The Metropolis adjusted Langevin Algorithm for log-concave probability measures in high dimensions

Andreas Eberle

June 9, 2011

1 INTRODUCTION

$$U(x) = \frac{1}{2}|x|^2 + V(x), \qquad x \in \mathbb{R}^d, \qquad V \in C^4(\mathbb{R}^d),$$

$$\mu(dx) = \frac{1}{Z} e^{-U(x)} \lambda^d(dx) = \frac{(2\pi)^{d/2}}{Z} e^{-V(x)} \gamma^d(dx),$$

$\gamma_d = N(0, I_d)$ standard normal distribution in $\mathbb{R}^d$.

AIM :

- Approximate Sampling from μ.
- Rigorous error and complexity estimates, $d \to \infty$.

RUNNING EXAMPLE: TRANSITION PATH SAMPLING

$$dY_t \;=\; dB_t - \nabla H(Y_t)\, dt\,, \qquad Y_0 = y_0 \in \mathbb{R}^n,$$

$\mu\;=$ conditional distribution on $C([0,T],\mathbb{R}^n)$ of $(Y_t)_{t\in[0,T]}$ given $Y_T = y_T$.

By Girsanov‘s Theorem:

$$\mu(dy) \;=\; Z^{-1}\, \exp(-V(y))\, \gamma(dy),$$

$\gamma\;=$ distribution of Brownian bridge from y_0 to y_T,

$$V(y) \;=\; \int_0^T \left(\frac{1}{2}\Delta H(y_t) + |\nabla H(y_t)|^2\right) dt.$$

Finite dimensional approximation via Karhunen-Loève expansion:

$$\gamma(dy) \;\to\; \gamma^d(dx)\,, \qquad V(y) \;\to\; V_d(x) \qquad \rightsquigarrow \text{ setup above}$$

MARKOV CHAIN MONTE CARLO APPROACH

- Simulate an ergodic Markov process (X_n) with stationary distribution μ.
- n large: $P \circ X_n^{-1} \approx \mu$
- Continuous time: *(over-damped) Langevin diffusion*

$$dX_t \;=\; -\frac{1}{2}X_t\,dt - \frac{1}{2}\nabla V(X_t)\,dt \,+\, dB_t$$

- Discrete time: *Metropolis-Hastings Algorithms, Gibbs Samplers*

METROPOLIS-HASTINGS ALGORITHM

(Metropolis et al 1953, Hastings 1970)

$$\mu(x) \ := \ Z^{-1} \exp(-U(x)) \qquad \text{density of } \mu \text{ w.r.t. } \lambda^d,$$

$p(x, y)$ stochastic kernel on $\mathbb{R}^d$ proposal density, > 0,

ALGORITHM

1. Choose an initial state X_0.
2. For $n := 0, 1, 2, \ldots$ do
   - Sample $Y_n \sim p(X_n, y)dy, \ U_n \sim \text{Unif}(0, 1)$ independently.
   - If $U_n < \alpha(X_n, Y_n)$ then accept the proposal and set $X_{n+1} := Y_n$; else reject the proposal and set $X_{n+1} := X_n$.

METROPOLIS-HASTINGS ACCEPTANCE PROBABILITY

$$\alpha(x,y) \;=\; \min\left(\frac{\mu(y)p(y,x)}{\mu(x)p(x,y)},1\right) \;=\; \exp\left(-G(x,y)^+\right), \quad x,y\in\mathbb{R}^d,$$

$$G(x,y)=\log\frac{\mu(x)p(x,y)}{\mu(y)p(y,x)}=U(y)-U(x)+\log\frac{p(x,y)}{p(y,x)}=V(y)-V(x)+\log\frac{\gamma^d(x)p(x,y)}{\gamma^d(y)p(y,x)}$$

- (X_n) is a time-homogeneous Markov chain with transition kernel

$$q(x,dy)=\alpha(x,y)p(x,y)dy+q(x)\delta_x(dy), \;\; q(x)=1-q(x,\mathbb{R}^d\setminus\{x\}).$$

- *Detailed Balance:*

$$\mu(dx)\,q(x,dy) \;=\; \mu(dy)\,q(y,dx).$$

PROPOSAL DISTRIBUTIONS FOR METROPOLIS-HASTINGS

$x \mapsto Y_h(x)$ proposed move, $h > 0$ step size,

$$p_h(x,dy) = P[Y_h(x) \in dy] \text{ proposal distribution,}$$
$$\alpha_h(x,y) = \exp(-G_h(x,y)^+) \text{ acceptance probability.}$$

