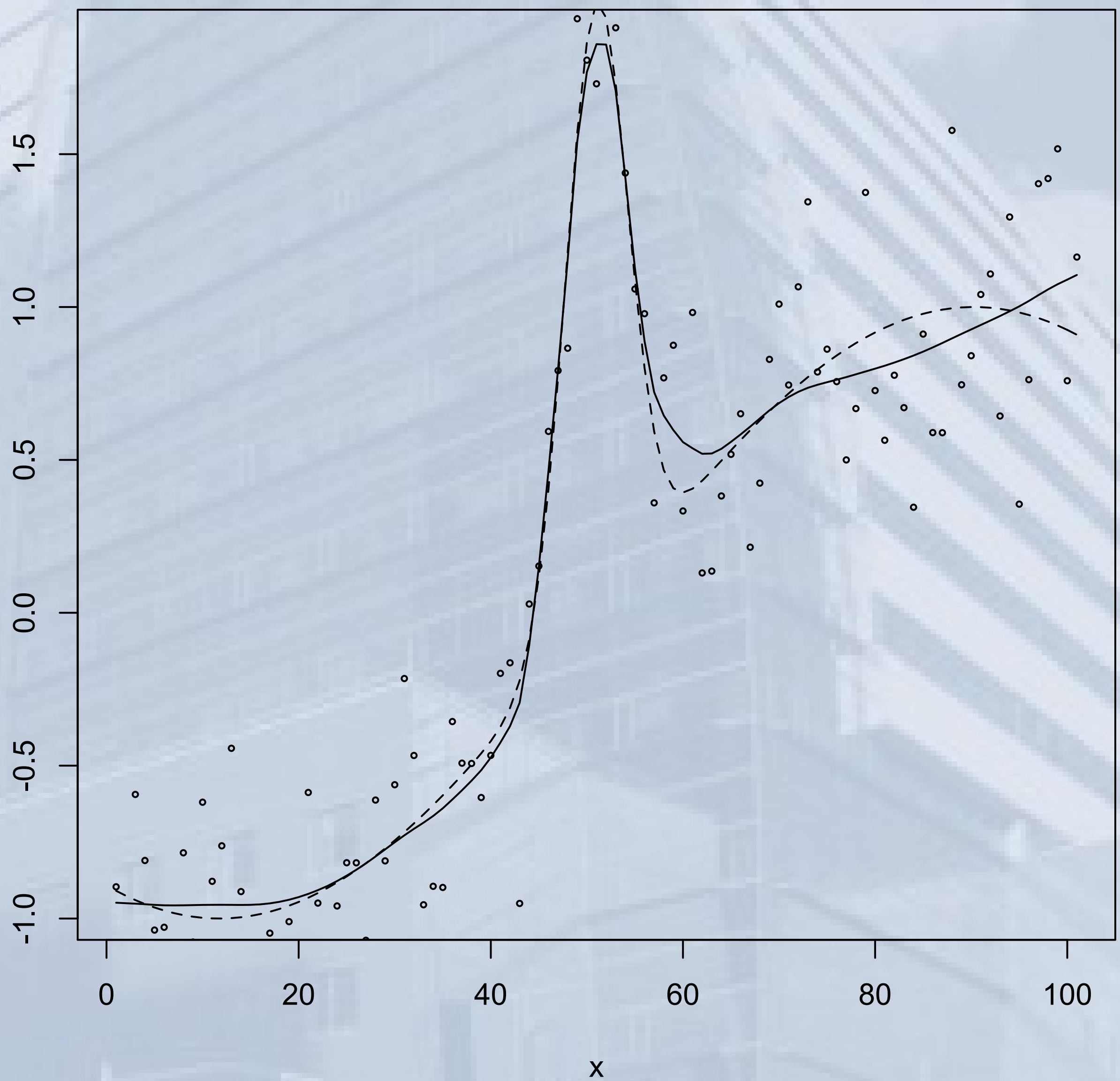
Simulation study

- Ordinary smoothing spline (use `smooth.spline()` in R)



