Stability of Partially Implicit Langevin Schemes and Their MCMC Variants

Bruno Casella
McKinsey & Company
Milan, Italy
e-mail: Bruno_Casella@mckinsey.com

Gareth Roberts
Department of Statistics
University of Warwick, UK
e-mail: gareth.o.roberts@warwick.ac.uk

Osnat Stramer*
Department of Statistics
and Actuarial Science
University of Iowa, IA, USA
e-mail: stramer@stat.uiowa.ed

September 8, 2009

Abstract

A broad class of implicit or partial implicit time discretizations for the Langevin diffusion are considered and used as proposals for the Metropolis-Hastings algorithm. Ergodic properties of our proposed schemes are studied. We show that introducing implicitness in the discretization leads to a process that often inherits the convergence rate of the continuous time process. These contrast with the behavior of the naive or Euler-Maruyama discretization, which can behave badly even in simple cases. We also show that our proposed chains, when used as proposals for the Metropolis-Hastings algorithm, preserve geometric ergodicity of their implicit Langevin schemes and thus behave better than the local linearization of the Langevin diffusion. We illustrate the behavior of our proposed schemes with examples. Our results are described in detail in one dimension only, although extensions to higher dimensions are also described and illustrated.

KEY WORDS: Langevin diffusions; ergodicity; implicit Euler schemes: discrete approximation

*To whom correspondence should be addressed

AMS Subject Classifications: 60J60, 60Jxx

1 Introduction

This paper concerns the use of implicit discretisation methods to improve the performances of the Langevin sampling. As generic Markov chain Monte Carlo (MCMC) simulation tools, Langevin algorithms were originally proposed by Doll et al. (1978). Our interest is motivated by applications in Bayesian statistics when we need to simulate from a posterior distribution whose normalising constant cannot be computed exactly. In this context, Langevin methods were popularised by Besag (1994) and Roberts and Tweedie (1996a). These methods have proved successful in many areas of statistical application such as in spatial statisics (see for example Christensen et al., 2006), typically giving rise to much more rapid mixing than vanilla methods such as the Metropolis-Hastings Random Walk (RWM) algorithm. However, Langevin methods are often less stable than their simpler competitors. For instance Roberts and Tweedie (1996a) demonstrate that the basic Langevin algorithm fails to be geometrically ergodic on light tailed target densities.

Within the context of ordinary differential equations, implicit methods have been shown to be more stable than traditional methods (see for example Stuart and Humphries, 1996). Our purpose here is to demonstrate similar results in the stochastic setting. We study the stability of partial implicit Langevin algorithms, as well as stability of their Metropolis-Hastings algorithms.

1.1 The problem

Suppose we want to simulate from a continuous density function π on $\mathbf{R}^m$ which we know only up to a constant factor k, that is we know the unnormalised function $\pi_u = k\,\pi$. Our schemes are based on Langevin diffusions which are constructed so that in continuous time it converges to π. The Langevin m-dimensional diffusion process $l = \{l_t : t \geq 0\}$ is defined as a solution to the stochastic differential equation (SDE):

$$dl_t = \frac{1}{2}\nabla \log \pi\,(l_t)\,dt + dW_t; \qquad l_0 = a \tag{1}$$

where ∇ denotes the usual gradient differential operator, and $W = \{W_t : t \geq 0\}$ is an m-dimensional standard Brownian motion. Under appropriate non-explosivity conditions, π is the unique ergodic measure of the process $l = \{l_t\}_{t\geq 0}$. Therefore, a natural way to simulate from π is to reproduce the long time behavior of l. Unfortunately, direct simulation from (1) is usually infeasible since we don't have an explicit expression for the transition law of the Langevin diffusion (though see Beskos et al., 2008a, for relatively small dimensional problems).

We therefore consider a discrete-time approximation $L := \{L_n : n \in \mathbb{N}\}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$, of the Langevin diffusion process l with step-size δ. In many situations, the discretisation scheme L inherits desirable stability properties from its parent diffusion, at least for sufficiently small discretisation intervals δ. In this case, its invariant measure will typically be close to, but not exactly given by, the target density π. To address this issue, a natural strategy is to supplement the discretisation scheme with a Metropolis-Hastings rejection step which enforces the correct invariant distribution.

We use the following terminology. We call the *unadjusted Langevin scheme* L, and the *Metropolis adjusted* or *Metropolis-Hastings Langevin scheme* $M = \{M_n : n \in \mathbb{N}\}$. M is generated by a Metropolis-Hastings algorithm employing the Langevin scheme L as a proposal. Otherwise we will refer to both generically as Langevin schemes.

1.2 The motivating example

In its standard form, the Langevin algorithm used by Roberts and Tweedie (1996a) proposes to approximate l by the δ-step Euler discretisation. This assumes that the drift function $\frac{1}{2}\nabla \log \pi(\cdot)$ is constant on small time intervals of length $\delta > 0$ and leads to the discrete time chain

$$L_{n+1} = L_n + \frac{1}{2}\nabla \log \pi(L_n)\delta + \sqrt{\delta}\xi_{n+1}, \qquad \xi_{n+1} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, I); \quad n \in \mathbb{N} \tag{2}$$

where $\mathcal{N}(0, I)$, I is the identity matrix, denotes the law of the m-dimensional standard normal random vector. As in Roberts and Tweedie (1996a), we term the Unadjusted Langevin Algorithm L (2) as ULA and its Metropolis-Hastings version M as MALA (Metropolis-Adjusted Langevin Algorithm). Under some regularity conditions, ULA inherits the ergodic behavior of the diffusion l when the tail of the target distribution is approximately Gaussian but looses this nice property for lighter tails.

This happens for example when π belongs to the following class of one-dimensional densities:

$$\mathcal{A}_d' = \left\{\pi : \pi(x) \propto e^{-\gamma|x|^{d+2}};\ x \in \Re, \gamma > 0\right\}, \quad d \geq 0 \tag{3}$$

As proved in Roberts and Tweedie (1996a), the Langevin diffusion (1) is always π-geometrically ergodic, in contrast to ULA that is transient for $d > 0$ or $d = 0$ with time interval $\delta > 1/\gamma$. Moreover, the transience of ULA leads to unstable behavior of MALA. Figures 1 and 2, provide simple graphical illustration of these behaviors. We consider the model $d\,l_t = -2\,l_t^3\,dt + dW_t$ corresponding to the target density $\pi(x) \propto \exp\left(-x^4\right)$. Figure 1 displays an exact trajectory of the diffusion l with a starting point $l_0 = 5$. The trajectory was produced by implementing the recent Exact Algorithm 3 (Beskos et al., 2008a) on a very fine discrete grid ($\delta = 0.01$).

[Figure 1 about here.]

The diffusion trajectory returns rapidly to the mode 0 when in the tails and tends to stay around 0. Figure 2 displays the ULA schemes L with discretisation interval $\delta = 0.1$ and starting point $L_0 = 0$ (*plot A*) and $L_0 = 5$ (*plot B*), the MALA scheme corresponding to the ULA scheme in plot B (*plot C*) and the Metropolis-Hastings Random Walk (RWM) with variance 0.1 and starting point 5 (*plot D*).

[Figure 2 about here.]

It is clear that, unlike the Langevin trajectory, ULA becomes explosive as we move from the center ($L_0 = 0$) to the tails ($L_0 = 5$). Notice that this transient behavior can be reproduced for any $\delta > 0$: i.e. for any $\delta > 0$, we can find a value L^* such that, for any $|L_0| > |L^*|$ the chain will explode with high probability. In this context the Metropolis-Hastings mechanism (designed to guarantee π-stationarity) is likely to reject the proposed moves and therefore, typically the MALA chain gets "stuck" (Figure 2, plot C). Notice that the RWM (Figure 2, plot D) performs much better: remarkably the use of a more "naive" proposal would allow us to avoid the stability problems arising from the Euler discretisation. In fact for $d > 0$, the RWM algorithm is geometrically ergodic (Mengersen and Tweedie, 1996, Theorem 3.2) in contrast to the MALA scheme that is not geometrically ergodic (Roberts and Tweedie, 1996a, Theorem 4.3).

1.3 Scope of the paper

The loss of ergodicity under the Euler discretisation occurs for essentially the same reason that the standard Euler method is unstable for stiff dissipative ODEs. In the deterministic context, the problem is cured by introducing implicitness in the discretisation. In this paper we show that the same remedy works in the context of Langevin sampling.

We study the theoretical properties of different versions of the partial implicit scheme. We show that for $\pi \in \mathcal{A}'_d,\ d \geq 0$, unlike ULA and MALA, our proposed schemes have robust stability properties. We shall demonstrate this by concentrating on the derivation of a geometric

Foster-Lyapunov drift inequality. These inequalities control the stability of excursions to the tails of the target distribution. Together with appropriate non-pathological properties in the "center" of the space, the drift condition implies geometric convergence to stationarity, and we shall generally couch our results in this way. We give results for both the unadjusted and Metropolis-adjusted cases. Whilst we work initially within the family $\mathcal{A}'_d,\ d \geq 0$, eventually we extend the approach to encompass a much larger class of target distributions.

The paper is organised as follows. In Section 2 we introduce different implicit Langevin algorithms. In Section 3 after introducing the basic notation, we state the results on the ergodic properties of partial implicit Langevin schemes for $\pi \in \mathcal{A}'_d,\ d \geq 0$. In Section 4 we firstly introduce the general Metropolis-Hastings algorithms and then study the ergodic properties of the Metropolis Adjusted Implicit Langevin Algorithm. We then prove that we can avoid pathological behavior of these algorithms by adjusting the Langevin schemes. In Section 5 we describe how to construct flexible geometrically ergodic Langevin algorithms to deal with more general target densities. We finish in Section 6 with some concluding remarks and directions for future research.

