



Optimal scaling for the transient phase of the random walk Metropolis Hastings algorithm

Benjamin Jourdain

Joint work with T. Lelièvre and B. Miasojedow

Computational methods for statistical mechanics - ICMS - 6 june
2014



Metropolis Hastings algorithm

- Sampling of a target probability measure with density p on \mathbb{R}^n
- At time k , given X_k^n , $\hat{X}_{k+1}^n \sim q(X_k^n, y) dy$, where $q(x, y)$ Markov density kernel on \mathbb{R}^n ,
- With probability $\frac{p(\hat{X}_k^n)q(\hat{X}_k^n, X_k^n)}{p(X_k^n)q(X_k^n, \hat{X}_k^n)} \wedge 1$ move accepted i.e. $X_{k+1}^n = \hat{X}_{k+1}^n$.
Otherwise $X_{k+1}^n = X_k^n$.
- $(X_k^n)_{k \geq 0}$ Markov chain with reversible measure $p(x) dx$.
- Gaussian random walk :
 - $\hat{X}_{k+1}^n = X_k^n + \sigma G_{k+1}$ where $(G_k)_{k \geq 1}$ 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 $\alpha(x, y) = \frac{p(y)}{p(x)} \wedge 1$.

How to choose σ in function of the dimension n ?



Previous work: *Roberts, Gelman, Gilks 97*

Two fundamental assumptions:

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

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 $(X_{[nt]}^{1,n})_{t \geq 0}$.

Under regularity assumptions on V , as $n \rightarrow \infty$, $(X_{[nt]}^{1,n})_{t \geq 0} \xrightarrow{(d)} (X_t)_{t \geq 0}$ 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: scaling of the variance proposal.

Question: How to choose ℓ ?

- 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

$$\mathbb{E}[\alpha(X_k^n, \hat{X}_{k+1}^n)] = \int_{\mathbb{R}^n \times \mathbb{R}^n} \underbrace{e^{\sum_{i=1}^n (V(x_i) - V(y_i))} \wedge 1}_{\alpha(x,y)} q_n(x, y) e^{-\sum_{i=1}^n V(x_i)} dx dy$$
$$\rightarrow_{n \rightarrow \infty} a(\ell) = 2\Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}}(V')^2 \exp(-V)}}{2}\right) \in (0, 1).$$

We observe that $a(\ell^*) \simeq 0.234$, whatever V .

This justifies a **constant acceptance rate strategy**, with a target acceptance rate of approximately 25%.



A few references on scaling

- **(H1) + (H2) and various proposals:** Gaussian RWM *Roberts Gelman Gilks 1997*, MALA *Roberts Rosenthal 1997*, nonGaussian RWM *Neal Roberts 2011*, RWM discontinuous target *Neal Roberts Yuen 2012*, Multiple try MCMC *Bédard Douc Moulines 2012*, Delayed rejection MCMC *Bédard Douc Moulines 2013*, Hybrid Monte Carlo *Beskos Pillai Roberts Sanz-Serna Stuart 2013*.
- **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, *Ottobre Stuart 2014*

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



Outline of the talk

- 1 Optimal scaling of the transient phase of RWMH
- 2 Longtime convergence of the nonlinear SDE
- 3 Optimization strategies for the RWMH algorithm



Main result : RWMH target $p(x) = \prod_{i=1}^n \exp(-V(x_i))$

Let $(G_k^i)_{i,k \geq 1}$ be i.i.d. $\sim \mathcal{N}_1(0, 1)$ indep. $(U_k)_{k \geq 1}$ i.i.d. $\sim \mathcal{U}[0, 1]$.

$$\left\{ \begin{array}{l} X_{k+1}^{i,n} = X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i \mathbf{1}_{\mathcal{A}_{k+1}}, \quad 1 \leq i \leq n, \\ \text{with } \mathcal{A}_{k+1} = \left\{ U_{k+1} \leq e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i))} \right\}. \end{array} \right.$$

From now on, we assume that V is C^3 with V'' and $V^{(3)}$ bounded and m is a probability measure on \mathbb{R} such that $\int_{\mathbb{R}} (V')^4(x) m(dx) < +\infty$.

