Optimal scaling of the transient phase of Metropolis Hastings algorithms

Tony Lelièvre
Ecole des Ponts and INRIA

Joint work with B. Jourdain and B. Miasojedow

MCMSki, Chamonix, 8 January 2014
Outline of the talk

Introduction

Optimal scaling of the transient phase of RWMH

Longtime convergence of the nonlinear SDE

Optimization strategies for the RWMH algorithm
Introduction

Optimal scaling of the transient phase of RWMH

Longtime convergence of the nonlinear SDE

Optimization strategies for the RWMH algorithm
Metropolis Hastings algorithm

The aim of the MH algorithm is to sample a target probability measure, say with density $p$ on $\mathbb{R}^n$.

Algorithm: iterate on $k \geq 0$,

- **Proposition**: At time $k$, given $X^n_k$, propose a move to $\hat{X}^n_{k+1} \sim q(X^n_k, y) \, dy$, where $q(x, y)$ Markov density kernel on $\mathbb{R}^n$,

- **Accept/Rejection**: Accept the move ($X^n_{k+1} = \hat{X}^n_{k+1}$) with probability $\alpha(X^n_k, \hat{X}^n_k)$, where

\[
\alpha(x, y) := \frac{p(y)q(y, x)}{p(x)q(x, y)} \wedge 1.
\]

Otherwise, reject the move ($X^n_{k+1} = X^n_k$).

$(X^n_k)_{k \geq 0}$ is a reversible Markov chain wrt $p(x) \, dx$.

The efficiency of the algorithm crucially depends on the choice of the proposal distribution $q$. 
In the following, we focus on the Gaussian random walk proposal (RWM):

- \( \hat{X}^{n}_{k+1} = X^n_k + \sigma G^k_{k+1} \) where \((G^k_k)_{k \geq 1}\) i.i.d. \( \sim \mathcal{N}_n(0, I_n) \)
- \( q(x, y) = \frac{1}{(2\pi \sigma^2)^{n/2}} \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right) = q(y, x) \)
- Acceptance probability \( \alpha(x, y) = \frac{p(y)}{p(x)} \wedge 1 \).

Another standard choice: one step of overdamped Langevin (MALA):

- \( \hat{X}^{n}_{k+1} = X^n_k + \frac{\sigma^2}{2} (\nabla \ln p)(X^n_k) + \sigma G^k_{k+1} \) where \((G^k_k)_{k \geq 1}\) i.i.d. \( \sim \mathcal{N}_n(0, I_n) \)
- \( q(x, y) \neq q(y, x) \).

Question: How to choose \( \sigma \) as a function of the dimension \( n \)?
Previous work: \textit{Roberts, Gelman, Gilks 97}

Two fundamental assumptions:

- (H1) \textbf{Product target:} $p(x) = p(x_1, \ldots, x_n) = \prod_{i=1}^{n} e^{-V(x_i)}$,
- (H2) \textbf{Stationarity:} $X_0^n = (X_0^{1,n}, \ldots, X_0^{n,n}) \sim p(x)dx$ and thus
  \[
  \forall k, \ X_k^n = (X_k^{1,n}, \ldots, X_k^{n,n}) \sim p(x)dx.
  \]

Then, pick the first component $X_k^{1,n}$, choose

\[
\sigma_n = \frac{\ell}{\sqrt{n}},
\]

and rescale the time accordingly (diffusive scaling) by considering $(X_{\lfloor nt \rfloor}^{1,n})_{t \geq 0}$.

Under regularity assumptions on $V$, as $n \to \infty$, $(X_{\lfloor nt \rfloor}^{1,n})_{t \geq 0} \overset{(d)}{\Rightarrow} (X_t)_{t \geq 0}$

unique solution of the SDE

\[
dX_t = -h(\ell) \frac{1}{2} V'(X_t) \, dt + \sqrt{h(\ell)} \, dB_t,
\]

where $h(\ell) = 2\ell^2 \Phi \left( -\frac{\ell \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}}{2} \right)$ with $\Phi(x) = \int_{-\infty}^{x} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}$. 
Previous work: Roberts, Gelman, Gilks 97

Practical counterparts: (i) scaling of the variance proposal, (ii) scaling of the number of iterations.

Question: How to choose $\ell$ ?