- **Random Walk Proposals** ($\rightsquigarrow$ Random Walk Metropolis)

$$\begin{aligned} Y_h(x) &= x + \sqrt{h} \cdot Z, \qquad Z \sim \gamma^d, \\ p_h(x,dy) &= N(x, h \cdot I_d), \\ G_h(x,y) &= U(y) - U(x). \end{aligned}$$

- **Ornstein-Uhlenbeck Proposals**

$$\begin{aligned} Y_h(x) &= \left(1 - \frac{h}{2}\right) x + \sqrt{h - \frac{h^2}{4}} \cdot Z, \qquad Z \sim \gamma^d, \\ p_h(x,dy) &= N((1-h/2)x, (h-h^2/4) \cdot I_d), \qquad \text{det. balance w.r.t. } \gamma^d \\ G_h(x,y) &= V(y) - V(x). \end{aligned}$$

- **Euler Proposals** ($\rightsquigarrow$ **M**etropolis **A**djusted **L**angevin **A**lgorithm)

$$Y_h(x) \;=\; \left(1-\frac{h}{2}\right)x - \frac{h}{2}\nabla V(x) + \sqrt{h}\cdot Z, \qquad Z \sim \gamma^d.$$

(Euler step for Langevin equation $dX_t = -\frac{1}{2}X_t\,dt - \frac{1}{2}\nabla V(X_t)\,dt + dB_t$)

$$\begin{aligned} p_h(x,dy) &= N((1-\frac{h}{2})x - \frac{h}{2}\nabla V(x), h\cdot I_d), \\ G_h(x,y) &= V(y)-V(x)-(y-x)\cdot(\nabla V(y)+\nabla V(x))/2 \\ &\quad +h(|\nabla U(y)|^2 - |\nabla U(x)|^2)/4. \end{aligned}$$

REMARK. Even for $V \equiv 0$, γ^d is not a stationary distribution for p_h^{Euler}. Stationarity only holds asymptotically as $h \to 0$. This causes substantial problems in high dimensions.

- **Semi-implicit Euler Proposals** ($\rightsquigarrow$ Semi-implicit MALA)

$$
\begin{aligned}
Y_h(x) &= \left(1-\frac{h}{2}\right)x - \frac{h}{2}\nabla V(x) + \sqrt{h-\frac{h^2}{4}}\cdot Z, \qquad Z \sim \gamma^d, \\
p_h(x,dy) &= N((1-\frac{h}{2})x - \frac{h}{2}\nabla V(x), (h-\frac{h^2}{4})\cdot I_d) \qquad (= p_h^{OU} \text{ if } V \equiv 0) \\
G_h(x,y) &= V(y) - V(x) - (y-x)\cdot(\nabla V(y) + \nabla V(x))/2 \\
&\quad + \frac{h}{8-2h}\left((y+x)\cdot(\nabla V(y) - \nabla V(x)) + |\nabla V(y)|^2 - |\nabla V(x)|^2\right).
\end{aligned}
$$

REMARK. Semi-implicit discretization of Langevin equation

$$
\begin{aligned}
dX_t &= -\frac{1}{2}X_t\,dt - \frac{1}{2}\nabla V(X_t)\,dt + dB_t \\
X_{n+1} - X_n &= -\frac{\varepsilon}{2}\frac{X_{n+1}+X_n}{2} - \frac{\varepsilon}{2}\nabla V(X_n) + \sqrt{\varepsilon}Z_{n+1}, \qquad Z_i \text{ i.i.d. } \sim \gamma^d
\end{aligned}
$$

Solve for X_n and substitute $h = \varepsilon/(1+\varepsilon/4)$:

$$X_{n+1} \;=\; \left(1 - \frac{h}{2}\right) X_n - \frac{h}{2}\,\nabla V(X_n) + \sqrt{h - \frac{h^2}{4}} \cdot Z_{n+1}.$$

KNOWN RESULTS FOR METROPOLIS-HASTINGS IN HIGH DIMENSIONS

- Scaling of acceptance probabilities and mean square jumps as $d \to \infty$
- Diffusion limits as $d \to \infty$
- Ergodicity, Geometric Ergodicity
- Quantitative bounds for mixing times, rigorous complexity estimates