Theorem 1

Assume that $\forall n \geq 1, X_0^{1,n}, \dots, X_0^{n,n}$ are i.i.d. $\sim m$. Then the process $(X_{[nt]}^{1,n})_{t \geq 0}$ converges in distribution to the unique solution of the SDE nonlinear in the sense of McKean: $X_0 \sim m$,

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

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



The functions Γ and \mathcal{G}

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

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

For $a > 0$, $\Gamma(a, a) = 2\mathcal{G}(a, a) = 2\ell^2 \Phi(-\ell\sqrt{a}/2)$.

If $m(dx) = e^{-V(x)} dx$, $\forall t \geq 0$, $X_t \sim e^{-V(x)} dx$ ($X_k^{i,n} \sim e^{-V(x)} dx$) and

$$\mathbb{E}[(V'(X_t))^2] = \int_{\mathbb{R}} V'(V'e^{-V}) = \int_{\mathbb{R}} V'(-e^{-V})' = \int_{\mathbb{R}} V'' e^{-V} = \mathbb{E}[V''(X_t)]$$

$$dX_t = \sqrt{h(\ell)} dB_t - h(\ell) \frac{1}{2} V'(X_t) dt \text{ with } h(\ell) = 2\ell^2 \Phi\left(-\frac{\ell}{2} \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}\right)$$



The functions Γ and \mathcal{G}

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

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

For $a > 0$, $\Gamma(a, a) = 2\mathcal{G}(a, a) = 2\ell^2 \Phi(-\ell\sqrt{a}/2)$.

If $m(dx) = e^{-V(x)} dx$, $\forall t \geq 0$, $X_t \sim e^{-V(x)} dx$ ($X_k^{i,n} \sim e^{-V(x)} dx$) and

$$\mathbb{E}[(V'(X_t))^2] = \int_{\mathbb{R}} V'(V' e^{-V}) = \int_{\mathbb{R}} V'(-e^{-V})' = \int_{\mathbb{R}} V'' e^{-V} = \mathbb{E}[V''(X_t)]$$

$$dX_t = \sqrt{h(\ell)} dB_t - h(\ell) \frac{1}{2} V'(X_t) dt \text{ with } h(\ell) = 2\ell^2 \Phi\left(-\frac{\ell}{2} \sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}\right)$$



Properties of Γ and \mathcal{G}

Lemma 2

- 1 $\forall (a, b) \in [0, +\infty] \times \mathbb{R}$, $0 \leq \mathcal{G}(a, b) \leq \Gamma(a, b) \leq I^2$,
- 2 *the function Γ is continuous on $[0, +\infty] \times \mathbb{R}$ and such that $\inf_{(a,b) \in [0, +\infty] \times [\inf V'', \sup V'']} \Gamma(a, b) > 0$,*
- 3 *the function \mathcal{G} is continuous on $\{[0, +\infty] \times \mathbb{R}\} \setminus \{(0, 0)\}$,*
- 4 $\exists C < +\infty$, $\forall (a, b)$ and $(a', b') \in [0, +\infty] \times [\inf V'', \sup V'']$,

$$\begin{aligned} |\Gamma(a, b) - \Gamma(a', b')| + (\sqrt{a} \wedge \sqrt{a'}) |\mathcal{G}(a, b) - \mathcal{G}(a', b')| \\ \leq C \left(|b' - b| + |a' - a| + |\sqrt{a'} - \sqrt{a}| \right). \end{aligned}$$

$\sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])$ and $\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) V'(X_t)$ the
coefs of the SDE have the same regularity in terms of
 $(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])$ by 2+4 \Rightarrow uniqueness by comp. $d(X_t - \tilde{X}_t)^2$



