Scaling of MCMC in Singular Parameter Limits of OU Processes

David White

Friday 20th March 2009

Acknowledgments

- This work undertaken within WMI from Sept 06 to date.
- Computational resources provided by the Warwick Computer Science Dept (Simon Hammond).
- Thanks to Andrew Stuart, Gareth Roberts, Jochen Voss and Frank Pinski.

The aim of this work is to understand how to sample measures efficiently defined on function spaces of the form:

$$\frac{d\mu}{d\mu_0}(x) \propto \exp\left(-\Phi\left(x;k\right)\right)$$

where Φ has singular dependence as $k \to \infty$.

$\mu_0 = \mathcal{N}\left(0, C\right)$ where $L = -C^{-1}$.

Outline

Problem Specification

The OU Bridge Problem

$$\begin{aligned} \text{SDE}: \quad & \frac{dx}{du} = -kx(u) + \sigma \frac{dW}{du} \\ \text{Initial Condition}: \quad & x(0) = x^- \\ \text{Final Condition}: \quad & x(\ell) = x^+ \end{aligned}$$

- Three parameters may be varied, k, σ and ℓ.
- Fix $\ell = 1$, $\sigma = 1$ and $x^- = x^+ = 0$.
- The single parameter k controls the difficulty of the sampling problem.

Problem Specification

The OU Bridge Problem

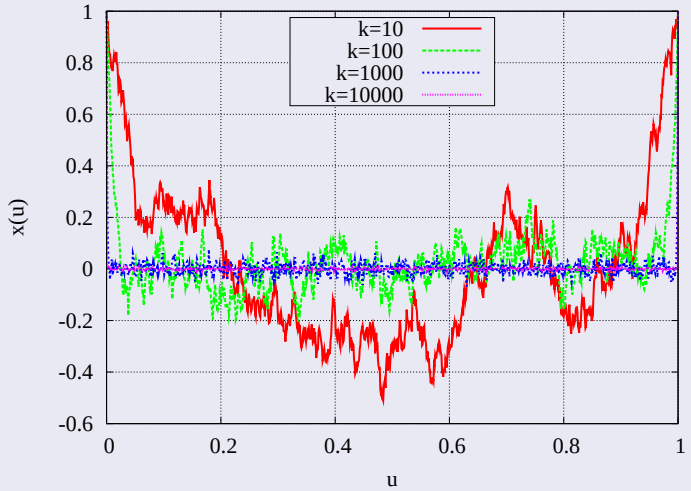
$$\text{SDE}: \quad \frac{dx}{du} = -kx(u) + \sigma \frac{dW}{du}$$

$$\text{Initial Condition}: \quad x(0) = x^-$$

$$\text{Final Condition}: \quad x(\ell) = x^+$$

- Three parameters may be varied, k, σ and ℓ.
- Fix $\ell = 1$, $\sigma = 1$ and $x^- = x^+ = 0$.
- The single parameter k controls the difficulty of the sampling problem.

Example Paths



Simpliest Proposal - Brownian Bridge Interpretation

$$y(u) = \left(1 - \beta^2\right)^{\frac{1}{2}} x(u) + \beta b(u)$$

- $x(u)$ is the current function.
- $y(u)$ is the proposed function.
- $b(u)$ is a Brownian bridge with zero endpoints.
- $\beta \in (0, 1)$ controls expected variation between x and y in each step.

Simpliest Proposal - SPDE Interpretation

SPDE with correct invarient measure and Dirichlet Boundary conditions $x(t,0) = x^-$ and $x(t,\ell) = x^+$

$$\text{SPDE Proposal}: \quad \frac{\partial x}{\partial t} = -x + \sqrt{2C}\frac{\partial w}{\partial t}$$

$$\text{FD Discretisation}: \quad \frac{y-x}{\Delta t} = -\frac{x+y}{2} + \sqrt{\frac{2C}{\Delta t}}\xi$$

$$\text{Useful Form}: \quad y = \frac{2-\Delta t}{2+\Delta t}x + \frac{\sqrt{8\Delta t}}{2+\Delta t}\sqrt{C}\xi$$

- $\sqrt{C}\xi$ corresponds to Brownian bridge.
- Δt controls the variation of the solution from the proposal.
- $\beta = \frac{\sqrt{8\Delta t}}{2+\Delta t}$