2 Partial Implicit Langevin Algorithms

We consider target densities $\pi \in \mathcal{A}'_d,\ d \geq 0$ as defined in (3). Within this setting, we can write explicitly the SDE of the Langevin diffusion (1):

$$dl_t = \frac{1}{2}\, k_d \,\mathrm{sign}\,(l_t)\, |l_t|^{\,d+1}\, dt + dW_t; \qquad k_d \;:=\; -\gamma(d+2),\ d \geq 0. \tag{4}$$

We defined our first Unadjusted Partial Implicit Langevin Algorithm (UPILA1) as the partial implicit Euler discretization scheme for the diffusion in (4):

$$\begin{aligned} \text{UPILA1:}\quad L_{n+1} \;&=\; L_n + \frac{1}{2}k_d \left(\theta\, \mathrm{sign}(L_{n+1})\, |L_{n+1}|^{d+1} \;+\; \hat{\theta}\, \mathrm{sign}(L_n)\, |L_n|^{d+1}\right) \delta \\ &+\; \sqrt{\delta}\, \xi_{n+1}, \end{aligned} \tag{5}$$

where $\hat{\theta} = 1 - \theta$ and $\{\xi_{n+1}, n \in \mathbb{N}\}$ are i.i.d random variables with zero mean, unit variance and positive, continuous density (with respect to Lebesgue measure μ^{Leb}) on $\Re$. Implicit Euler schemes for a multidimensional diffusion are defined in Kloeden and Platen (1992), Section 15.4. The crucial feature of (5) is the introduction of a parameter θ which controls the relative weight of the implicit component in the discretisation of the drift and can be interpreted as a measure of the "degree of implicitness". Schemes of this type are also called stochastic θ-schemes.

In fact (5) is the stochastic analogous of the well known θ-method used in the deterministic context. In particular the special cases $\theta = 0$, $\theta = 0.5$, and $\theta = 1$ give respectively the explicit Euler scheme, the stochastic generalization of the trapezoidal method and the backward Euler method. Stability properties of the stochastic system (5) have been investigated in Saito and Mitsui (1996) and Higham (2000) in the linear case ($d = 0$). Note that for $d \neq 0$ and $\theta \neq 0$, simulation from (5) implies the inversion of the non-linear function

$$F(u) = u - \frac{1}{2}\,\theta\, k_d\, \delta\, \mathrm{sign}(u)\, |u|^{d+1} \tag{6}$$

at each iteration of the Markov chain. This can be easily done for the one-dimensional case but can be a substantial limiting factor in view of high-dimensional applications of UPILA. In this perspective it is more realistic to resort to linearly partial implicit discretisation schemes. The main idea is to decompose the drift function $\nabla \log \pi(u)$ in a linear and a non-linear component:

$$\nabla \log \pi(u) = k_d \mathrm{sign}\,(u)\, |u|^{\,d+1} = k_d\, |u|^d\, u$$

and then to treat explicitly the non-linear component and partially implicitly the linear one. The idea of incorporating the implicitness parameter θ only in the linear part of the drift for Langevin diffusions is in Beskos et al. (2008b). We consider the following Unadjusted Partial Implicit Langevin Algorithms:

$$\text{UPILA2:} \quad L_{n+1} = L_n + \frac{1}{2}\, k_d\, |L_n|^d \left(\theta L_{n+1} + \hat{\theta} L_n\right) \delta + \sqrt{\delta}\xi_{n+1}.$$

UPILA2 introduces implicitness in the discretisation and preserves an explicit characterisation of the chain's dynamics with obvious computational advantages. UPILA2 is exponentially ergodic (see Section 3) under some conditions on the implicit parameter θ. However, since the variance of UPILA2 tends to 0 when $d > 0$ as $|x| \to +\infty$, it tends to propose a deterministic-like move toward the center when in the tails and thus its Metropolis adjusted chain may loose this nice geometric rate of convergence for target distributions with Gaussian or lighter tails. The use of the two-steps discretisation strategy (sometimes called the split-step method) is a way to overcome this difficulty when $d > 0$. We refer to this scheme as UPILA3. It splits the discretisation problem into two stages as follows.

$$\text{UPILA3 } (d > 0): \qquad \begin{aligned} L_* &= L_n + \frac{1}{2}\, k_d\, |L_n|^d \left(\theta L_* + \hat{\theta} L_n\right) \delta \\ L_{n+1} &= L_* + \sqrt{\delta}\xi_{n+1} \end{aligned}$$

In the first stage a linear implicit discretisation to the drift component of (1) is employed. At this stage the numerical problem is equivalent to the approximation of an ODE since no randomness is involved. In the second stage, the resulting dynamic system is perturbed by a Gaussian noise which accounts for the effects of the Wiener component in (1). An example of a discretisation scheme using the split-step technique can be found in Mattingly et al. (2002), where the drift is approximated by an Euler implicit scheme. The dynamics of UPILA2 and UPILA3 can be represented as follows:

$$L_{n+1} = \mu(L_n) + \sigma_h(L_n)\xi_{n+1}, \qquad h = 1,2 \tag{7}$$

where

$$\mu(x) = \left(\frac{1 + \hat{\theta}\frac{1}{2}\, k_d\ |x|^d\, \delta}{1 - \theta\frac{1}{2}\, k_d\ |x|^d\, \delta}\right) x, \tag{8}$$

$$\sigma_1^2(x) = \left[\left(\frac{\sqrt{\delta}}{1 - \theta\,\frac{1}{2}\, k_d\ |x|^d\, \delta}\right)\right]^2, \qquad \sigma_2^2(x) = \delta, \tag{9}$$

and $\sigma_h(x)$, $h = 1,2$ stands for UPILA2/3 respectively.

Remark 1. *Notice that the ULA chain arises as a particular case of UPILA1, UPILA2, and UPILA3 with $\theta = 0$ and $\xi_{n+1} \sim \mathcal{N}(0,1)$. Ergodic results for ULA chains are in Roberts and Tweedie (1996a).*

3 Ergodic Properties of Unadjusted Implicit Schemes

3.1 Notation and basic definitions

Let us consider a generic scalar Markov chain $X := \{X_n : n \in \mathbb{N}\}$ on a state space $(\Re, \mathcal{B}(\Re))$ where $\mathcal{B}(\Re)$ is the Borel $\sigma-$algebra on $\Re$. For any $x \in \Re$, let $\mathbb{P}^n(x,\cdot) : \mathcal{B}(\Re) \to [0,1]$ be its n-step transition kernel:

$$\mathbb{P}^n(x,A) = \Pr\left(X_n \in A \mid X_0 = x\right); \quad A \in \mathcal{B}(\Re),\ n \in \mathbb{N}^+$$

with the convention that $\mathbb{P}(x,\cdot) \equiv \mathbb{P}^1(x,\cdot)$. Given a non-trivial probability measure p on $(\Re, \mathcal{B}(\Re))$ we say that $\mathbb{P}^n$ converges to p or that the chain X is p-ergodic if, for p-a.e. x:

$$\|\mathbb{P}^n(x,\cdot) - p(\cdot)\| \stackrel{n\to+\infty}{\longrightarrow} 0 \tag{10}$$

where for any signed measure M on $(\Re, \mathcal{B}(\Re))$ the symbol $\|M\|$ denotes its total variation norm:

$$\|M\| = \sup_{A \in \mathcal{B}(\Re)} M(A)$$

Analogously, we say that $\mathbb{P}^n$ converges exponentially to p or that the chain X is p-geometrically ergodic if, for p-a.e. x there exists a constant $r < 1$ and a function $m(\cdot)$ such that, for any $n \in \mathbb{N}^+$:

$$\|\mathbb{P}^n(x, \cdot) - p(\cdot)\| \leq m(x)\, r^n$$

3.2 Convergence properties

Our treatment of geometric ergodicity follows the classic approach of Meyn and Tweedie (1993). The first step is to establish communication properties of the chain (i.e. μ^{Leb}-irreducibility and aperiodicity) and minorisation condition for small sets. We set $\mathbb{Q}$ to represent the transition kernel of the unadjusted Langevin discretisation L with discretisation interval δ and by q its Lebesgue transition density.

Lemma 1. *Let us assume that $\pi \in \mathcal{A}_d^{'}$, $d \geq 0$. For any $\theta \in [0,1]$, UPILA1, UPILA2 and UPILA3 are μ^{Leb}-irreducible, aperiodic and all compact sets are small.*

Proof: We denote the density function of ξ_{n+1} by φ. We consider UPILA1 first. L is defined as a solution to (5) or equivalently as a solution to,

$$F(L_{n+1}) = \mu_F(L_n) + \sqrt{\delta}\xi_{n+1}, \tag{11}$$

where $\mu_F(x) = F(x) + \frac{1}{2}\, k_d \operatorname{sign}(x)\; |x|^{d+1}\, \delta$, and F is defined in (6). It is now easy to check that the transition density of UPILA1 is,

$$\text{UPILA1: } q(x,y) \;=\; \varphi\left(\frac{F(y) - \mu_F(x)}{\sqrt{\delta}}\right)\frac{1}{\sqrt{\delta}}F'(y),$$

where

$$F'(y) = 1 - \frac{1}{2}\, k_d\, \theta\; (\operatorname{sign}(y))^{d+1}\; y^{d+1}\delta.$$

We next consider the chains UPILA2 and UPILA3. From (7), the transition densities of UPILA2 and UPILA3 are given by:

$$\text{UPILA2/3}: q(x,y) = \varphi\left(\frac{y - \mu(x)}{\sigma(x)}\right)\frac{1}{\sigma_h(x)}, \qquad h = 1, 2$$

The results now follow from basic results in Chapter 6 of Meyn and Tweedie (1993).