- The function $\ell \mapsto h(\ell) = 2\ell^2 \Phi \left(-\frac{\ell \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}}{2}\right)$ is maximum at $\ell^* \simeq \frac{2.38}{\sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}}$.

- Besides, the limiting average acceptance rate is

$$\mathbb{E}[\alpha(X^n_k, \hat{X}^n_{k+1})] = \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{\sum_{i=1}^n (V(x_i) - V(y_i))} \wedge 1 q(x, y) e^{-\sum_{i=1}^n V(x_i)} dx dy$$

$$\longrightarrow_{n \to \infty} \text{acc}(\ell) = 2\Phi \left(-\frac{\ell \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}}{2}\right) \in (0, 1).$$

Observe that $\text{acc}(\ell^*) \simeq 0.234$, whatever $V$.

This justifies a constant acceptance rate strategy, with a target acceptance rate of approximately 25%.
A few references


- **Beyond (H1):** i. but non i.d. components RWM *Bédard 2007,2009*; finite range interactions *Breyer Roberts 2000*; mean-field interaction *Breyer Piccioni Scarlatti 2004*; density w.r.t. i.i.d. *Beskos Roberts Stuart 2009*; infinite-dimensional target with density w.r.t. Gaussian field RWM *Mattingly, Pillai, Stuart 2012*, MALA *Pillai, Stuart, Thiery 2012*.

- **Beyond (H2):** *Christensen, Roberts, Rosenthal 2005* Partial results for RWM and MALA with Gaussian target, *Pillai, Stuart, Thiery 2013* modified RWM for infinite-dimensional target with density w.r.t. Gaussian field.

**Aim of this work:** Study of the limit $n \rightarrow \infty$ *without* the stationarity assumption (H2).
Introduction

Optimal scaling of the transient phase of RWMH

Longtime convergence of the nonlinear SDE

Optimization strategies for the RWMH algorithm
The limit $n \to \infty$ without (H2)

We consider the RWMH with target $p(x) = \prod_{i=1}^{n} \exp(-V(x_i))$: $(G^i_k)_{i,k \geq 1}$ are i.i.d. $\sim \mathcal{N}(0, 1)$ independent of $(U_k)_{k \geq 1}$ i.i.d. $\sim \mathcal{U}[0, 1]$, and

$$
\begin{align*}
X_{k+1}^{i,n} &= X_k^{i,n} + \frac{\ell}{\sqrt{n}} G^i_{k+1} 1_{A_{k+1}}, ~ 1 \leq i \leq n, \\
with A_{k+1} &= \left\{ U_{k+1} \leq e^{\sum_{i=1}^{n} (V(X_k^{i,n}) - V(X_k^{i,n} + \frac{\ell}{\sqrt{n}} G^i_{k+1}))} \right\}.
\end{align*}
$$

From now on, we assume that $V$ is $C^3$ with $V''$ and $V^{(3)}$ bounded.

Theorem

Assume that

1. $m$ is a probability measure on $\mathbb{R}$ s.t. $\int_{\mathbb{R}} (V')^4(x) m(dx) < +\infty$,
2. $\forall n \geq 1$, $X_0^{1,n}, \ldots, X_n^{n,n}$ are i.i.d. according to $m$.

Then the process $(X_{\lfloor nt \rfloor}^{1,n})_{t \geq 0}$ converges in distribution to the unique solution of the SDE nonlinear in the sense of McKean: $X_0 \sim m$,

$$
dX_t = -\mathcal{G}(a(t), b(t)) V'(X_t) dt + \Gamma^{1/2}(a(t), b(t)) dB_t
$$

with $a(t) = \mathbb{E}[(V'(X_t))^2]$, $b(t) = \mathbb{E}[V''(X_t)]$, and...
The functions $\Gamma$ and $\mathcal{G}$

$$\Gamma(a, b) = \begin{cases} \ell^2 \Phi \left( -\frac{\ell b}{2\sqrt{a}} \right) + \ell^2 e^{\frac{\ell^2(a-b)}{2}} \Phi \left( \ell \left( \frac{b}{2\sqrt{a}} - \sqrt{a} \right) \right) & \text{if } a \in (0, +\infty), \\ \frac{\ell^2}{2} & \text{if } a = +\infty, \\ \ell^2 e^{-\frac{\ell^2 b^+}{2}} & \text{where } b^+ = \max(b, 0) \text{ if } a = 0, \end{cases}$$

$$\mathcal{G}(a, b) = \begin{cases} \ell^2 e^{\frac{\ell^2(a-b)}{2}} \Phi \left( \ell \left( \frac{b}{2\sqrt{a}} - \sqrt{a} \right) \right) & \text{if } a \in (0, +\infty), \\ 0 & \text{if } a = +\infty \text{ and } 1_{\{b > 0\}} \ell^2 e^{-\frac{\ell^2 b}{2}} & \text{if } a = 0. \end{cases}$$
Remarks