SPDE Proposals

$$\text{Std MALA}: \quad \frac{\partial x}{\partial t} = Lx - k^2 x + \sqrt{2}\frac{\partial w}{\partial t}$$

$$\text{Prec MALA}: \quad \frac{\partial x}{\partial t} = -x - k^2 Cx + \sqrt{2C}\frac{\partial w}{\partial t}$$

$$\text{Std RWMH}: \quad \frac{\partial x}{\partial t} = Lx + \sqrt{2}\frac{\partial w}{\partial t}$$

$$\text{Prec RWMH}: \quad \frac{\partial x}{\partial t} = -x + \sqrt{2C}\frac{\partial w}{\partial t}$$

MCMC Proposals

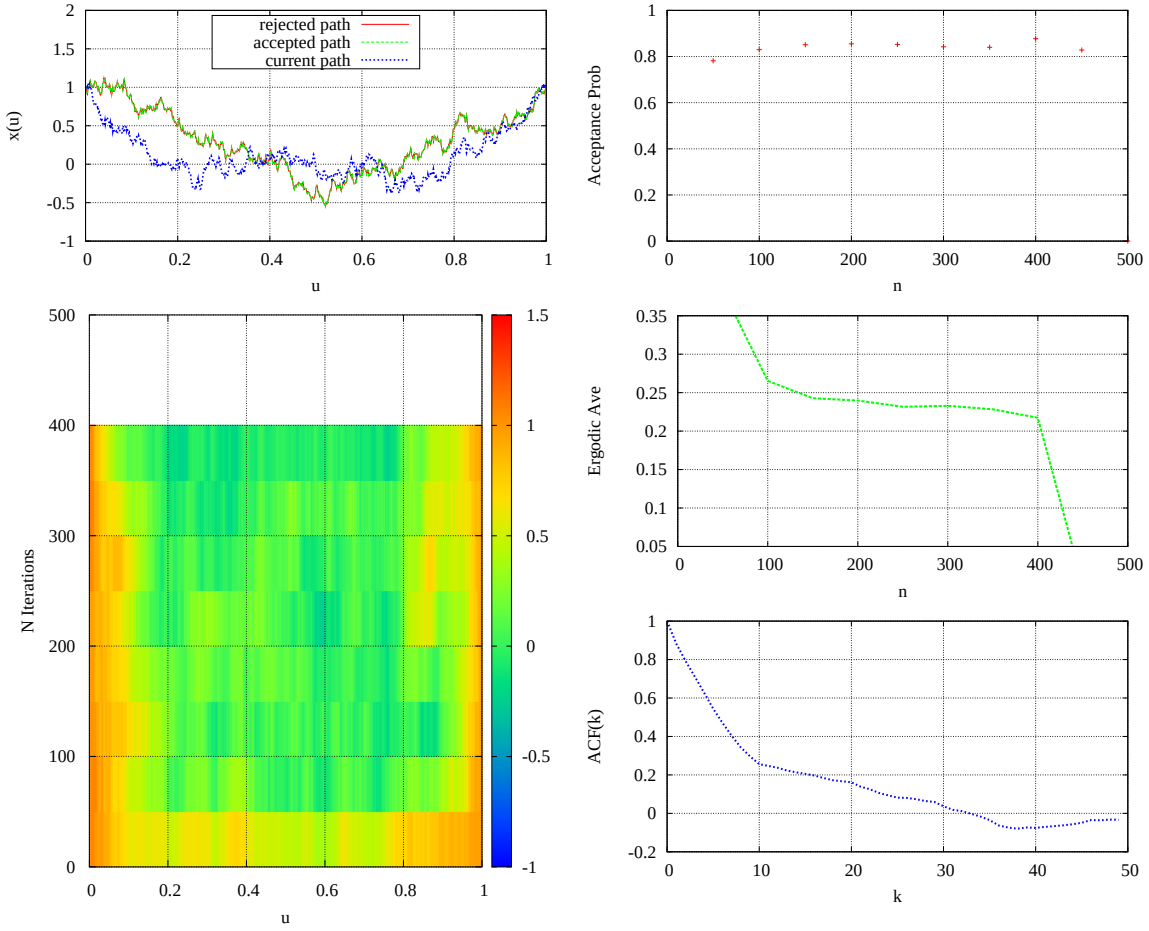
$$\text{Std MALA}: \quad (2-\Delta t L)\,y = \left(2+\Delta t L - 2k^2\Delta t\right)x + \sqrt{8\Delta t}\,\xi$$