$\square$

Thus, from Theorem 15.0.1 of Meyn and Tweedie (1993), for geometric ergodicity to hold it is sufficient to find a function $V : \Re \to [1, +\infty]$ such that for some $\lambda < 1$,

$$\limsup_{|x| \to +\infty} \frac{PV(x)}{V(x)} \leq \lambda, \tag{12}$$

where $PV(x) = \int q(x, y)V(y)dy$. In Theorem 1 we consider the case $d > 0$ (light tails) and in Theorem 2 we consider the case $d = 0$ (Gaussian case). We note that UPILA1 is the same as UPILA2 for the case $d = 0$ and UPILA3 is defined for $d > 0$ only. Thus, for $d = 0$ we need to consider UPILA2 only. For the rest of the results in this Section we assume that

$$\xi_{n+1} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1) \qquad n \in \mathbb{N} \tag{13}$$

We consider heavier tail distributions for the noise in Section 4.

Theorem 1. *Let us consider target densities $\pi \in \mathcal{A}_d^{'}$, $d > 0$.*

- *UPILA1 is geometrically ergodic if and only if $\theta \geq \frac{1}{2}$.*
- *UPILA2 and UPILA3 are geometrically ergodic if and only if $\theta \geq \frac{1}{2}$ for $d \in (0, 1]$ and if and only if $\theta > \frac{1}{2}$ for $d > 1$.*

Proof: We prove that the required conditions are sufficient. The converse implication (necessary conditions) follows from the results in Lemma 2. We propose the following drift functions: $V(u) := e^{s\,|F(u)|}$, $s > 0$ for UPILA1 (where $F(\cdot)$ is defined in (6)) and $V(u) := e^{s\,|u|}$, $s > 0$ for UPILA2 and UPILA3. We consider the limit (12) when $x \to +\infty$; the negative case follows by symmetry.

Let us consider UPILA1 first. Since the function $F(\cdot)$ is continuous and monotone increasing there exists $x_+ > 0$ such that for any $x > x_+$, $F(x) > 0$. In this context for any δ, $s > 0$, $\theta \in [0, 1]$ and $x > x_+$:

$$\begin{aligned}
\frac{PV(x)}{V(x)} &= e^{-s\,|F(x)|}\,\mathbb{E}\left[e^{s\,|F(L_{n+1})|} \mid L_n = x\right] \\
&\leq e^{-s\,F(x)}\left(\mathbb{E}\left[e^{s\,F(L_{n+1})} \mid L_n = x\right] + \mathbb{E}\left[e^{-s\,F(L_{n+1})} \mid L_n = x\right]\right)
\end{aligned}$$

$$L_1 \overset{def}{=} \lim_{x \to +\infty} \frac{PV(x)}{V(x)} \leq c \lim_{x \to +\infty} \left(e^{s\,\frac{1}{2}\,k_d\,x^{d+1}\delta} + e^{s\left(\frac{1}{2}\,k_d\,x^{d+1}(\theta - \frac{1}{2})\delta - 2\,x\right)}\right),$$

where c is a positive constant. Thus, L_1 is equal to 0 for $\theta \geq \frac{1}{2}$, and $+\infty$ otherwise. An analogous argument leads to the following inequality for UPILA2 and UPILA3:

$$\frac{PV(x)}{V(x)} \leq e^{-s\,x + \frac{s^2}{2}D}\left(e^{-s\,\mu(x)} + e^{+s\,\mu(x)}\right)$$

for some positive constant $D > 0$. It is easy to check that

$$L_2 \stackrel{def}{=} \lim_{x \to +\infty} \frac{PV(x)}{V(x)} = \begin{cases} 0 & \text{if } \theta > \frac{1}{2} \text{ or } \theta = \frac{1}{2} \text{ and } d < 1 \\ \\ < 1 & \text{if } \theta = \frac{1}{2} \text{ and } d = 1 \text{ for small } s \\ \\ \infty & \text{otherwise} \end{cases}$$

The proof follows from the drift condition (12).

□

Theorem 2. *Let us consider target densities $\pi \in \mathcal{A}_d'$, $d = 0$. UPILA2 is geometrically ergodic for $\theta \geq \frac{1}{2}$ or $\theta < \frac{1}{2}$ with $\delta < \frac{2}{k_0(2\theta - 1)}$.*

PROOF: We apply the same framework as in the proof of Theorem 1. Thus choosing the drift function $V(u) = e^{s\,|u|}$, $s > 0$, we obtain that the drift condition holds when $\theta \geq \frac{1}{2}$, or when $\theta < \frac{1}{2}$ and $\delta < \frac{2}{k_0(2\theta-1)}$.

□

The following Lemma provides a characterisation of the behavior of UPILA1, UPILA2 and UPILA3 when they are not geometrically ergodic and thus completes the proofs of Theorem 1 and 2.

Lemma 2. *Let us consider the class of target densities $\pi \in \mathcal{A}_d'$, $d \geq 0$.*

1. *For any $\theta < 1/2$,*
   - *(a) If $d > 0$, UPILA1, UPILA2 and UPILA3 are transient.*
   - *(b) If $d = 0$, UPILA1 (=UPILA2) is transient for $\delta > \frac{2}{k_0(2\theta-1)}$.*
2. *If $\theta = 1/2$ and $d > 1$, UPILA2 and UPILA3 are not geometrically ergodic while UPILA2 is ergodic.*

PROOF: In the Appendix.

The table provides a summary of the convergence results for UPILA1, UPILA2, and UPILA3.

[1] Notice that the results for UPILA1 are identical to UPILA2/3 with the only exception of the case $d > 1$, $\theta = \frac{1}{2}$ where UPILA1 is geometrically ergodic.

Summary of the results for UPILA1, UPILA2/3			
	$\theta < 1/2$	$\theta = 1/2$	$\theta > 1/2$
$d = 0$ (UPILA1)	G.E. ($\delta < \delta^*$)	G.E.	G.E.
$d = 0$ (UPILA1)	T ($\delta > \delta^*$)	G.E.	G.E.
$d \in (0, 1]$ (UPILA1, UPILA2/3)	T	G.E.	G.E.
$d > 1$ (UPILA1)	T	G.E.	G.E.
$d > 1$ (UPILA2)	T	E.	G.E.

Generally speaking, when the implicit component dominates the explicit component ($\theta > \frac{1}{2}$), the Langevin scheme preserves the drift condition and geometrically ergodicity follows. Figure 3 illustrates the superior performance of partial implicit schemes with $\theta > \frac{1}{2}$ with respect to the standard Euler discretisation ($\theta = 0$) in Figure 2. It represents typical trajectories of UPILA1 and UPILA2 with $\theta = 0.7$ and $\delta = 0.1$ for the model $d\,l_t = -2\,l_t^3 + dW_t$ with a starting point 5.

[Figure 3 about here.]

4 Metropolis Adjusted Implicit Langevin Algorithms

4.1 Introduction

A generic recipe for construction of a Markov chain with desired stationary density π is the Metropolis-Hastings construction. Given the current state x, we propose a move to y according to the transition density $q(x, y)$ and accept the proposed move with probability $\alpha(x, y)$:

$$\alpha(x, y) = \begin{cases} 1 \wedge \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} & \text{if } \pi(x)q(x, y) > 0 \\ 1 & \text{if } \pi(x)q(x, y) = 0 \end{cases}$$

The resulting Metropolis-Hastings kernel $\mathbb{P}$ is given by:

[1] Notes on the table: T = transient, G.E. = Geometrically Ergodic, E. = Ergodic (but not geometrically ergodic). Furthermore $\delta^* = \frac{2}{k_0(2\theta-1)}$.

$$\mathbb{P}(x, A) \;=\; \int_A p(x, y)dy \;+\; r(x)\, \mathbb{I}_{\{A\}}(x), \quad A \in \mathcal{B}(\Re)$$

where the function $r(x)$ is the Metropolis Hastings rejection probability from the state x and $p(x, y)$ is the Metropolis-Hastings off-diagonal transition density:

$$r(x) = 1 - \int_{\Re} p(x, y)\, dy; \quad p(x, y) = \begin{cases} q(x, y)\, \alpha(x, y) & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

The Metropolis-Hastings correction guarantees π-stationarity of the Metropolis-Hastings Langevin algorithm through the detailed balance condition: i.e. for any $A \in \mathcal{B}(\Re)$,

$$\pi(x)p(x, y) = \pi(y)p(y, x) \;\Rightarrow\; \int_{\mathbf{R}} \pi(x)\mathbb{P}(x, A)dx = \int_A \pi(y)dy$$

Irreducibility and aperiodicity of the Metropolis adjusted chain follows from irreducibility and aperiodicity of the unadjusted chain under a positivity assumption on the target density π. From Lemma 1, the Metropolis adjusted Lengevin chain is μ^{Leb}-irreducible and aperiodic and therefore $\pi-$ergodic. However, in general the adjusted Metropolis-Hastings algorithm does not necessarily inherit the geometric ergodicity of the unadjusted algorithm. In fact if

$$\lim_{|x| \to +\infty} r(x) \;=\; \lim_{|x| \to +\infty} \mathbb{P}\left(x, \, \{x\}\right) \;=\; 1 \tag{14}$$

then, from Theorem 5.1 of Roberts and Tweedie (1996b), the Metropolis-Hastings algorithm is not geometrically ergodic. Good examples of this phenomenon can be found in Stramer and Tweedie (1999) and Hansen (2003). On the other hand, if

$$\limsup_{|x| \to +\infty} r(x) \;=\; \limsup_{|x| \to +\infty} \mathbb{P}\left(x, \, \{x\}\right) \;=\; 0 \tag{15}$$

then, from Theorem 3.1 of Stramer and Tweedie (1999), the Metropolis-Hastings algorithm is geometrically ergodic.