- **Limiting acceptance rate**: \( t \mapsto P(A_{\lfloor nt \rfloor}) \) converges to \( t \mapsto \text{acc}(a(t), b(t)) \) where \( a(t) = \mathbb{E}[(V'(X_t))^2] \), \( b(t) = \mathbb{E}[V''(X_t)] \) and
  \[
  \text{acc}(a, b) = \frac{1}{\ell^2} \Gamma(a, b).
  \]

- **Stationary case**: If \( m(dx) = e^{-V(x)} dx \), then \( \forall t \geq 0 \ X_t \sim e^{-V(x)} dx \) and \( a(t) = \mathbb{E}[(V'(X_t))^2] = \int_{\mathbb{R}} V'(V' e^{-V}) = \int_{\mathbb{R}} V'(-e^{-V})' = \int_{\mathbb{R}} V'' e^{-V} = \mathbb{E}[V''(X_t)] = b(t) \) are constant. Using the fact that for \( a > 0 \), \( \Gamma(a, a) = 2G(a, a) = 2\ell^2 \Phi(-\ell \sqrt{a}/2) \), we are back to the dynamics

  \[
  dX_t = -h(\ell) \frac{1}{2} V'(X_t) dt + \sqrt{h(\ell)} dB_t
  \]

  with \( h(\ell) = 2\ell^2 \Phi \left(-\ell \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)} \right) \).
Propogation of chaos

- One can actually prove a propagation of chaos result.

Definition

A sequence \((\chi_1^n, \ldots, \chi_n^n)_{n \geq 1}\) of exchangeable random variables is said to be \(\nu\)-chaotic if for fixed \(k \in \mathbb{N}^*\), the law of \((\chi_1^n, \ldots, \chi_k^n)\) converges in distribution to \(\nu^{\otimes k}\) as \(n\) goes to \(\infty\).

The processes \(((X_{[nt]}^1, \ldots, X_{[nt]}^n)_{t \geq 0})_{n \geq 1}\) are \(P\)-chaotic where \(P\) is the law of the unique solution to the SDE nonlinear in the sense of McKean: \(X_0 \sim m\)

\[
dX_t = -G(a(t), b(t)) V'(X_t) \, dt + \Gamma^{1/2}(a(t), b(t)) \, dB_t.
\]

with \(a(t) = \mathbb{E}[(V'(X_t))^2]\) and \(b(t) = \mathbb{E}[V''(X_t)]\).

- The assumption on the IC may then be replaced by: the initial positions \((X_{0,1}^n, \ldots, X_{0,n}^n)_{n \geq 1}\) are exchangeable, \(m\)-chaotic and s.t. \(\sup_n \mathbb{E}[(V'(X_{0,1}^n))^4] < \infty\).
Proof

The proof is based on:

- A weak formulation of the nonlinear SDE (martingale problem)
- Tightness arguments

This is a mean field limit.
Introduction

Optimal scaling of the transient phase of RWMH

Longtime convergence of the nonlinear SDE

Optimization strategies for the RWMH algorithm
We would like to understand the longtime behavior of the nonlinear SDE
\[ dX_t = -G(a(t), b(t))V'(X_t)dt + \Gamma^{1/2}(a(t), b(t)) dB_t, \]
where \( a(t) = \mathbb{E}[(V'(X_t))^2] \) and \( b(t) = \mathbb{E}[V''(X_t)]. \)

**Proposition**

*The probability measure \( e^{-V(x)} dx \) is the unique invariant measure for this SDE.*
Denoting by $\psi_t$ the density of $X_t$, one has

$$
\begin{aligned}
\partial_t \psi_t &= \partial_x \left( G(a[\psi_t], b[\psi_t]) V' \psi_t + \frac{1}{2} \Gamma(a[\psi_t], b[\psi_t]) \partial_x \psi_t \right), \\
a[\psi_t] &= \int (V'(x))^2 \psi_t(x) \, dx, \\
b[\psi_t] &= \int V''(x) \psi_t(x) \, dx.
\end{aligned}
$$