Mean field interaction

Let $(x_1, \dots, x_n) \in \mathbb{R}^n$, $\eta^n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $(G^i)_{1 \leq i \leq n}$ i.i.d. $\sim \mathcal{N}_1(0, 1)$.

$$\begin{aligned} \mathcal{E}_{k+1} &\stackrel{\text{def}}{=} \mathbb{E} \left(\left(\sum_{i=1}^n (V(x_i) - V(x_i + \frac{\ell G^i}{\sqrt{n}})) + \overbrace{\sum_{i=1}^n (V'(x_i) \frac{\ell G^i}{\sqrt{n}} + \frac{V''(x_i) \ell^2}{2n})}^{\sim \mathcal{N}_1(\frac{\ell^2}{2} \langle \eta^n, V'' \rangle, \ell^2 \langle \eta^n, (V')^2 \rangle)} \right)^2 \right) \\ &= \frac{\ell^4}{4n^2} \mathbb{E} \left(\left(\sum_{i=1}^n (V''(x_i)(1 - (G^i)^2) - V^{(3)}(x_i) \frac{\ell}{3\sqrt{n}} (G^i)^3) \right)^2 \right) \end{aligned}$$

with $\chi_i \in [x_i, x_i + \frac{\ell G^i}{\sqrt{n}}]$ only depending on G^i . For $i \neq j$,

$$\begin{aligned} \mathbb{E}[V''(x_i)(1 - (G^i)^2) \{V''(x_j)(1 - (G^j)^2) - V^{(3)}(\chi_j) \frac{\ell}{3\sqrt{n}} (G^j)^3\}] \\ = V''(x_i) \mathbb{E}[1 - (G^i)^2] \mathbb{E}[\dots] = 0 \end{aligned}$$

With boundedness of V'' and $V^{(3)}$, one concludes $\mathcal{E}_{k+1} \leq \frac{C}{n}$.



Mean field interaction

Let now $\mu_k^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}$. The evolution of the RWM algorithm writes

$$X_{k+1}^{i,n} = X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i \mathbf{1}_{\{U_{k+1} \leq e^{\ell \sqrt{\langle \mu_k^n, (V')^2 \rangle}} G_{k+1} - \frac{\ell^2}{2} \langle \mu_k^n, V'' \rangle + O(n^{-1/2})\}}, \quad 1 \leq i \leq n$$

where $G_{k+1} \sim \mathcal{N}_1(0, 1)$ independent of the positions up to time n and such that

$$\mathbb{E} \left(G_{k+1}^i G_{k+1} \right) = \frac{\mathbb{E}(V'(X_k^{i,n}))}{\sqrt{n}}.$$

Gaussian calculations + diffusion approximation techniques lead to Theorem 1



- 1 Optimal scaling of the transient phase of RWMH
- 2 Longtime convergence of the nonlinear SDE
- 3 Optimization strategies for the RWMH algorithm



Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) V'(X_t) dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) dB_t.$$

Proposition 3

The probability measure $e^{-V(x)} dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing $m(dx) = e^{-V(x)} dx$ in the main theorem, one obtains that this measure is invariant.
- - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞ ,
 - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
 - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V'' \psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.



Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) V'(X_t) dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) dB_t.$$

Proposition 3

The probability measure $e^{-V(x)} dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing $m(dx) = e^{-V(x)} dx$ in the main theorem, one obtains that this measure is invariant.
 - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞ ,
 - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
 - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V'' \psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.



Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) V'(X_t) dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)]) dB_t.$$

Proposition 3

The probability measure $e^{-V(x)} dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing $m(dx) = e^{-V(x)} dx$ in the main theorem, one obtains that this measure is invariant.
- - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞ ,
 - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
 - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V'' \psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.



Fokker-Planck equation

Denoting by ψ_t the density of X_t , one has

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

Question 1: Does ψ_t converge to $\psi_\infty = \exp(-V)$?

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



Fokker-Planck equation

To analyze the longtime behavior, we use entropy techniques.