The later result suggests a simple way to construct a geometrically ergodic Metropolis-Hastings chain by selecting a geometrically ergodic unadjusted Langevin scheme and ensuring that the Metropolis-Hastings rejection probability tends to 0 as we move further in the tails. Unfortunately, condition (15) is often not satisfied.

In Section 4.2 we employ UPILA1, UPILA2, and UPILA3 schemes as proposals for the Metropolis-Hastings algorithms when $\pi \in \{\mathcal{A}'_d, \;\; d \geq 0\}$. We call these algorithms MAPILA1, MAPILA2, and MAPILA3 respectively (Metropolis Adjusted Implicit and Partial Implicit Langevin Algorithm). We choose the distribution of the noise term ξ_{n+1} in such a way that the unadjusted scheme is geometrically ergodic and that condition (15) is satisfied.

4.2 Convergence properties

We firstly consider the light tails case ($d > 0$).

Theorem 3. *Let us assume that $\pi \in \mathcal{A}'_d$, $d > 0$. Under assumption (13),*

(a) MAPILA3 is geometrically ergodic for all $\theta > \frac{1}{2}$.

(b) MAPILA1 and MAPILA2 are not geometrically ergodic for all $0 \leq \theta \leq 1$.

PROOF: The proof follows by showing that condition (15) holds for MAPILA3 while condition (14) holds for MAPILA1 and MAPILA2. Part (a) follows from Theorem 5.1 of Roberts and Tweedie (1996b) and Part (b) follows from Theorem 3.1 of Stramer and Tweedie (1999). We omit the details as they require simple tedious algebra calculations.

□

Using UPILA3 with Gaussian tails (13) as the proposal chain guarantees geometric ergodicity of its Metropolis adjusted scheme, MAPILA2. Yet, it looses the nice property of UPILA2, that the variance of the transition density is state dependent. One way to obtain geometric ergodicity of MAPILA1 and MAPILA2 is to choose, as in Stramer and Tweedie (1999),

$$\frac{\xi_{n+1}}{\sqrt{\nu/(\nu-2)}} \overset{\text{i.i.d}}{\sim} t(\nu) \tag{16}$$

where $t(\nu)$ denotes the law of the t distribution with $\nu > 2$ degrees of freedom so that the variance is finite. This proposal has thicker tails than the normal distribution to help prevent the sampler from getting stuck in the tails. Our simulation results were robust to the choice of ν. We show that this variation guarantees the geometric ergodicity of UPILA1 and UPILA2.

Theorem 4. *Let us assume that $\pi \in \mathcal{A}'_d$, $d > 0$ and consider MAPILA1 and MAPILA2 schemes under the assumption (16). MAPILA1 and MAPILA2 are geometrically ergodic for all $\theta > \frac{1}{2}$.*

PROOF: We firstly note that for $\theta > \frac{1}{2}$, UPILA1 and UPILA2 are geometric ergodic under the heavier tails assumption (16). This follows as in Section 3.2 by choosing the drift functions: $V(u) = |F(u)| + 1$ for UPILA1 (where $F(\cdot)$ is defined in (6)) and the drift functions: $V(u) = u^2 + 1$ for UPILA2. The proof that (15) holds for both schemes requires simple algebra and is omitted.

$\square$

We next consider Gaussian tails ($d = 0$). We show that there is no need to adjust the distribution of the noise to obtain geometric ergodicity. Recall that for $d = 0$ we have one algorithm, MAPILA2. Again, the proof is simple and is omitted.

Theorem 5. *Let us assume that $\pi \in \mathcal{A}'_d$, $d = 0$. MAPILA2 is geometrically ergodic for $\theta = \frac{1}{2}$ and for all $\theta < \frac{1}{2}$ with $\delta < \frac{2}{k_0(2\theta-1)}$. In particular for $\theta = \frac{1}{2}$ the Metropolis-Hastings rejection probability $r(x)$ is equal to 0 for any $x \in \Re$.*

Remark 2. *The behavior of MAPILA2 for $d = 0$ and $\theta = \frac{1}{2}$ is somewhat surprising as the Metropolis step is always accepted. This implies that UPILA2 and MAPILA2 coincide or, equivalently, that the Metropolis-Hastings correction is not really needed since UPILA2 is able to reproduce exactly the ergodic behavior of the Langevin diffusion.*

We illustrate the ergodic behavior of MAPILA1, MAPILA2, MAPILA3, and RWM schemes with starting point 200 for $\pi(x) \propto \exp\left(-x^4\right)$. We simulate MAPILA1, MAPILA2 (each with noise assumption (16) with $\nu = 30$) and MAPILA3 (with a Gaussian noise (13)) with $\delta = 0.1$, and implicit parameter $\theta = 0.7$. We also simulate RWM with $\mathcal{N}(0, 0.1)$ noise. The trace plots of the steps taken by the four algorithms appear in Figure 4. It is clear that all three Metropolis adjusted partial implicit Langevin algorithms with $\theta = 0.7$ hit neighborhood of 0 more rapidly than the RWM scheme.

We next assess the behavior of a single long series for the later four algorithms. We have used all algorithms with $100,000$ steps; we start from 0 and discard $10,000$ steps to eliminate the effect of the initial point. The estimated histograms for π (red lines) and the true density function π (black lines) appear in Figure 5. All three Metropolis adjusted partial implicit Langevin algorithms with $\theta = 0.7$ perform slightly better than the RWM algorithm.

[Figure 4 about here.]

[Figure 5 about here.]

5 Extensions of Partial Implicit schemes

5.1 More general class of partial implicit schemes

The investigation of ergodic properties of partial implicit schemes within the class $\mathcal{A}'_d,\ d \geq 0$ shows how to preserve the drift condition by reinforcing the implicit component in the discretization of the Langevin diffusion. The same effect applies to all those target densities whose tail behavior is analogous to the tail behavior of $\pi \in \mathcal{A}'_d, \quad d \geq 0$. Thus we consider a more general class of target densities of the form $\{\mathcal{A}_d,\ d \geq 0\}$ where, for any $d \geq 0$, $\mathcal{A}_d$ includes all those target densities π characterized by the following limits:

$$\lim_{u \to +\infty} \frac{\nabla \log \pi(u)}{u\ u^d} \quad = k_d^+ \in (-\infty, 0)$$
$$\lim_{u \to -\infty} \frac{\nabla \log \pi(u)}{u\ \ |u|^d} \quad = k_d^- \in (-\infty, 0)$$

Note that $\mathcal{A}'_d \subset \mathcal{A}_d$. We can write the dynamics of UPILA2 for $\mathcal{A}_d$ in the following way:

$$\text{UPILA2:} \qquad L_{n+1} = \left(\frac{1 + \hat{\theta} A\left(L_n\right) \delta}{1 - \theta A\left(L_n\right) \delta} \right) L_n + \left(\frac{\sqrt{\delta}}{1 - \theta A\left(L_n\right) \delta} \right) \xi_{n+1},$$

where $A(u) = \frac{1}{2} \frac{\nabla \log \pi(u)}{u},\ u \neq 0$. Clearly, UPILA2 scheme for $\mathcal{A}_d$ satisfies the same drift condition as for $\mathcal{A}'_d$ and thus, under regularity conditions is able to recover geometric ergodicity behavior from the Langevin diffusion.

We illustrate the ergodic behavior of MAPILA2 ($\delta = 0.1,\ \theta = 0.7$) and RWM schemes for $\pi(x) \ \propto\ \exp\left(-x^{4} + x^2\right)$. We note that communication properties for MAPILA2 hold. To obtain geometric ergodicity we use MAPILA2 with a tail noise (16) with $\nu = 30$. Figure 6 displays a trace plot of the algorithm with starting point 5. Clearly MAPILA2 hits a neighborhood of 0 quite rapidly. We again assess the behavior of a single long series for MAPILA2 and RWM with $100,000$ steps; we start from 0 and discard $10,000$ steps to eliminate the effect of the initial point. Figure 7 depicts the estimated histograms. Clearly, MAPILA2 scheme describes each mode better than the RWM scheme.

[Figure 6 about here.]

[Figure 7 about here.]

5.2 Using partial implicit local linearization

Based on Shoji and Ozaki (1998) we now define the partial implicit local linear approximation for a Langevin diffusion (1). We firstly review the explicit scheme. Over the time interval

$[n\delta, (n+1)\delta]$, we use a first order Taylor expansion for the drift $\frac{1}{2}\nabla \log \pi(l_t)$. The Langevin process l_t is then approximated by the linear process $\tilde{l}_t$, defined as a solution to

$$d\tilde{l}_t = \Big(\frac{1}{2}\nabla \log \pi(x) + \frac{1}{2}\nabla^2 \log \pi(x)(\tilde{l}_t - x)\Big)dt + dW_t, \quad n\delta \le t \le (n+1)\delta \tag{17}$$

where $\tilde{l}_{n\delta} = x$ and ∇^2 is the second-order partial derivative. The explicit local linearization scheme L_{n+1} given L_n is defined as a solution to the linear stochastic differential equation (17) at time $(n+1)\delta$.