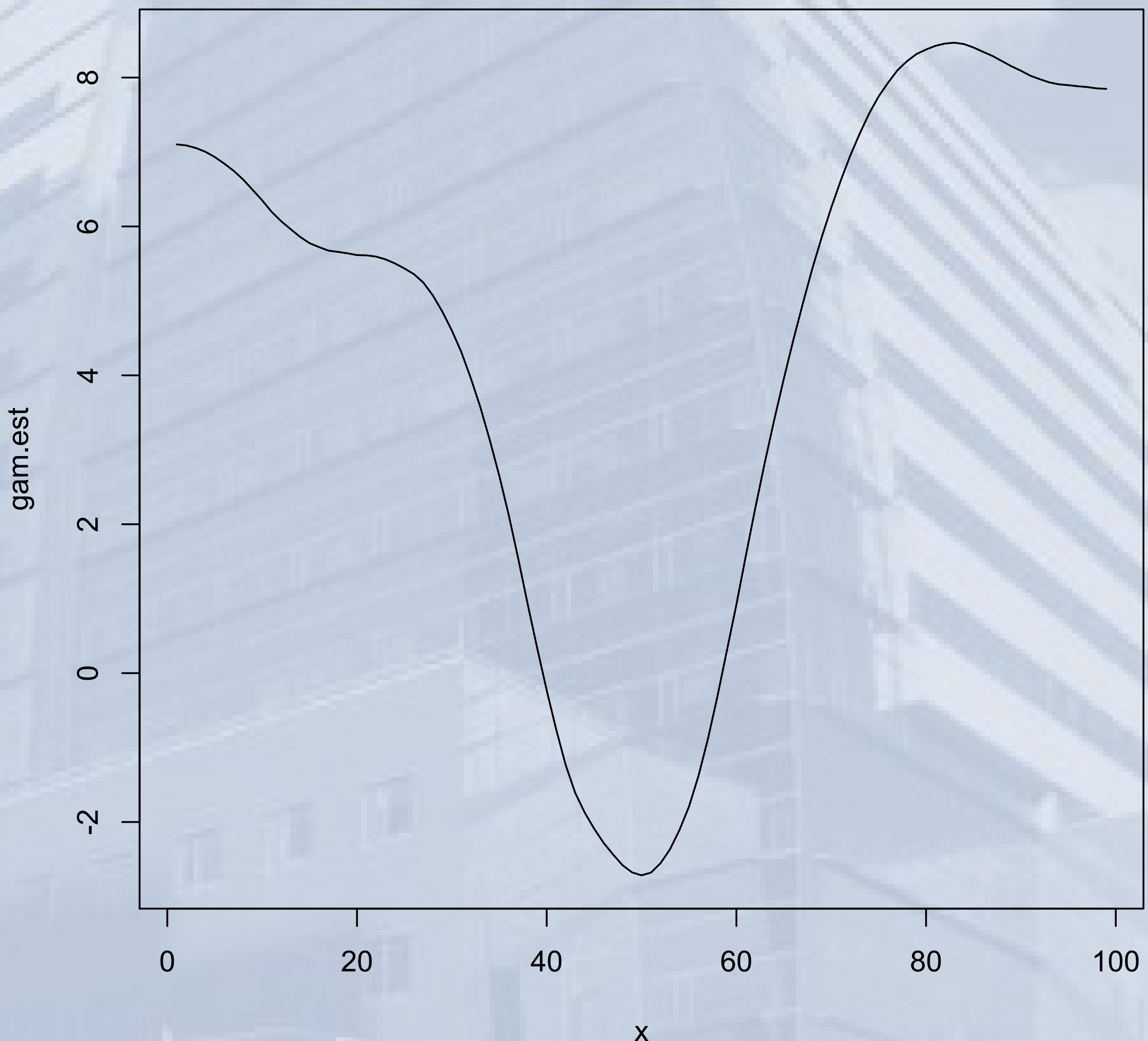
Simulation study

- BASS SDE v1 (use MCMC)