$$\text{Prec MALA}: \quad (2+\Delta t)\,y = \left(2-\Delta t - 2k^2\Delta t C\right)x + \sqrt{8\Delta t C}\,\xi$$

$$\text{Std RWMH}: \quad (2-\Delta t L)\,y = (2+\Delta t L)\,x + \sqrt{8\Delta t}\,\xi$$

$$\text{Prec RWMH}: \quad (2+\Delta t)\,y = (2-\Delta t)\,x + \sqrt{8\Delta t C}\,\xi$$

Algorithm Demo



Outline

Efficiency Measures

Average Acceptance Probability

- Easy to measure.
- Widely used rule of thumb.
- No idea which value is optimal.

Efficiency Measures

Average Acceptance Probability

- Easy to measure.
- Widely used rule of thumb.
- No idea which value is optimal.

Autocorrelation Functions

$$\mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} h\left(X_n\right)\right) = \frac{\sigma^2}{N}\left(1 + 2\sum_{n=1}^{N}\frac{N-n}{N}\frac{\mathrm{Cov}\left(h\left(X_m\right), h\left(X_{m+n}\right)\right)}{\mathrm{Var}\left(h\left(X_m\right)\right)}\right)$$

- Difficult to measure accurately.
- Requires a huge numbers of samples for large lags.
- Proved that smallest IACF implies best sampler.
- May vary for different statistics $h\left(X_m\right)$.

Outline

Theoretical Acceptance Probability

Aim of the following calculation to choose Δt in accordance with k so that $\mathbb{E}[\alpha] \sim \mathcal{O}(1)$ as $k \to \infty$.

If $\lim_{k\to\infty} \mathbb{E}[\alpha] = 0$ or $\lim_{k\to\infty} \mathbb{E}[\alpha] = 1$ then the sampler is performing sub-optimally.

Theoretical Acceptance Probability

Lemma

The Metropolis-Hastings acceptance probability α may be written in the form as $\alpha(x, y) = \exp(\min(0, R(x, y)))$. The following bounds apply to the expectation of α $\forall \lambda > \sqrt{\mathbb{E}[R^2]}$ and $\gamma < 0$.

1. $\mathbb{E}[\alpha] \geq \exp(-\lambda)\left(1 - \frac{\mathbb{E}[R^2]}{\lambda^2}\right)$
2. $\mathbb{E}[\alpha] \leq \mathbb{P}[R \geq \gamma] + \exp(\gamma)$

Theoretical Acceptance Probability

Lemma

The Metropolis-Hastings acceptance probability α may be written in the form as $\alpha(x,y) = \exp(\min(0, R(x,y)))$. The following bounds apply to the expectation of α $\forall \lambda > \sqrt{\mathbb{E}[R^2]}$ and $\gamma < 0$.

1. $\mathbb{E}[\alpha] \geq \exp(-\lambda)\left(1 - \frac{\mathbb{E}[R^2]}{\lambda^2}\right)$
2. $\mathbb{E}[\alpha] \leq \mathbb{P}[R \geq \gamma] + \exp(\gamma)$

Preconditioned RWMH

$$
\begin{aligned}
R(x,y) &= \frac{4k^2\Delta t}{(2+\Delta t)^2}\int_0^\ell x^2 du - \frac{4k^2\Delta t}{(2+\Delta t)^2}\int_0^\ell \left(\sqrt{C}\xi\right)^2 du \\
&- k^2\sqrt{8\Delta t}\frac{2-\Delta t}{(2+\Delta t)^2}\int_0^\ell x\sqrt{C}\xi du
\end{aligned}
$$

$\mathbb{E}[R]$

$\mathbb{E}[R]$

Use the KL basis to describe bridges with zero endpoints.

$$\varphi^{(p)}(u) = \sqrt{\frac{2}{\ell}} \sin\left(\frac{\pi p u}{\ell}\right)$$

In this basis x and $\sqrt{C}\xi$ have the expansions.