**Question 1:** Does $\psi_t$ converges to $\psi_\infty = \exp(-V)$?

**Question 2:** Is it possible to optimize the convergence, by appropriately choosing $\ell$ (recall that the variance of the proposal is $\ell^2/n$, and thus that $\Gamma(a, b) = \Gamma(a, b, \ell)$ and $G(a, b) = G(a, b, \ell)$)?
Entropy techniques

To analyze the longtime behavior, we use entropy techniques.

**Definition**

The probability measure $\nu$ satisfies a log-Sobolev inequality with constant $\rho > 0$ (in short LSI($\rho$)) if, for any probability measure $\mu$ absolutely continuous wrt $\nu$,

$$H(\mu|\nu) \leq \frac{1}{2\rho} I(\mu|\nu)$$

where

- $H(\mu|\nu) = \int \ln \left( \frac{d\mu}{d\nu} \right) d\mu$ is the Kullback-Leibler divergence (or relative entropy) of $\mu$ wrt $\nu$,

- $I(\mu|\nu) = \int \left| \nabla \ln \left( \frac{d\mu}{d\nu} \right) \right|^2 d\mu$ is the Fisher information of $\mu$ wrt $\nu$.

Roughly speaking, $e^{-V}$ satisfies LSI($\rho$) for some $\rho > 0$ if $V$ has at least quadratic growth at $\infty$.

In the Gaussian case $V(x) = \frac{x^2 + \ln(2\pi)}{2}$, $\exp(-V)$ satisfies LSI(1).
Entropy techniques

Recall the nonlinear FP equation:

$$\partial_t \psi_t = \partial_x \left( G(a[\psi_t], b[\psi_t]) V' \psi_t + \frac{1}{2} \Gamma(a[\psi_t], b[\psi_t]) \partial_x \psi_t \right).$$

We can prove exponential convergence of $\psi_t$ to the invariant density $\psi_\infty = e^{-V}$ in entropy.

**Theorem**

*If $X_0$ admits a density $\psi_0$ s.t. $\mathbb{E}[(V'(X_0))^2] < +\infty$ and $H(\psi_0|\psi_\infty) < \infty$, then*

$$\frac{d}{dt} H(\psi_t|\psi_\infty) \leq -\frac{b[\psi_t] \Gamma(a[\psi_t], b[\psi_t]) - 2a[\psi_t] G(a[\psi_t], b[\psi_t])}{2(b[\psi_t] - a[\psi_t])} I(\psi_t|\psi_\infty) < 0.$$  

*If moreover $\psi_\infty = e^{-V}$ satisfies LSI($\rho$), then there exists a positive and non-increasing function $\lambda : [0, +\infty) \to (0, +\infty)$ such that $\forall t \geq 0$

$$H(\psi_t|\psi_\infty) \leq e^{-t \lambda(H(\psi_0|\psi_\infty))} H(\psi_0|\psi_\infty).$$
Elements of proof

Writing $a, b$ for $a[\psi_t], b[\psi_t]$, one has

$$\frac{d}{dt} H(\psi_t|\psi_\infty) = \int_\mathbb{R} \partial_t \psi_t \ln \psi_t + \int_\mathbb{R} V \partial_t \psi_t$$

$$= - \frac{\Gamma(a, b)}{2} I(\psi_t|\psi_\infty) + (a - b)^2 \frac{2G(a, b) - \Gamma(a, b)}{2(b - a)} I(\psi_t|\psi_\infty),$$

where $\frac{2G(a,b) - \Gamma(a,b)}{2(b-a)} \geq 0$. Moreover,

$$(a - b)^2 = \left( \int_\mathbb{R} (V')^2 \psi_t - \int_\mathbb{R} V'' \psi_t \right)^2 = \left( \int_\mathbb{R} V'(V \partial_t \psi_t + \partial_x \psi_t) \right)^2$$

$$= \left( \int_\mathbb{R} V' \partial_x \ln(\psi_t/e^{-V})\psi_t \right)^2 \leq a I(\psi_t|\psi_\infty).$$