Ergodic properties of the explicit local linearization scheme are derived in Stramer and Tweedie (1999) for the one-dimensional case and in Hansen (2003) for the multi-dimensional case. It is shown that for a large class of light tail distributions, the local linearization scheme is geometrically ergodic. Yet, it does not inherit geometric ergodicity in complete generality and furthermore, it does not in general lead to geometric ergodicity of the Metropolis Hastings algorithm.

We propose the following partial implicit scheme based on the explicit local linearization scheme (17):

$$\begin{aligned} L_{n+1} &= L_n + \Big(\frac{1}{2}\nabla \log \pi(L_n) - \frac{1}{2}\nabla^2 \log \pi(L_n) L_n\Big)\delta \\ &+ \frac{1}{2}\nabla^2 \log \pi(L_n)(\theta L_{n+1} + \hat{\theta} L_n)\delta + \sqrt{\delta}\,\xi_{n+1} \end{aligned} \tag{18}$$

The dynamics of (18) can be represented as follows:

$$L_{n+1} = \mu(L_n) + \Sigma^{\frac{1}{2}}(L_n)\xi_{n+1},$$

where

$$\begin{aligned} \mu(x) &= x + \Big(I - \frac{1}{2}\nabla^2 \log \pi(x)\theta\delta\Big)^{-1}\Big(\frac{1}{2}\nabla \log \pi(x)\delta\Big) \\ \Sigma(x) &= \delta\Big(I - \frac{1}{2}\nabla^2 \log \pi(x)\theta\delta\Big)^{-2} \end{aligned}$$

The ergodic behavior of the partial implicit local linear approximations are only illustrated here, but they do indicate that the partial implicit approach can be expected to work well in higher-dimension.

We firstly illustrate this algorithm with the one-dimensional model, $d\,l_t = -2\,l_t^3\,dt + dW_t$ corresponding to the target density $\pi(x) \propto \exp\big(-x^4\big)$. For this example,

$$\mu(x) = x - \frac{2x^3\delta}{1+6x^2\theta\delta} \qquad \Sigma(x) = \frac{\delta}{(1+6x^2\theta\delta)^2}$$

It is easy to check that the partial implicit local linearization scheme is π-geometrically ergodic for all $\theta > \frac{1}{3}$. We now assess the behavior of a single long series for the partial implicit local linearization algorithm with $\theta = 0.4$, $\delta = 0.1$ and compare it to UPILA2 with $\theta = 0.7$, $\delta = 0.1$. We have used the two algorithms with $100,000$ steps; we start from 0 and discard $10,000$ steps to eliminate the effect of the initial point. Figure 8 demonstrates the estimated histograms for π using the partial implicit local linearization scheme (red lines), UPILA2 scheme (green lines) and the true density function π (black lines).

[Figure 8 about here.]

We next illustrate the advantage of using the partial implicit local linearization scheme over the explicit local linearization and the Euler schemes with the bivariate Langevin diffusion model (1) corresponding to the target density $\pi(x) \propto \exp(-2(x_1^4 + x_2^4 - x_1^2 x_2^2))$, $x = (x_1, x_2) \in \Re^2$. For this example,

$$\frac{1}{2}\nabla \log(\pi(x)) = \begin{bmatrix} -4x_1^3 + 2x_1x_2^2 \\ -4x_2^3 + 2x_2x_1^2 \end{bmatrix}, \quad \frac{1}{2}\nabla^2 \log(\pi(x)) = \begin{bmatrix} -12x_1^2 + 2x_2^2 & 4x_1x_2 \\ 4x_1x_2 & -12x_2^2 + 2x_1^2 \end{bmatrix}.$$

From Hansen (2003), the Langevin diffusion process is geometric ergodic while the explicit local linearization chain is transient. It is also easy to check that the explicit Euler scheme is transient. It can be shown that the partial implicit local linearization scheme L for this example is exponentially ergodic for implicit parameter θ that is bigger than a certain threshold. We omit the details and illustrate our results in the following two Figures. Figure 9 is a trace plot of the steps taken by the partial implicit local linearization scheme L with $\theta = 0.5$, $\delta = 0.1$ and different starting points. The arrows indicate the end of each step, showing rapid convergence. Figure 10 is a trace plot of the steps taken by L for $5,000$ iterations with a starting point $(0,0)'$. It is clear from Figures 9 and 10 that the process L hits a neighborhood of $(0,0)'$ (the mode of π) quite rapidly, and then proceeds to move around this mode. Figure 10 provides an approximation for the shape of π.

[Figure 9 about here.]

[Figure 10 about here.]

6 Conclusion

We have defined different partial implicit Langevin algorithms with implicitness parameter $0 \le \theta \le 1$. We have shown that by introducing implicitness in the discretization, our proposed chains have better ergodic properties than the explicit Euler scheme (where $\theta = 0$). Furthermore, ergodic properties of a partial implicit scheme can be preserved when used as a proposal for the Metropolis-Hastings algorithm by choosing a heavier tail distribution than the Gaussian distribution for the noise.

All of our results are described in detail and illustrated in one dimension. We also outline possible extension of our results to the multi-dimensional case that is based on the local linearization scheme. The study of ergodic properties of the multi-dimensional case is left as a future research.

Appendix: proof of Lemma 2

Preliminaries

Let us introduce some preliminary notation. Given a generic Markov chain $X := \{X_n : n \in \mathbb{N}\}$ let us define:

$$\tau_X(A) := \inf_{t \in \mathbb{N}^+} \{n : X_n \in A\}; \qquad A \in \mathcal{B}(\Re)$$

and, correspondingly:

$$\mathbb{F}^X_{x,A}(n) := Pr\left(\tau_X(A) \leq n \mid X_0 = x\right)$$

the conditional distribution function of $\tau_X(A)$. Consider now another Markov chain $X' = \{X'_n : n \in \mathbb{N}\}$ on the same probability space. For a given set $A \in \mathcal{B}(\Re)$ and value $x \in \Re$, we say that $\mathbb{F}^X_{x,A}$ *dominates stochastically* $\mathbb{F}^{X'}_{x,A}$ and write $\mathbb{F}^X_{x,A} \overset{s.d.}{\rightarrow} \mathbb{F}^{X'}_{x,A}$ if, for any $n \in \mathbb{N}^+$, $\mathbb{F}^X_{x,A}(n) \leq \mathbb{F}^{X'}_{x,A}(n)$. We state the following useful Lemma.

Lemma 3. *Let us suppose that the chain X' is transient (not geometrically ergodic). If there exists a set C such that C is small for both X and X' and for any $x \in C^c$, $\mathbb{F}^X_{x,C} \overset{s.d.}{\rightarrow} \mathbb{F}^{X'}_{x,C}$ then the chain X is also transient (not geometrically ergodic).*

PROOF: Transience (not geometric ergodicity) of X follows from Theorem 8.3.6 (Theorem 15.0.1) of Meyn and Tweedie (1993) by a stochastic comparison argument (see for example Meyn and Tweedie, 1993, page. 221, 222).

□

Proof of statement 1(a)

Let us consider the following class of SETAR (*self-exciting threshold autoregressive*) models $S := \{S_n : n \in \mathbb{N}\}$:

$$S_{n+1} = \gamma_j + \phi_j\, S_n + \sqrt{\delta}\xi_{n+1}; \qquad S_n \in Q_j,\ \ j = 1, 2, \ldots, M \tag{19}$$

where $Q_j = (q_j, q_{j+1}]$, $-\infty = q_1 < q_2 < \ldots < q_M = +\infty$. Conditions on the parameters under which this model is transient are well known (Meyn and Tweedie, 1993, page. 222). In particular S is transient if one of the following two conditions hold:

$$\phi_1 \;<\; 0\,; \quad \phi_1\, \phi_M \;=\; 1\,; \quad \gamma_M \;+\; \phi_M\, \gamma_1 < 0 \tag{20}$$

$$\phi_1 \;<\; 0\,; \quad \phi_1\, \phi_M > 1 \tag{21}$$

To prove transience of UPILA1 we shall focus initially on the chain $F = \{F(L_n) :\ n \in \mathbb{N}\}$ with dynamics (11). For $d > 0$, $\theta < 1/2$ and fixed $\epsilon > 0$, we can find sufficiently large x_0 such that for any x with $|x| > x_0$,

$$\begin{aligned} \mathbb{E}\left(F(L_{n+1}) \mid L_n = x\right) \;&<\; -F(x) - \epsilon \;\;\text{for}\;\;\; x \,>\, x_0 \\ \mathbb{E}\left(F(L_{n+1}) \mid L_n = x\right) \;&>\; -F(x) + \epsilon \;\;\text{for}\;\;\; x \,<\, -x_0 \end{aligned}$$

We compare F with the transient SETAR model S (19) with parameters $M = 2$, $\gamma_1 = -\epsilon$, $\phi_1 = -1$ and $\gamma_2 = 0$, $\phi_2 = -1$ corresponding to the sets $Q_1 = (-\infty, -x_0]$ and $Q_2 = (x_0, +\infty]$ and satisfying (20). By choosing $C = [-x_0, +x_0]$, it is clear, that $\mathbb{F}^F_{x,C} \overset{s.d.}{\to} \mathbb{F}^S_{x,C}$ for any $x \in C^c$. Transience of F follows from Lemma 3 which trivially implies transience of UPILA1.