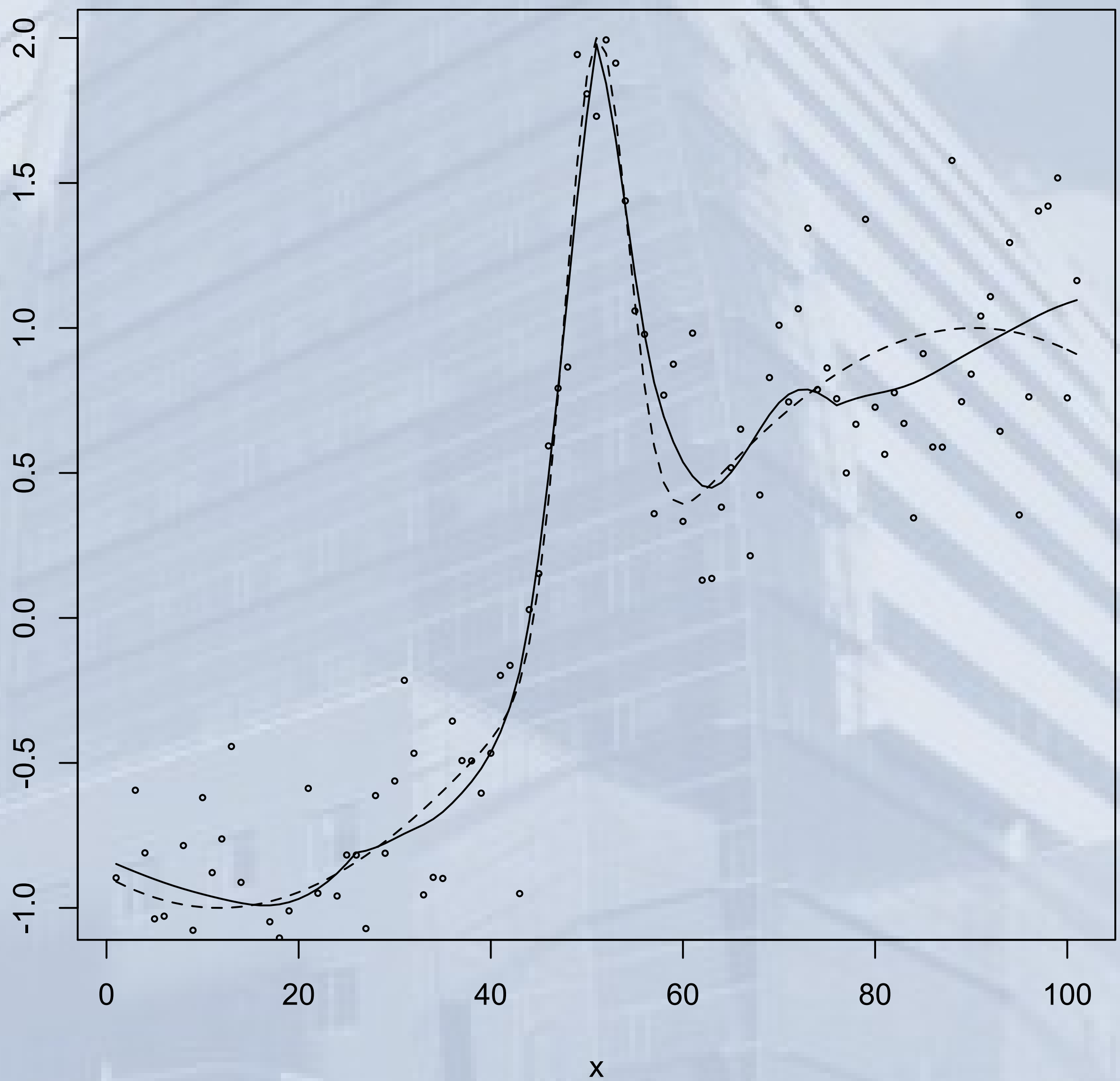
Simulation study

- BASS SDE v1 (use MCMC)



Simulation study

- BASS SDE v2 (use INLA)



Discussion

Discussion

- The model does not work for functions with sharp peaks or discontinuities. (See Bolin (2011) for the SDE with non-Gaussian noises.)

Discussion

- The model does not work for functions with sharp peaks or discontinuities. (See Bolin (2011) for the SDE with non-Gaussian noises.)
- A better MCMC algorithm?

Discussion

- The model does not work for functions with sharp peaks or discontinuities. (See Bolin (2011) for the SDE with non-Gaussian noises.)
- A better MCMC algorithm?
- Two dimensional extension: thin-plate splines.

Key references

- Yue, Lindgren, Simpson and Rue (2012). "Bayesian adaptive smoothing spline using stochastic differential equations." *(In progress.)*
- Yue, Speckman and Sun (2012). "Priors for Bayesian adaptive spline smoothing." *Annals of the Institute of Statistical Mathematics* **64**(3), 577-613.
- Lindgren, Rue, and Lindström (2011). "An explicit link between gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion)." *Journal of the Royal Statistical Society: Series B* **73**, 423-498.
- Lindgren and Rue (2008). "On the second-order random walk model for irregular locations." *Scandinavian Journal of Statistics* **35**, 691-700.