Optimal Scaling and diffusion limits

- *Roberts, Gelman, Gilks 1997*: Diffusion limit for RWM with product target, $h = O(d^{-1})$
- *Roberts, Rosenthal 1998*: Diffusion limit for MALA with product target, $h = O(d^{-1/3})$
- *Beskos, Roberts, Stuart, Voss 2008*: Semi-implicit MALA applied to Transition Path Sampling, Scaling $h = O(1)$
- *Beskos, Roberts, Stuart 2009*: Optimal Scaling for non-product targets
- *Mattingly, Pillai, Stuart 2010*: Diffusion limit for RWM with non-product target, $h = O(d^{-1})$
- *Pillai, Stuart, Thiéry 2011*: Diffusion limit for MALA with non-product target, $h = O(d^{-1/3})$

Geometric ergodicity for MALA

- *Roberts, Tweedie 1996*: Geometric convergence holds if ∇U is globally Lipschitz but fails in general
- *Bou Rabee, van den Eijnden 2009*: Strong accuracy for truncated MALA
- *Bou Rabee, Hairer, van den Eijnden 2010*: Convergence to equilibrium for MALA at exponential rate up to term exponentially small in time step size

BOUNDS FOR MIXING TIME, COMPLEXITY

Metropolis with ball walk proposals

- *Dyer, Frieze, Kannan 1991*: $\mu = Unif(K)$, $K \subset \mathbb{R}^d$ convex
  $\Rightarrow$ Total variation mixing time is polynomial in d and $diam(K)$
- *Applegate, Kannan 1991, ... , Lovasz, Vempala 2006*: $U : K \to \mathbb{R}$ concave, $K \subset \mathbb{R}^d$ convex
  $\Rightarrow$ Total variation mixing time is polynomial in d and $diam(K)$

Metropolis adjusted Langevin :

- No rigorous complexity estimates so far.
- Classical results for Langevin diffusions. In particular: If μ is strictly log-concave, i.e.,

$$\exists\, \kappa > 0 : \quad \partial^2 U(x) \geq \kappa \cdot I_d \qquad \forall\, x \in \mathbb{R}^d$$

then

$$d_K(\,\text{law}(X_t)\,,\,\mu\,) \;\leq\; e^{-\kappa t}\, d_K(\,\text{law}(X_0)\,,\,\mu\,).$$

If μ is strictly log-concave, i.e.,

$$\exists\, \kappa > 0 : \quad \partial^2 U(x) \geq \kappa \cdot I_d \qquad \forall\, x \in \mathbb{R}^d$$

then

$$d_K(\,\mathrm{law}(X_t)\,,\, \mu\,) \;\leq\; e^{-\kappa t}\, d_K(\,\mathrm{law}(X_0)\,,\, \mu\,).$$

- Bound is independent of dimension, sharp !
- Under additional conditions, a corresponding result holds for the Euler discretization.
- This suggests that comparable bounds might hold for MALA, or even for Ornstein-Uhlenbeck proposals.

2 Main result and strategy of proof

Semi-implicit MALA:

$$Y_h(x) \;=\; \left(1-\frac{h}{2}\right)x-\frac{h}{2}\,\nabla V(x)+\sqrt{h-\frac{h^2}{4}}\cdot Z, \qquad Z\sim\gamma^d,\; h>0,$$

Coupling of proposal distributions $p_h(x,dy)$, $x\in\mathbb{R}^d$,

$$W_h(x) \;=\; \begin{cases} Y_h(x) & \text{if } U\leq \alpha_h(x,Y_h(x)) \\ x & \text{if } U>\alpha(x,Y(x,\tilde{x})) \end{cases}, \quad U\sim Unif(0,1) \text{ independent of } Z,$$

Coupling of MALA transition kernels $q_h(x,dy)$, $x\in\mathbb{R}^d$.

We fix a radius $R \in (0, \infty)$ and a norm $\|\cdot\|_-$ on $\mathbb{R}^d$ such that $\|x\|_- \le |x|$ for any $x \in \mathbb{R}^d$, and set

$$d(x, \tilde{x}) := \min(\|x - \tilde{x}\|_-, R), \qquad B := \{x \in \mathbb{R}^d : \|x\|_- < R/2\}.$$

EXAMPLE: Transition Path Sampling

- $|x|_{\mathbb{R}^d}$ is finite dimensional projection of Cameron-Martin norm

$$|x|_{CM} = \left(\int_0^T \left| \frac{dx}{dt} \right|^2 dt \right)^{1/2}.$$

- $\|x\|_-$ is finite dimensional approximation of supremum or L^2 norm.