To prove transience of UPILA2 it suffices to notice that, for $d > 0$ and $\theta < 1/2$, UPILA2 tends to behave as a deterministic explosive system, when in the tails. In fact when $|x|$ is large its conditional mean $\mu(x)$, defined in (8) is approximately $-\frac{\hat{\theta}}{\theta}x < -x$ and its conditional variance $\sigma_1^2(x)$ defined in (9) converges to 0 when $|x| \to \infty$.

To prove transience of UPILA3 we use analogous stochastic comparison arguments as for UPILA2. We observe that for $d > 0$,

$$\mathbb{E}\left(L_{n+1} \mid L_n = x\right) = \mu(x) \overset{def}{=} \beta(x)\, x \;\approx -\frac{\hat{\theta}}{\theta}x,$$

for $|x|$ sufficiently large, where $\mu(x)$ is defined in (8) and $\beta(x) = \frac{1+\hat{\theta}\frac{1}{2}\, k_d\, |x|^d \delta}{1-\theta\frac{1}{2}\, k_d\, |x|^d \delta}$. For $\theta < \frac{1}{2}$, we choose $\epsilon > 0$ such that $-\frac{\hat{\theta}}{\theta} + \epsilon < -1$. There exists x_0 s.t for any x with $|x| \geq x_0$, $\beta(x) < -1 - \epsilon$. We consider the SETAR model defined by $Q_1 := (-\infty, -x_0]$, $Q_2 := (x_0, +\infty]$, $\phi_1 = \beta(-x_0)$

and $\phi_2 = \beta(x_0)$. By construction $\phi_1, \phi_2 < -1$ so that the SETAR model is transient from (21). Furthermore, setting $C = [-x_0, +x_0]$, we have $\mathbb{F}^L_{x,C} \overset{s.d.}{\to} \mathbb{F}^S_{x,C}$ for any $x \in C^c$. Finally transience of UPILA3 follows from Lemma 3.

Proof of statement 1(b)

For $d = 0$, UPILA1(=UPILA2) schemes behave as $AR(1)$ models. It is therefore transient when $|\mu(L_n)| > 1$, where $\mu(\cdot)$ is defined as in (8). This is true for $\theta < \frac{1}{2}$ and $\delta > \frac{2}{k_0\,(2\theta-1)}$.

Proof of statement 2

To prove that UPILA2 and UPILA3 are not geometrically ergodic we adapt an argument from Roberts and Tweedie (1996a) (Theorem 3.2 (a)). For simplicity we outline the argument just for UPILA3; for UPILA2 the conclusion follows similarly. We observe that for any $\epsilon > 0$, we can find a sufficiently large x_0 such that for any $x \geq x_0$:

$$\begin{aligned} \mathbb{E}\left(L_{n+1} \mid L_n = x\right) &> -x + \epsilon \qquad \text{for} \qquad x \geq x_0 \\ \mathbb{E}\left(L_{n+1} \mid L_n = x\right) &> -x - \epsilon \qquad \text{for} \qquad x \leq -x_0 \end{aligned}$$

Consider the random walk with negative drift $R := \{R_n : \ n \in \mathbb{N}\}$ defined by:

$$R_{n+1} = R_n - \epsilon + \sqrt{\delta}\xi_{n+1}.$$

Clearly, if $C = [-x_0, +x_0]$, then $\mathbb{F}^L_{x,C} \overset{s.d.}{\to} \mathbb{F}^R_{x,C}$ for any $x \in C^c$. Thus, according to Lemma 3, in order to complete the argument it suffices to prove that the chain R is not geometrically ergodic. The random walk R can be seen as the δ-embedded chain of the Brownian motion with constant drift $-\epsilon/\delta$. Thus the time taken by the random walk to hit any compact set C is at least as long as the time taken by the continuous process. We can now use Bachelier-Levy formula (as in Roberts and Tweedie, 1996a, Theorem 2.4) to show that these hitting times do not have exponential tails. Therefore, the chain R is not geometrically ergodic and the conclusion follows from Lemma 3.

In order to prove ergodicity (10) of UPILA2 it suffices (Meyn and Tweedie, 1993, Theorem 13.0.1) to find a positive function V, such that, there exist $b < +\infty$ and a small set C satisfying the following condition:

$$\Delta V\,(x) \overset{def}{=} PV(x) - V(x) < \, -1 \, + \, b\,\mathbf{I}_{\{C\}}(x) \tag{22}$$

where $PV(x)$ is defined in (12). Condition (22) ensures the return to regenerative sets and thus the convergence of the Markov chain but, unlike (12), it does not guarantee geometric rate of convergence. We take $V(x) = e^{s|x|}$ as the drift criterion so that $\lim_{x\to+\infty} \Delta V(x) = -\infty$ as $|x| \to \infty$ for $d > 1$. Therefore, UPILA2 is ergodic by (22).

$\square$

References

J. E. Besag (1994). "Comments on "representation of knowledge in complex systems" by U. Grenander and M.I. Miller." *J. Roy. Statist. Soc. Ser. B* **56**, 591–592.

A. Beskos, O. Papaspiliopoulos and G. O. Roberts (2008a). "A new factorisation of diffusion measure and sample path reconstruction." *Methodol. Comput. Appl. Probab.* **10**, 85–104.

A. Beskos, G. O. Roberts, A. Stuart and J. Voss (2008b). "MCMC methods for diffusion bridges." *Stochastic and Dynamics* **8**, 319–350.

O. F. Christensen, G. O. Roberts and M. Sköld (2006). "Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models." *J. Comput. Graph. Statist.* **15**, 1–17.

J. D. Doll, P. J. Rossky and K. L. Friedman (1978). "Brownian dynamics as smart Monte Carlo simulation." *J. Chemical Physics* **69**, 4628–4633.

N. R. Hansen (2003). "Geometric ergodicity of discrete-time approximations to multivariate diffusions." *Bernoulli* **9**, 725–743.

D. J. Higham (2000). "Mean-square and asymptotic stability of the stochastic theta method." *SIAM J. Numer. Anal.* **38**, 753–769 (electronic).

P. Kloeden and E. Platen (1992). *Numerical solutions of stochastic differential equations*. Berlin: Springer-Verlag.

J. C. Mattingly, A. M. Stuart and D. J. Higham (2002). "Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise." *Stochastic Process. Appl.* **101**, 185–232.

K. L. Mengersen and R. L. Tweedie (1996). "Rates of convergence of the Hastings and Metropolis algorithms." *Ann. Statist.* **24**, 101–121.

S. P. Meyn and R. L. Tweedie (1993). *Markov chains and stochastic stability*. Communications and Control Engineering Series. London: Springer-Verlag London Ltd.

G. O. Roberts and R. L. Tweedie (1996a). "Exponential convergence of Langevin distributions and their discrete approximations." *Bernoulli* **2**, 341–363.

G. O. Roberts and R. L. Tweedie (1996b). "Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms." *Biometrika* **83**, 95–110.

Y. Saito and T. Mitsui (1996). "Stability analysis of numerical schemes for stochastic differential equations." *SIAM J. Numer. Anal.* **33**, 2254–2267.

I. Shoji and T. Ozaki (1998). "Estimation for nonlinear stochastic differential equations by a local linearization method." *Stochastic Anal. Appl.* **16**, 733–752.

O. Stramer and R. L. Tweedie (1999). "Langevin-type models. II. Self-targeting candidates for MCMC algorithms." *Methodol. Comput. Appl. Probab.* **1**, 307–328.

A. M. Stuart and A. R. Humphries (1996). *Dynamical systems and numerical analysis*, volume 2 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge: Cambridge University Press.

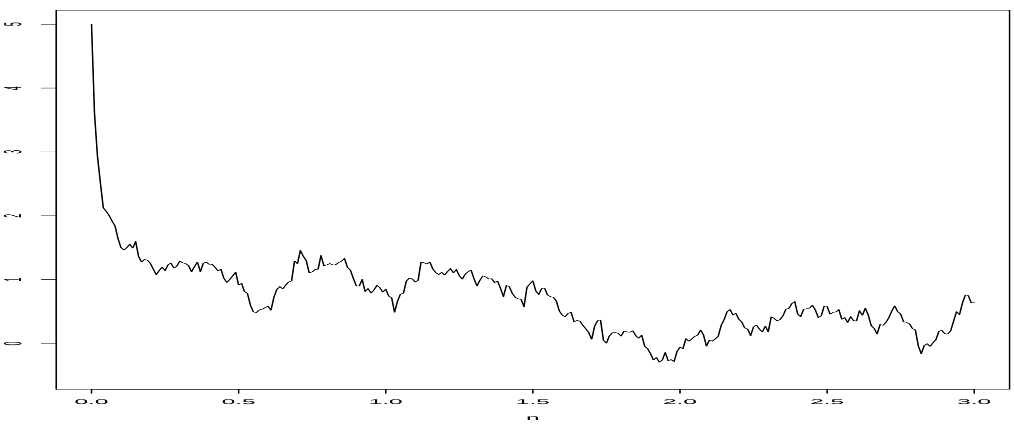

Figure 1: Exact trajectory of the Langevin diffusion: $d\,l_t = -2\,l_t^3\,dt\,+\,dW_t$ with $l_0 = 5$