Hence

$$\frac{d}{dt} H(\psi_t|\psi_\infty) \leq - \frac{b\Gamma(a, b) - 2aG(a, b)}{2(b - a)} I(\psi_t|\psi_\infty).$$

If $\psi_\infty$ satisfies LSI($\rho$), then (i) $- I(\psi_t|\psi_\infty) \leq -2\rho H(\psi_t|\psi_\infty)$ and (ii) using the fact that $t \mapsto H(\psi_t|\psi_\infty)$ is decreasing, $\forall t \geq 0$, $2\rho \frac{b\Gamma(a, b) - 2aG(a, b)}{2(b - a)} \geq \lambda(H(\psi_0|\psi_\infty)) > 0$. 
Introduction

Optimal scaling of the transient phase of RWMH

Longtime convergence of the nonlinear SDE

Optimization strategies for the RWMH algorithm
Strategies to optimize the convergence of RWMH

We want to choose $\ell$ in order to accelerate the convergence to equilibrium. Two natural strategies: (i) optimize the exponential rate of convergence to zero of $H(\psi_t|\psi_\infty)$ (ii) choose $\ell$ in order to obtain a constant average acceptance rate.

Preliminary remark: When $b \leq 0$, one has
\[
\frac{d}{dt} H(\psi_t|\psi_\infty) \leq -\frac{\Gamma(a,b)}{2} \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t \quad \text{with} \quad \lim_{\ell \to \infty} \Gamma(a, b) = +\infty.
\]
So one should choose $\ell$ as large as possible.

From now on, suppose that $b > 0$ (recall that in the longtime limit $b = a > 0$).

We have:
\[
\frac{d}{dt} H(\psi_t|\psi_\infty) \leq -\frac{b\Gamma(a,b) - 2aG(a,b)}{2(b - a)} \underbrace{I(\psi_t|\psi_\infty)}_{\frac{1}{b} F(\frac{a}{b}, \sqrt{b}, \ell)} < 0,
\]

where
\[
F(s, \ell) = \begin{cases} 
\ell^2 e^{-\frac{\ell^2}{2}} & \text{if } s = 0, \\
2\ell^2 \left( \frac{1 + \frac{\ell^2}{4}}{2} \Phi \left( -\frac{\ell}{2} \right) - \frac{\ell}{2\sqrt{2\pi}} e^{-\frac{\ell^2}{8}} \right) & \text{if } s = 1, \\
\frac{\ell^2}{1-s} \left( \Phi \left( -\frac{\ell}{2\sqrt{s}} \right) + (1-2s)e^{\frac{\ell^2(s-1)}{2}} \Phi \left( \frac{\ell}{2\sqrt{s}} - \ell \sqrt{s} \right) \right) & \text{if } 0 < s \neq 1.
\end{cases}
\]
Choice of $\ell$ maximizing the exponential rate of cv

**Lemma**

Let $b > 0$. Then $\ell \geq 0 \mapsto \frac{1}{b} F\left(\frac{a}{B}, \ell \sqrt{b}\right)$ admits a unique maximum at point $\ell^* (a, b)$. Moreover

$$\ell^* (a, b) = \frac{1}{\sqrt{b}} \ell^* \left(\frac{a}{b}\right)$$

where for any $s \geq 0$, $\ell^* (s)$ realizes the unique maximum of $\ell \mapsto F(s, \ell)$. The function $s \mapsto \ell^* (s)$ is continuous on $[0, +\infty)$ and

- $\tilde{\ell}^* (a, b) \sim_{a/b \to 0} \frac{\ell^* (0)}{\sqrt{b}} = \frac{\sqrt{2}}{\sqrt{b}}$.
- $\tilde{\ell}^* (a, b) \sim_{a/b \to 1} \frac{\ell^* (1)}{\sqrt{b}}$.
- $\tilde{\ell}^* (a, b) \sim_{a/b \to +\infty} \frac{x^* \sqrt{a}}{b}$ where $x^* \approx 1.22$.