$$x(u) = \sum_{p=1}^{\infty} \frac{\chi_p}{\sqrt{\lambda^{(p)} + k^2}} \varphi^{(p)} \text{ where } \chi_p \sim \mathcal{N}(0,1)$$

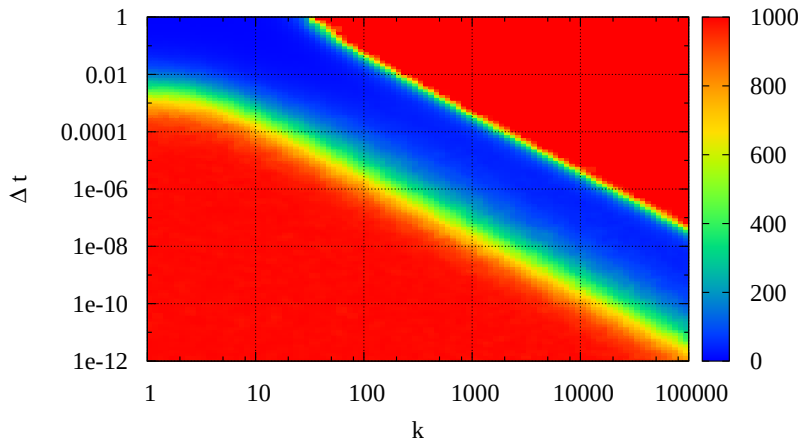
$$\sqrt{C}\xi(u) = \sum_{p=1}^{\infty} \frac{\xi_p}{\sqrt{\lambda^{(p)}}} \varphi^{(p)} \text{ where } \xi_p \sim \mathcal{N}(0,1)$$

$$\mathbb{E}[R] = \frac{2\Delta t}{(2+\Delta t)^2}\left(\frac{k\ell}{\tanh(k\ell)} - 1\right) - \frac{2}{3}\frac{k^2\ell^2\Delta t}{(2+\Delta t)^2}$$