Definition 4

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

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

- $H(\mu|\nu) = \int \ln \left(\frac{d\mu}{d\nu} \right) d\mu$ is the Kullback-Leibler divergence (or relative entropy) of μ wrt ν ,
- $I(\mu|\nu) = \int \left| \nabla \ln \left(\frac{d\mu}{d\nu} \right) \right|^2 d\mu$ is the Fisher information of μ wrt ν .



Convergence to the invariant density $\psi_\infty = e^{-V}$

Theorem 5

If X_0 admits a density ψ_0 s.t. $\mathbb{E}[(V'(X_0))^2] < +\infty$ and $H(\psi_0|\psi_\infty) < \infty$, then

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

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

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

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

In the Gaussian case $V(x) = \frac{x^2 + \ln(2\pi)}{2}$, $\mathcal{N}_1(0, 1)$ satisfies $LSI(1)$.



Elements of proof

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

$$\begin{aligned} \frac{d}{dt} H(\psi_t | \psi_\infty) &= \int_{\mathbb{R}} \partial_t \psi_t \ln \psi_t + \int_{\mathbb{R}} V \partial_t \psi_t \\ &= -\frac{\Gamma(a, b)}{2} I(\psi_t | \psi_\infty) + (a - b)^2 \times \left\{ \frac{2\mathcal{G}(a, b) - \Gamma(a, b)}{2(b - a)} \right\}_{\geq 0} \\ (a - b)^2 &= \left(\int_{\mathbb{R}} (V')^2 \psi_t - \int_{\mathbb{R}} V'' \psi_t \right)^2 = \left(\int_{\mathbb{R}} V' (V' \psi_t + \partial_x \psi_t) \right)^2 \\ &= \left(\int_{\mathbb{R}} V' \partial_x \ln(\psi_t / e^{-V}) \psi_t \right)^2 \leq a \times I(\psi_t | \psi_\infty). \end{aligned}$$

Hence $\frac{d}{dt} H(\psi_t | \psi_\infty) \leq -\frac{b\Gamma(a, b) - 2a\mathcal{G}(a, b)}{2(b - a)} I(\psi_t | \psi_\infty)$. When ψ_∞ satisfies LSI(ρ), it satisfies the transport inequality $W_2^2(\psi_t, \psi_\infty) \leq \frac{2}{\rho} H(\psi_t | \psi_\infty)$. With $t \mapsto H(\psi_t | \psi_\infty) \searrow \Rightarrow \sup_t a[\psi_t] < C(H(\psi_0 | \psi_\infty))$ with $C \nearrow$.

$$\lambda(H(\psi_0 | \psi_\infty)) \stackrel{\text{def}}{=} \frac{1}{2\rho} \inf_{(a, b): a \leq C(H(\psi_0 | \psi_\infty))} \frac{b\Gamma(a, b) - 2a\mathcal{G}(a, b)}{2(b - a)} > 0.$$



- 1 Optimal scaling of the transient phase of RWMH
- 2 Longtime convergence of the nonlinear SDE
- 3 Optimization strategies for the RWMH algorithm



Decrease of the Kullback-Leibler divergence

When $b \leq 0$, one has $\frac{d}{dt} H(\psi_t | \psi_\infty) \leq -\frac{\Gamma(a,b)}{2} \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t$ with $\lim_{\ell \rightarrow \infty} \Gamma(a,b) = +\infty$. So one should choose ℓ as large as possible. From now on, suppose that $b > 0$ (recall that in the longtime limit $b = a > 0$).