**Remark:** Since $dV(X_t) = V'(X_t) \left(\sqrt{\Gamma(a, b)} dB_t - \mathcal{G}(a, b) V'(X_t))dt\right) + \frac{1}{2} \Gamma(a, b) V''(X_t) dt$, we have $\frac{d}{dt} \mathbb{E}[V(X_t)] = \frac{1}{2} (b \Gamma(a, b) - 2a \mathcal{G}(a, b))$ and $\tilde{\ell}^* (a, b)$ also maximizes $|\frac{d}{dt} \mathbb{E}[V(X_t)]|$.
Figure: Solid line: the function $s \mapsto \ell^*(s)$. Dashed line: the function: $s \mapsto x^* \sqrt{s}$. 
Comparison with constant acceptance rate strategies

Recall that the limiting mean acceptance rate is

$$\text{acc}(a, b, \ell) = \frac{1}{\ell^2} \Gamma(a, b) = G\left(\frac{a}{b}, \ell \sqrt{b}\right)$$

where

$$G(s, \ell) = \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + e^{\frac{\ell^2(s-1)}{2}} \Phi\left(\ell \left(\frac{1}{2\sqrt{s}} - \sqrt{s}\right)\right).$$

Lemma

For $s > 0$, the function $\ell \mapsto G(s, \ell)$ is decreasing. Moreover, for $\alpha \in (0, 1)$, the unique $\ell$ s.t. $\text{acc}(a, b, \ell) = \alpha$ is

$$\tilde{\ell}^\alpha(a, b) = \frac{1}{\sqrt{b}} \ell^\alpha\left(\frac{a}{b}\right)$$

where $\ell^\alpha(s)$ is the unique solution to $G(s, \ell^\alpha(s)) = \alpha$. Last,

- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \to 0} \frac{\sqrt{-2 \ln(\alpha)}}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \to 1} \frac{\ell^\alpha(1)}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \to \infty} -2\Phi^{-1}(\alpha) \frac{\sqrt{a}}{b}$. 
Comparison with constant acceptance rate strategies

**Remark 1**: Notice that \( \tilde{\ell}^*(a, b) = \frac{1}{\sqrt{b}} \ell^* \left( \frac{a}{b} \right) \) and \( \tilde{\ell}^\alpha(a, b) = \frac{1}{\sqrt{b}} \ell^\alpha \left( \frac{a}{b} \right) \) have the same scaling in \((a, b)\).

\( \rightarrow \) Constant acceptance rate strategy seems sensible.

**Remark 2**: Choice of \( \alpha \): how to choose \( \alpha \) to get \( \tilde{\ell}^*(a, b) \sim \tilde{\ell}^\alpha(a, b) \)?

- \( a/b \to 0 \): \( \alpha = \frac{1}{e} \approx 0.37 \).
- \( a/b \to 1 \): \( \alpha \) such that \( \ell^\alpha(1) = \ell^*(1) \), namely \( \alpha \approx 0.35 \).
- \( a/b \to \infty \): \( \alpha = \Phi(-x^*/2) \approx 0.27 \).

(Recall that the standard choice for the RWM under the stationarity assumption is \( \alpha = 0.234 \).)

\( \rightarrow \) Constant acceptance rate with \( \alpha \in (1/4, 1/3) \) seems sensible.

Let us plot the relative difference in terms of exponential rate of convergence, for the three values \( \alpha = \frac{1}{e} \approx 0.37 \), \( \alpha \approx 0.35 \) and \( \alpha = \Phi(-x^*/2) \approx 0.27 \).
Figure: $\frac{F(\frac{a}{b}, l^*(a, b)\sqrt{b}) - F(\frac{\alpha}{b}, l^\alpha(a, b)\sqrt{b})}{F(\frac{a}{b}, l^*(a, b)\sqrt{b})}$ as function of $a$ for $b = 1, 0.1, 10$ and $\alpha \approx 0.27$ solid line, $\alpha \approx 0.35$ dashed line, $\alpha = e^{-1} \approx 0.37$ dotted line. $\alpha \approx 0.27$ seems to be the best compromise.
Gaussian target: \( V(x) = \frac{1}{2} (x^2 + \ln(2\pi)) \)

Setting \( m(t) \overset{\text{def}}{=} \mathbb{E}[X_t] = \int_{\mathbb{R}} x \psi_t(x) dx \) and \( s(t) \overset{\text{def}}{=} \mathbb{E}[(X_t)^2] = \int_{\mathbb{R}} x^2 \psi_t(x) dx \), one has

\[
H(\psi_t|\psi_\infty) = \frac{1}{2} \left( s(t) - \ln(s(t) - m(t)^2) - 1 \right),
\]

\[
\frac{d}{dt} H(\psi_t|\psi_\infty) = \frac{1}{2} \left( F(s, \ell)(1 - s) - \frac{F(s, \ell)(1 - s) + 2mG(s, 1, \ell)}{s - m^2} \right).
\]

It is possible to compute numerically \( \ell^{\text{ent}}(m, s) \) maximizing \( \left| \frac{d}{dt} H(\psi_t|\psi_\infty) \right| \).