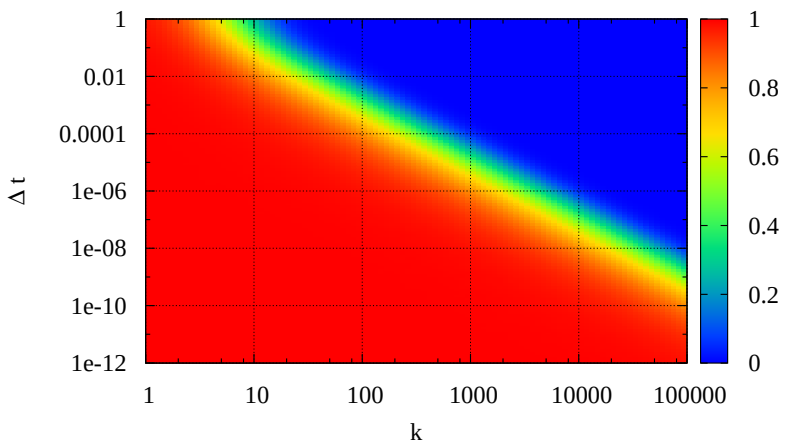
Outline

Numerical Results - Preconditioned RWMH

IACF

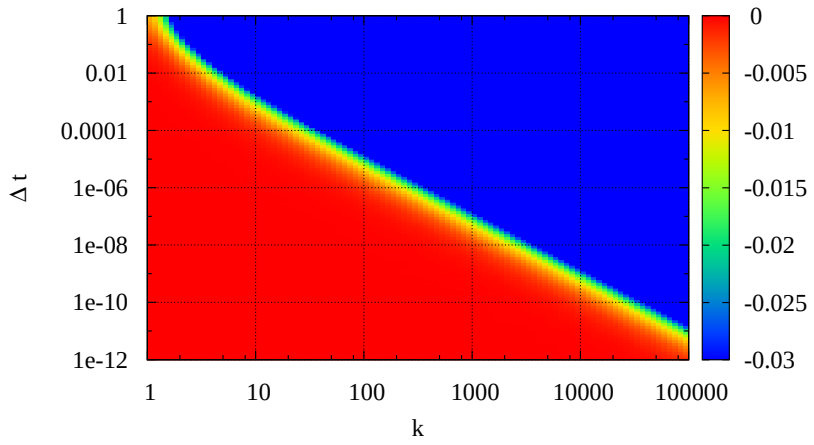


Acceptance Probability

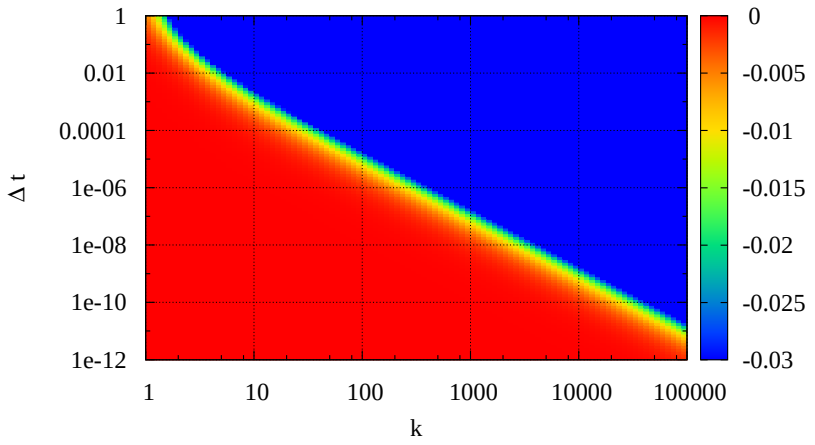


Numerical Results - Preconditioned RWMH

Numerical R

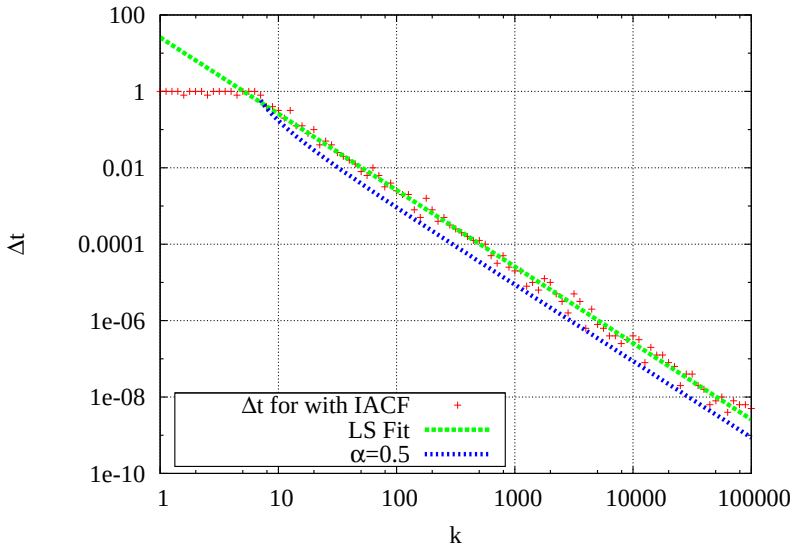


Theoretical R



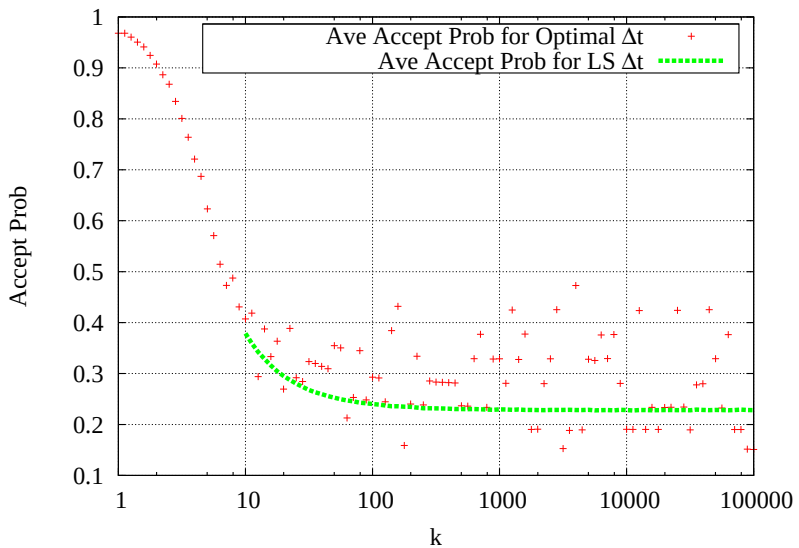
Numerical Results - Preconditioned RWMH

Optimal Δt scaling