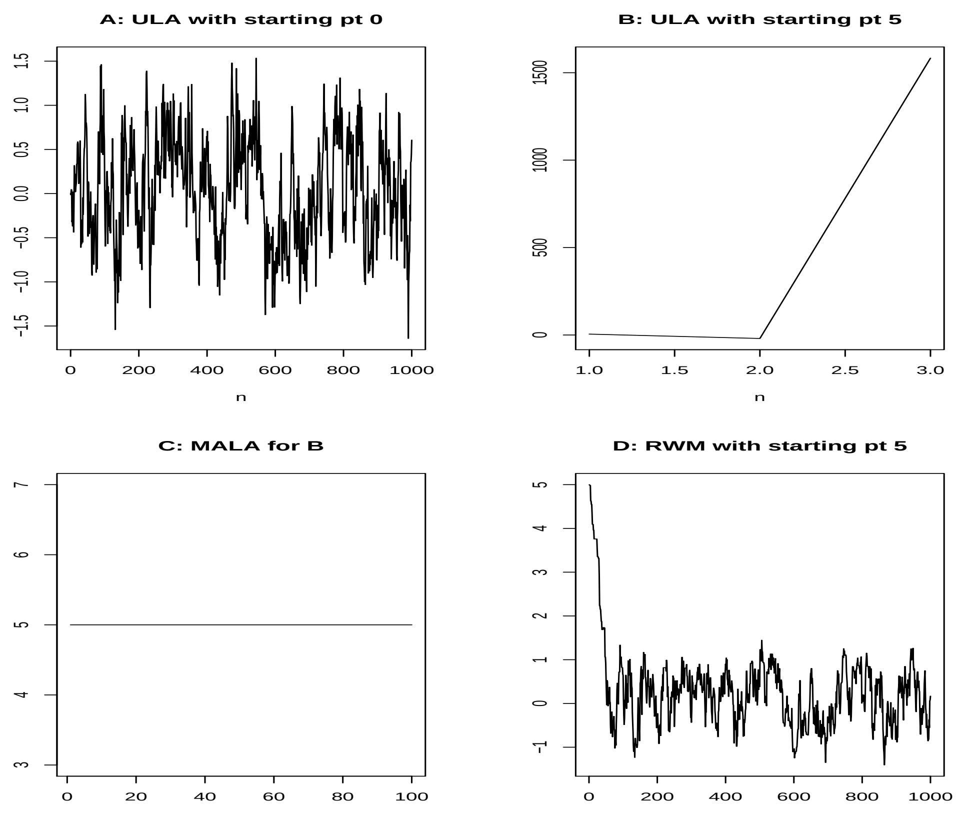

Figure 2: ULA schemes with discretisation interval $\delta = 0.1$ with starting point 0 (Plot A) and starting point 5 (Plot B), MALA scheme for Plot B (Plot C), and RWM with variance 0.1 and starting point 5 (plot D).

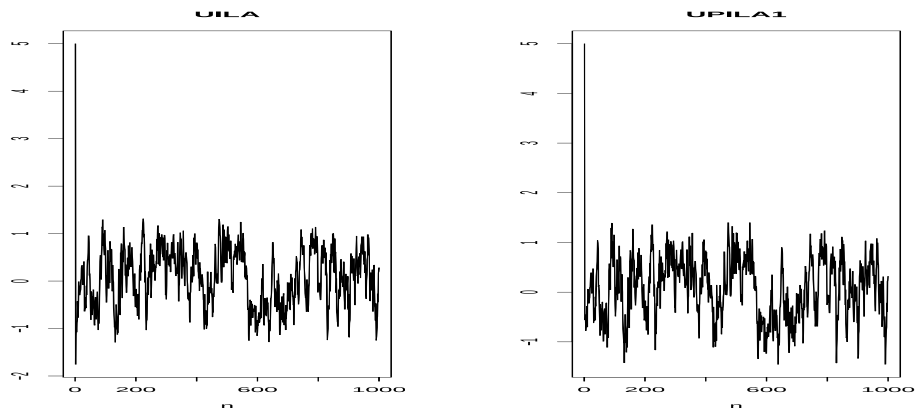

Figure 3: UPILA1 and UPILA2 scheme with discretisation interval $\delta = 0.1$, implicit parameter $\theta = 0.7$ and starting point 5.

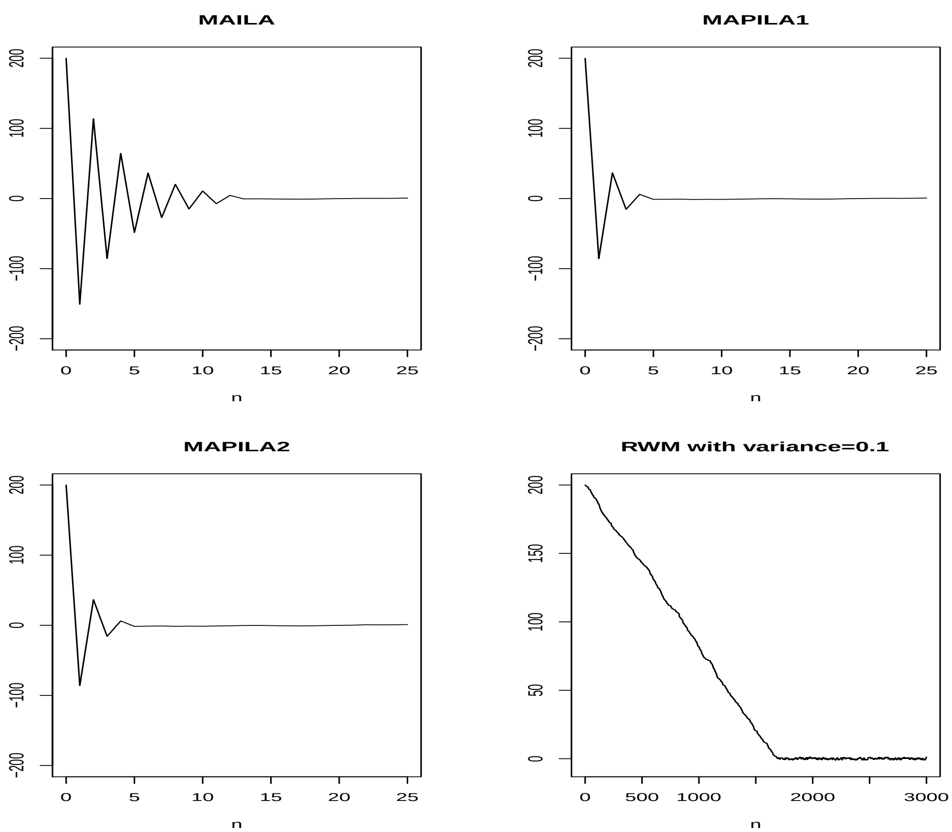

Figure 4: Trace plots for MAPILA1, MAPILA2, and MAPILA3 with $\delta = 0.1$ and $\theta = 0.7$, and RWM scheme with variance 0.1. The starting point for all schemes is 200.

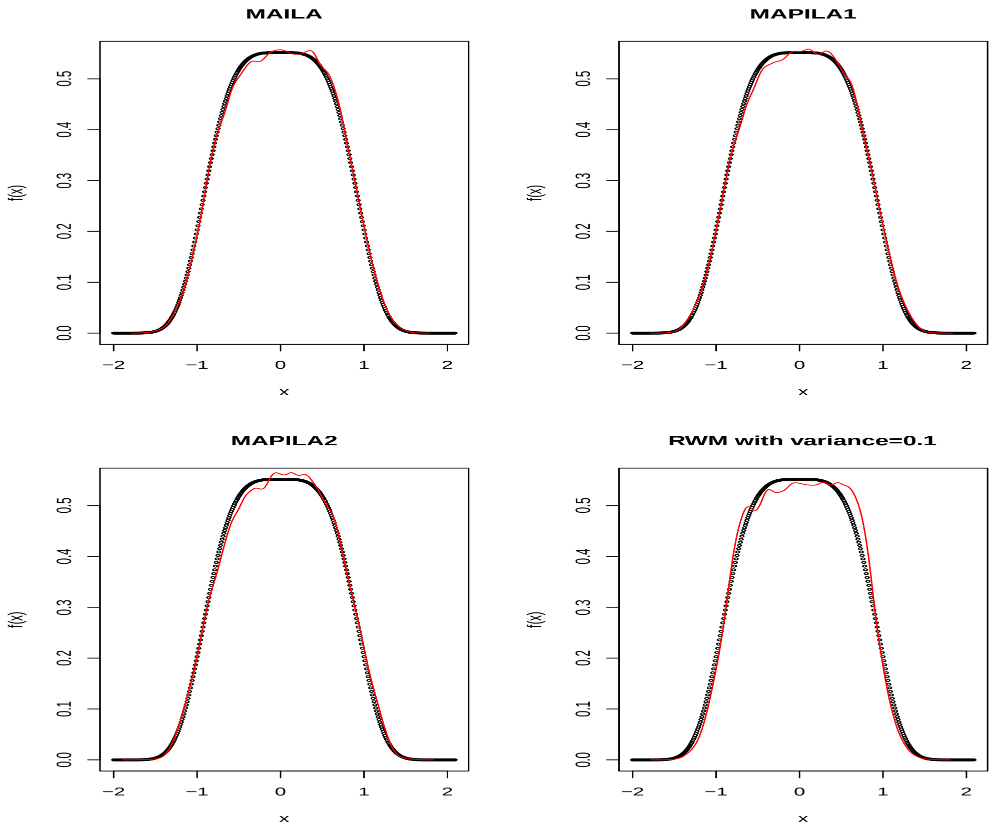

Figure 5: Estimated histograms for $\pi \propto \exp(-x^4)$ (red lines) and the true density function π (black lines). The estimated histograms are for MAPILA1, MAPILA2, MAPILA3 with $\delta = 0.1$ and $\theta = 0.7$, and RWM with variance 0.1. The estimated histograms are based on $N = 100,000$ iterations after a burn-in period of $10,000$ iterations.