$$\frac{d}{dt} H(\psi_t | \psi_\infty) \leq - \underbrace{\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{2(b-a)}}_{\frac{1}{b} F\left(\frac{a}{b}, \ell\sqrt{b}\right)} I(\psi_t | \psi_\infty) < 0,$$

where

$$F(s, \ell) = \begin{cases} \ell^2 e^{-\frac{\ell^2}{2}} & \text{if } s = 0 \\ 2\ell^2 \left(\left(1 + \frac{\ell^2}{4}\right) \Phi\left(-\frac{\ell}{2}\right) - \frac{\ell}{2\sqrt{2\pi}} e^{-\frac{\ell^2}{8}} \right) & \text{if } s = 1 \\ \frac{\ell^2}{1-s} \left(\Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + (1-2s)e^{\frac{\ell^2(s-1)}{2}} \Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) \right) & \text{if } 0 < s \neq 1 \end{cases}$$



Choice of ℓ maximizing the exponential rate of cv

Lemma 6

Let $b > 0$. Then $\tilde{\ell}^*(a, b) = \operatorname{argmax}_{\ell \geq 0} \frac{1}{b} F\left(\frac{a}{b}, \ell\sqrt{b}\right) = \frac{1}{\sqrt{b}} \ell^*\left(\frac{a}{b}\right)$ where for any $s \geq 0$, $\ell^*(s)$ realizes the unique maximum of $\ell \mapsto F(s, \ell)$. Moreover, $s \mapsto \ell^*(s)$ is continuous on $[0, +\infty)$ and

- $\tilde{\ell}^*(a, b) \underset{a/b \rightarrow 0}{\sim} \frac{\ell^*(0)}{\sqrt{b}} = \frac{\sqrt{2}}{\sqrt{b}}$.
- $\tilde{\ell}^*(a, b) \underset{a/b \rightarrow 1}{\sim} \frac{\ell^*(1)}{\sqrt{b}}$.
- $\tilde{\ell}^*(a, b) \underset{a/b \rightarrow +\infty}{\sim} \frac{x^* \sqrt{a}}{b}$ where $x^* \simeq 1.22$.

Notice that

$$dV(X_t) = V'(X_t) \left(\sqrt{\Gamma(a, b)} dB_t - \mathcal{G}(a, b) V'(X_t) dt \right) + \frac{1}{2} \Gamma(a, b) V''(X_t) dt$$

so that $\frac{d}{dt} \mathbb{E}[V(X_t)] = \frac{1}{2} (b\Gamma(a, b) - 2a\mathcal{G}(a, b)) = \frac{b-a}{b} F\left(\frac{a}{b}, \ell\sqrt{b}\right)$ and $\tilde{\ell}^*(a, b)$ also maximizes $\left| \frac{d}{dt} \mathbb{E}[V(X_t)] \right|$.



Comparison with constant acceptance rate strategies

The limiting mean acceptance rate in Theorem 1 is

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

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

Lemma 7

For $s > 0$, the function $\ell \mapsto H(s, \ell)$ is decreasing. Moreover, for $\alpha \in (0, 1)$, the unique ℓ 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 $H(s, \ell^\alpha(s)) = \alpha$. Last,

- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \rightarrow 0} \frac{\sqrt{-2 \ln(\alpha)}}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \rightarrow 1} \frac{\ell^\alpha(1)}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a, b) \sim_{a/b \rightarrow \infty} -2\Phi^{-1}(\alpha) \frac{\sqrt{a}}{b}$.



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) .

→ Constant acceptance rate strategy seems sensible.

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

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

(The standard choice for the RWM, under the stationarity assumption, is $\alpha = 0.234$.)