To assess the convergence, we compute

\[
t_0 \mapsto \hat{I}_{t_0,t_0+T}^m = \frac{1}{T} \sum_{k=t_0+1}^{t_0+T} \frac{X_{k,1}^{1,n} + \ldots + X_{k,n}^{1,n}}{n}
\]

\[
t_0 \mapsto \hat{I}_{t_0,t_0+T}^s = \frac{1}{T} \sum_{k=t_0+1}^{t_0+T} \frac{(X_{k,1}^{1,n})^2 + \ldots + (X_{k,n}^{n,n})^2}{n}.
\]
Figure: $t_0 \mapsto$ square bias of $(\hat{s}^{t_0, T+t_0}, \hat{m}^{t_0, T+t_0}, (X_0^{1,n}, \ldots, X_0^{n,n}) = (10, \ldots, 10), n = 100(\epsilon^{0.27} - A \rightarrow$ adaptive scaling Metropolis algorithm and $\epsilon^{0.27} - N \rightarrow$ numerical approximation of $\epsilon^{0.27}(s, 1).$}
Conclusions:

1. The constant $\ell$ strategy is bad;
2. The constant average acceptance rate strategy (using $\ell^\alpha$) leads to very close convergence curves compared to the optimal exponential rate of convergence strategy (using $\ell^*$);
3. The optimal exponential rate of convergence strategy is as good as the most optimal strategy one could design in terms of entropy decay (using $\ell^\text{ent}$).
Example of non Gaussian target

\[ V(x) = \begin{cases} 
(x - 1)^2(x + 1)^2 & \text{if } |x| \leq 1, \\
4x^2 - 8|x| + 4 & \text{otherwise.} 
\end{cases} \]

\[ I = \int_{\mathbb{R}} (V')^2 e^{-V} = 4.07 \text{ so that } \frac{2.38}{\sqrt{I}} = 1.18 \]

\[ X_0^{i,n} \text{ i.i.d. } \sim \mathcal{N}_1(1, 0.143) \text{ so that } \]
\[ \mathbb{E}[(V'(X_0^{1,n}))^2] = \mathbb{E}[V''(X_0^{1,n})] = 5.24 \]
The constant acceptance rate strategies are implemented using an adaptive scaling Metropolis algorithm.
References

Optimal scaling of the transient phase of MALA (1)

Consider the MALA algorithm:

\[
X_{k+1}^{i,n} = X_k^{i,n} + \left( \sigma_n G_{k+1} - \frac{\sigma_n^2}{2} V'(X_k^{i,n}) \right) Z_{k+1}^{i,n} \quad 1 \leq i \leq n
\]

where \( A_{k+1} = \begin{cases} U_{k+1} \leq e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + Z_{k+1}^{i,n})) + \frac{1}{2} [(G_{k+1}^i)^2 - (G_{k+1}^i - \frac{\sigma_n}{2} (V'(X_k^{i,n}) + V'(X_k^{i,n} + Z_{k+1}^{i,n}))^2) ]} \end{cases} \)

For \( \sigma_n = \frac{\ell}{n^{1/4}} \) and \(((X_0^{1,n}, \ldots, X_0^{n,n}))_{n \geq 1} m\text{-chaotic, one expects prop. of chaos for the processes} ((X_{\left\lfloor \sqrt{n}t \right\rfloor}^{1,n}, \ldots, X_{\left\lfloor \sqrt{n}t \right\rfloor}^{n,n})_{t \geq 0})_{n \geq 1} \) to the law of

\[
\begin{cases}
\frac{dX_t}{dt} = \sqrt{w(t, \ell)} dB_t - w(t, \ell) \frac{1}{2} V'(X_t) \ dt, \ X_0 \sim m(dx)
\end{cases}
\]

where \( w(t, \ell) = \ell^2 \left( e^{\frac{\ell^4}{8} \mathbb{E}((V'(X_t)^2 + V(4)^2 - 2 V(3)^2 V' - (V'')^2)(X_t)) \wedge 1} \right) \).