Gradient=-2.00027

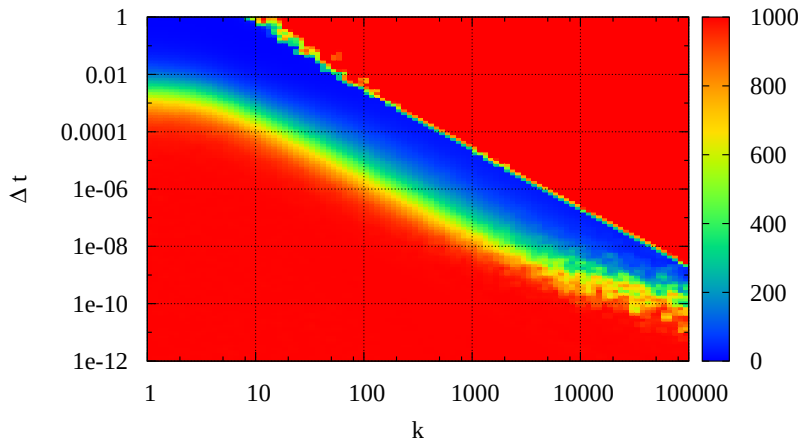
Optimal Acceptance Probability



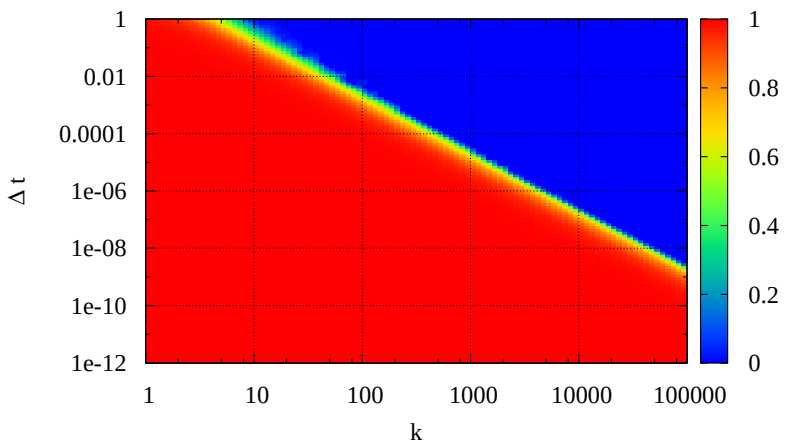
Limit=0.228

Numerical Results - Preconditioned MALA

IACF

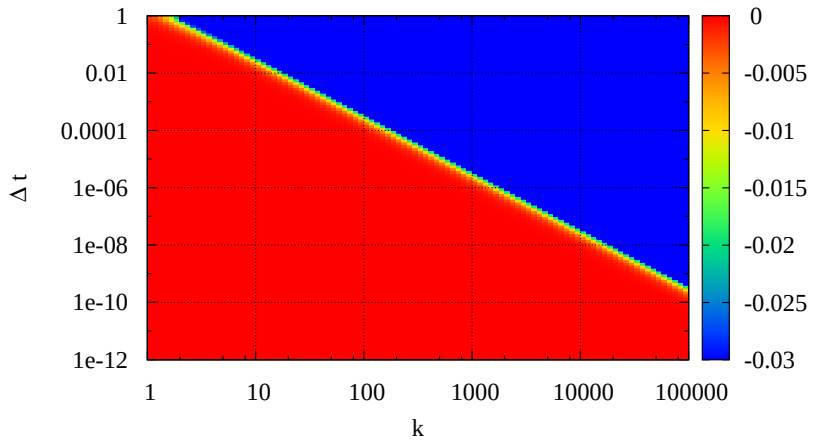


Acceptance Probability

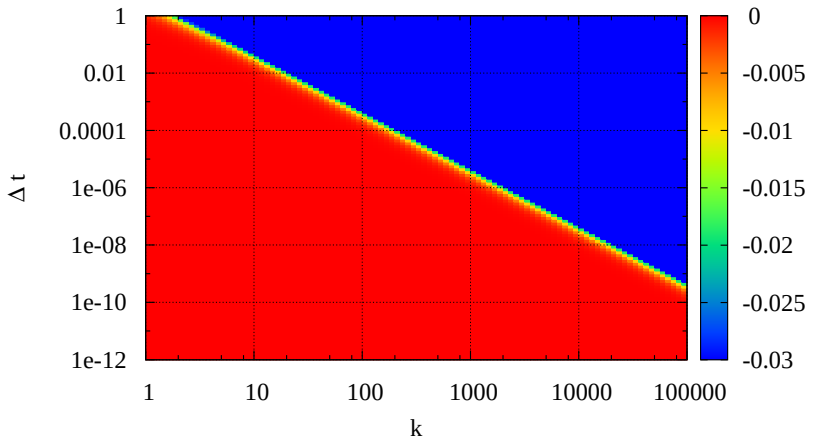


Numerical Results - Preconditioned MALA

Numerical R

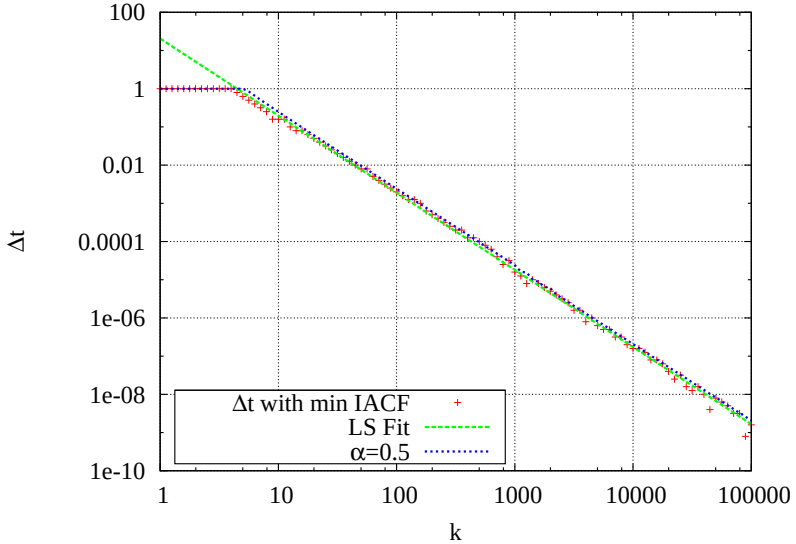


Theoretical R



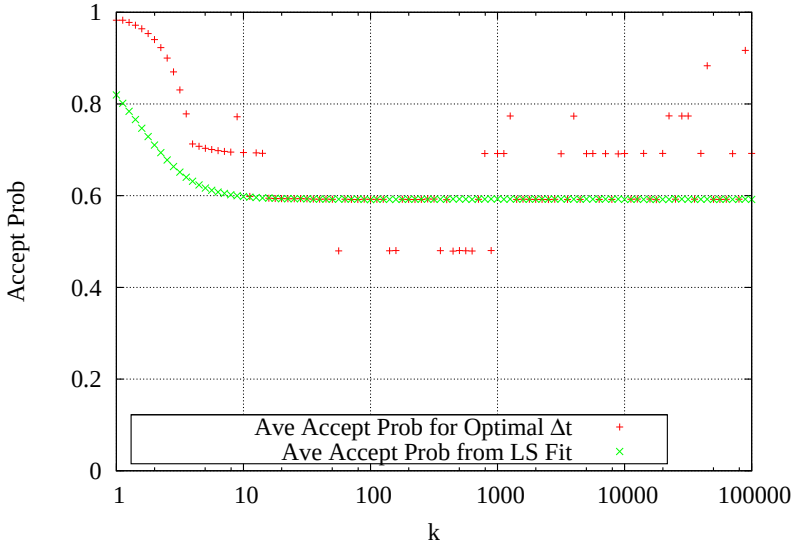
Numerical Results - Preconditioned MALA

Optimal Δt scaling



Gradient$=$ 2.0108

Optimal Acceptance Probability



Limit=0.591

Conclusions

- For this problem the preconditioned MALA is the best method.
- The standard RWMH is the worst method.
- Numerical results suggest that the optimal acceptance probabilities are 0.228 and 0.591 for RWMH and MALA respectively.

Thanks for Listening
Any questions/comments?