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

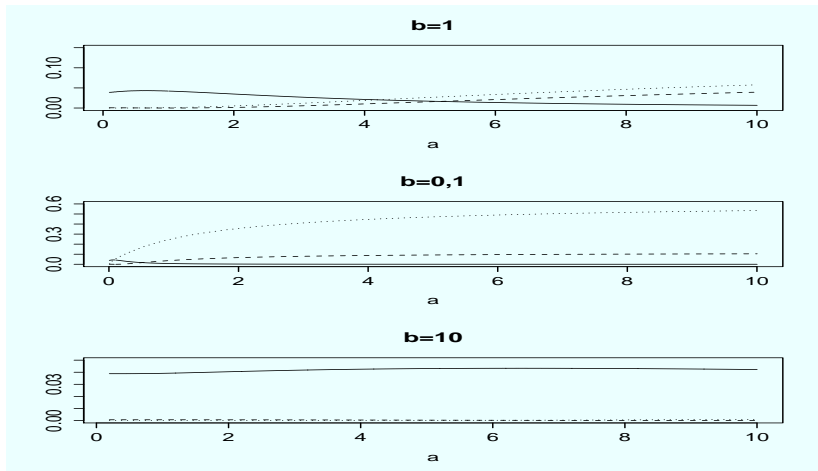


Figure: $\frac{F(\frac{a}{b}, \tilde{I}^*(a,b)\sqrt{b}) - F(\frac{a}{b}, \tilde{I}^\alpha(a,b)\sqrt{b})}{F(\frac{a}{b}, \tilde{I}^*(a,b)\sqrt{b})}$ as function of a for $b = 1, 0.1, 10$ and $\alpha \simeq 0.27$ solid line, $\alpha \simeq 0.35$ dashed line, $\alpha = e^{-1} \simeq 0.37$ dotted line.



Gaussian target : $V(x) = \frac{1}{2}(x^2 + \ln(2\pi))$

We assume that ψ_0 Gaussian $\Rightarrow \psi_t$ Gaussian.

Setting $m(t) \stackrel{\text{def}}{=} \mathbb{E}[X_t] = \int_{\mathbb{R}} x \psi_t(x) dx$ and

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

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

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

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

To assess the convergence, we compute

$$t_0 \mapsto \hat{I}_{t_0, t_0+T}^m = \frac{1}{T} \sum_{k=t_0+1}^{t_0+T} \frac{X_k^{1,n} + \dots + X_k^{n,n}}{n}$$

$$t_0 \mapsto \hat{I}_{t_0, t_0+T}^s = \frac{1}{T} \sum_{k=t_0+1}^{t_0+T} \frac{(X_k^{1,n})^2 + \dots + (X_k^{n,n})^2}{n}.$$

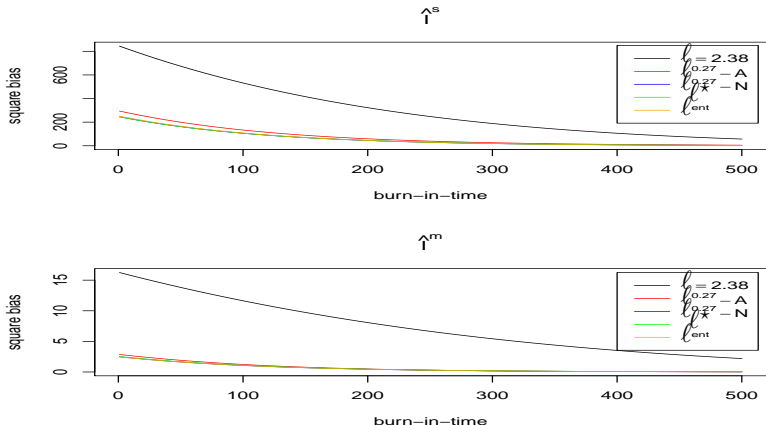


Figure: $t_0 \mapsto$ square bias of $(\hat{I}_{t_0, T+t_0}^s, \hat{I}_{t_0, T+t_0}^m)$, $(X_0^{1,n}, \dots, X_0^{n,n}) = (10, \dots, 10)$,
 $n = 100$ ($\ell^{0.27} - A \rightarrow$ adaptive scaling Metropolis algorithm and $\ell^{0.27} - N \rightarrow$ numerical approximation of $\ell^{0.27}(s, 1)$)



Conclusions:

- 1 The constant ℓ strategy is bad ;
- 2 The constant average acceptance rate strategy (using ℓ^α) leads to very close convergence curves compared to the optimal exponential rate of convergence strategy (using ℓ^*) ;
- 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 ℓ^{ent}).