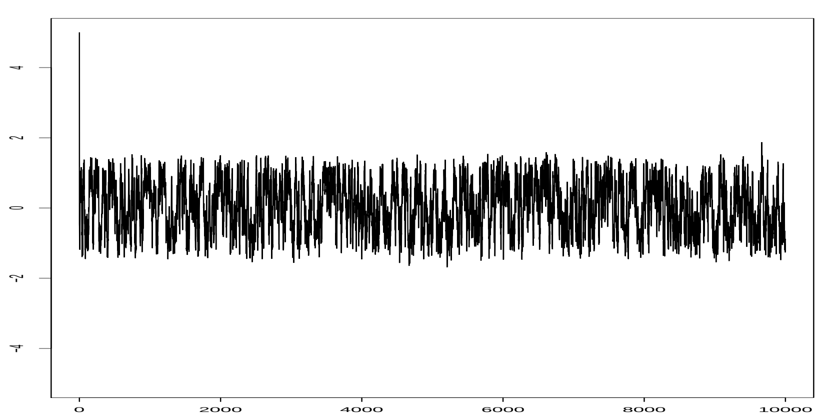

Figure 6: Trace plot for MAPILA2 with $\pi(x) \propto \exp\left(-x^4 + x^2\right)$, $\delta = 0.1$, $\theta = 0.7$ and starting point 5.

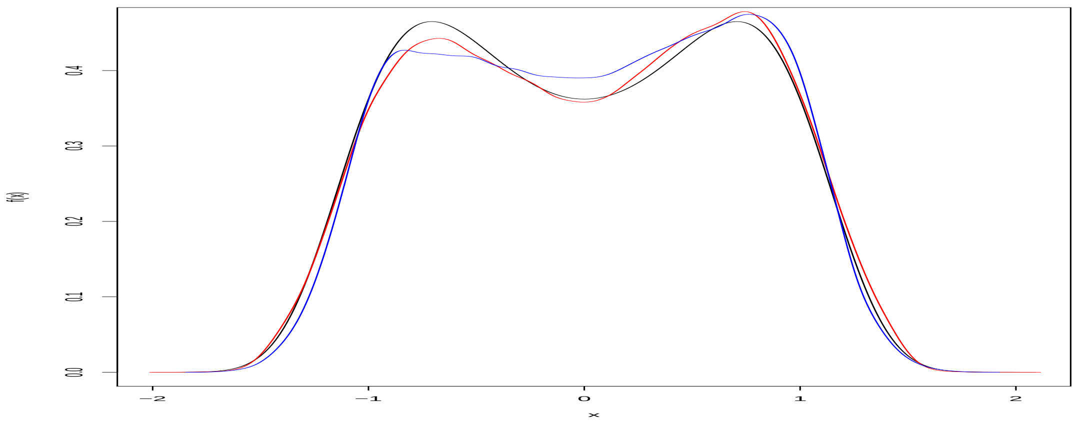

Figure 7: Histograms for MAPILA2 (red) with $\delta = 0.1$ and $\theta = 0.7$, RWM (blue) with variance 0.1, and the true density function $\pi(x) \propto \exp\left(-x^4 + x^2\right)$ (black lines). The histograms are based on $100,000$ iterations after a burn-in period of $10,000$ iterations.

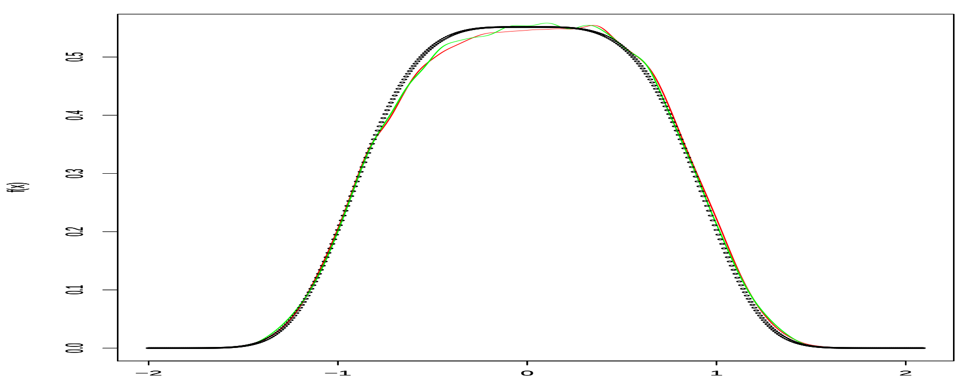

Figure 8: Histograms for the partial implicit local linearization with $\delta = 0.1$ and $\theta = 0.4$ (red line), MAPILA2 with $\delta = 0.1$ and $\theta = 0.7$ (green line), and the true density function π (black lines). The histograms are based on $100,000$ iterations after a burn-in period of $10,000$ iterations.

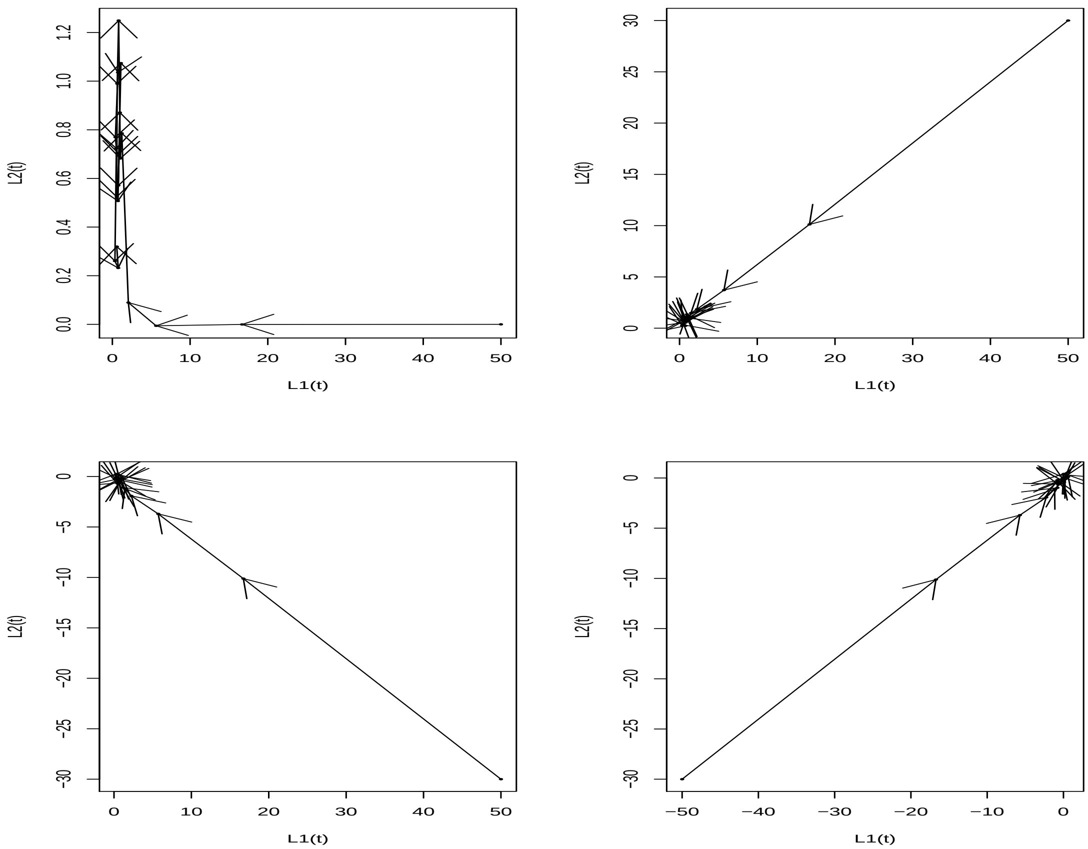

Figure 9: Partial implicit local linearization schemes with $\delta = 0.1$, $\theta = 0.5$, and different starting points.

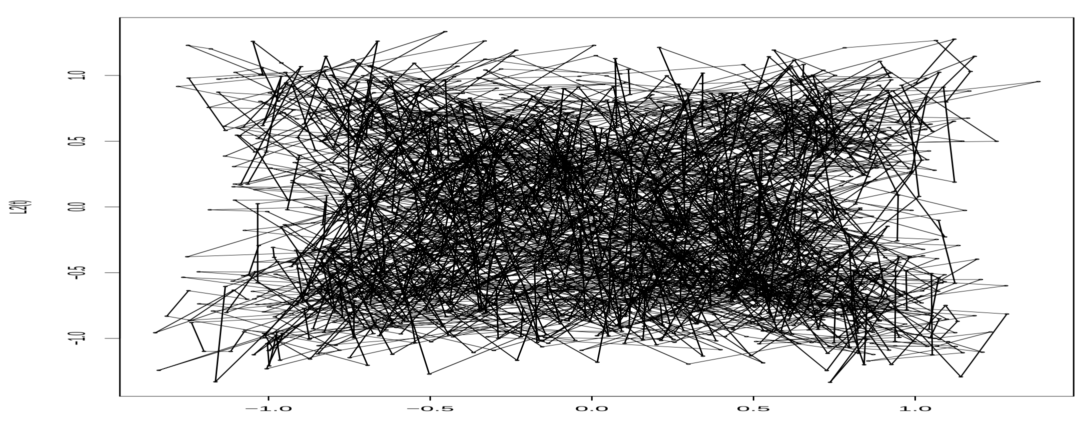

Figure 10: Partial implicit local linearization schemes with $\delta = 0.1$, $\theta = 0.5$ and a starting point $(0, 0)$ based on $5,000$ iterations .