Remark: If \( V(x) = \frac{x^2 + \ln(2\pi)}{2} \), then

\[
\frac{d}{dt} \mathbb{E}(X_t^2) = \ell^2 \left( e^{\frac{\ell^4}{8} \mathbb{E}(X_t^2) - 1} \wedge 1 \right) \left( 1 - \mathbb{E}(X_t^2) \right) \quad [\text{Christensen, Roberts, Rosenthal 2005}].
\]
Optimal scaling of the transient phase of MALA (2)

\[ w(t, \ell) = \ell^2 \left( e^{\frac{\ell^4}{8}} \mathbb{E}\left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_t) \right) \land 1 \right) \]

- on time intervals such that \[ \mathbb{E}\left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_t) \right) < 0, \] then \[ \ell \mapsto w(t, \ell) \] maximum at \[ \ell^* \approx 1.42 \frac{1}{\mathbb{E}^{1/4}((2V^{(3)} V' + (V'')^2 - (V')^2 V'' - V^{(4)})(X_t))} \]

- on time intervals such that \[ \mathbb{E}\left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_t) \right) = 0 \] (this is in particular the case at equilibrium), the correct scaling [Roberts, Rosenthal 1998] is

\[ \sigma_n = \frac{\ell}{n^{1/6}} \]

and one obtain a diffusive limit for \((X_{[n^{1/3} t]}^1, n)_{t \geq 0}\). At equilibrium, there exists an optimal \(\ell = \ell^*\) and \(\text{acc}(\ell^*) = 0.574\).

- on time intervals such that \[ \mathbb{E}\left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_t) \right) > 0, \] with the scaling \(\sigma_n = \frac{\ell}{n^{1/4}}\), we have \(w(t, \ell) = \ell^2 \to +\infty\) as \(\ell \to +\infty\). One should take \(\sigma_n \gg \frac{\ell}{n^{1/4}}\).
Optimal scaling of the transient phase of MALA (3)

The case $\mathbb{E} \left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2) (X_t) \right) > 0$: one should take $\sigma_n$ going to zero as slowly as possible.

Let us consider the Gaussian case $V(x) = (x^2 + \ln(2\pi))/2$, so that $\mathbb{E} \left( ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2) (X_t) \right) = \mathbb{E}(X_t^2 - 1)$.

**Proposition**

If the initial random variables $(X_0^{1,n}, \ldots, X_0^{n,n})$ are i.i.d. according to $m$ such that $\langle m, x^2 - 1 \rangle > 0$ and $\langle m, x^8 \rangle < +\infty$, and $\sigma_n$ satisfies:

$$\lim_{n \to \infty} \sigma_n = 0 \quad \text{and} \quad \lim_{n \to \infty} n\sigma_n^2 = +\infty,$$

then the processes $((X_{t/\sigma_n^2})_{t \geq 0}, \ldots, (X_{t/\sigma_n^2})_{t \geq 0})$ are Q-chaotic where Q denotes the law of the Ornstein-Uhlenbeck process $dX_t = dB_t - \frac{X_t}{2} dt$, $X_0 \sim m$. Moreover, the limiting mean acceptance rate is 1.

Remark: this result still holds if $\lim_{n \to \infty} n\sigma_n^2 = 0$. 
## In summary

<table>
<thead>
<tr>
<th></th>
<th>RWMH</th>
<th>MALA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equilibrium</strong></td>
<td>( \sigma_n = \frac{\ell}{\sqrt{n}} ), acc((\ell^*)) = 0.234</td>
<td>( \sigma_n = \frac{\ell}{n^{1/6}} ), acc((\ell^*)) = 0.574</td>
</tr>
<tr>
<td><strong>Transient</strong></td>
<td>( \sigma_n = \frac{\ell}{\sqrt{n}} ), acc((\ell^*)) = 0.27</td>
<td>( \sigma_n = \frac{\ell}{n^{1/4}} ), optimal (\ell) ??</td>
</tr>
</tbody>
</table>

In all cases, the associated timescale is the diffusive one:

\[
\left( X_{\lfloor t/\sigma_n^2 \rfloor}^{1,n} \right)_{t \geq 0}.
\]