GOAL:

$$E\left[d(W_h(x),W_h(\tilde{x}))\right] \;\le\; \left(1-\kappa h+Ch^{3/2}\right)\, d(x,\tilde{x}) \qquad \forall\, x,\tilde{x}\in B,\; h\in(0,1)$$

with explicit constants $\kappa, C \in (0,\infty)$ that do depend on the dimension d only through the moments

$$m_k \;:=\; \int_{\mathbb{R}^d} \|x\|_-^k \,\gamma^d(dx)\;, \qquad k\in\mathbb{N}.$$

CONSEQUENCE:

- Contractivity of MALA transition kernel q_h for small h w.r.t. Kantorovich-Wasserstein distance

$$d_K(\nu,\eta) \;=\; \sup_{X\sim\nu,Y\sim\eta} E[d(X,Y)]\,, \qquad \nu,\eta\in Prob(\mathbb{R}^d).$$

$$d_K(\nu q_h,\mu) \;\le\; (1-\kappa h+Ch^{3/2})\, d_K(\nu,\mu) + R\cdot(\nu(B^c)+\mu(B^c)).$$

- Upper bound for mixing time

$$T_{mix}(\varepsilon) \;=\; \inf\{n \geq 0 \,:\, d_K(\nu q_h^n, \mu) < \varepsilon \text{ for any } \nu \in Prob(\mathbb{R}^d)\}.$$

EXAMPLE: Transition Path Sampling

Dimension-independent bounds hold under appropriate assumptions.

STRATEGY OF PROOF: Let

$$A(x) := \{U \le \alpha_h(x, Y_h(x))\} \qquad \text{(proposed move from } x \text{ is accepted)}$$

Then

$$\begin{aligned} E[d(W_h(x), W_h(\tilde{x}))] \quad \le \quad & E[d(Y_h(x), Y_h(\tilde{x}));\ A(x) \cap A(\tilde{x})] \\ & + d(x, \tilde{x}) \cdot P[A(x)^C \cap A(\tilde{x})^C] \\ & + R \cdot P[A(x) \,\triangle\, A(\tilde{x})]. \end{aligned}$$

We prove under appropriate assumptions:

1. $E[d(Y_h(x), Y_h(\tilde{x}))] \;\le\; (1 - \kappa h) \cdot d(x, \tilde{x})$,
2. $P[A(x)^C] \;=\; E[1 - \alpha_h(x, Y_h(x))] \;\le\; C_1\, h^{3/2}$,
3. $P[A(x) \,\triangle\, A(\tilde{x})] \;\le E[|\alpha_h(x, Y_h(x)) - \alpha_h(\tilde{x}, Y_h(\tilde{x}))|] \;\le\; C_2\, h^{3/2}\, \|x - \tilde{x}\|_-$,

with explicit constants $\kappa, C_1, C_2 \in (0, \infty)$.

3 Contractivity of proposal step

PROPOSITION. Suppose there exists a constant $\alpha \in (0,1)$ such that

$$\|\nabla^2 V(x) \cdot \eta\|_- \;\leq\; \alpha \|\eta\|_- \qquad \forall x \in B, \eta \in \mathbb{R}^d. \tag{1}$$

Then

$$\|Y_h(x) - Y_h(\tilde{x})\|_- \;\leq\; \left(1 - \frac{1-\alpha}{2} h\right) \|x - \tilde{x}\|_- \qquad \forall x, \tilde{x} \in B,\; h > 0.$$

Proof.

$$\begin{aligned}
\|Y_h(x) - Y_h(\tilde{x})\|_- \quad &\leq \quad \int_0^1 \|\partial_{x-\tilde{x}} Y_h(tx + (1-t)\tilde{x})\|_- \, dt \\
&= \quad \int_0^1 \|(1 - \frac{h}{2})(x - \tilde{x}) - \frac{h}{2} \nabla^2 V(tx + (1-t)\tilde{x}) \cdot (x - \tilde{x})\|_- \, dt \\
&\leq \quad (1 - \frac{h}{2})\|x - \tilde{x}\|_- + \frac{h}{2} \alpha \|x - \tilde{x}\|_- \quad \square
\end{aligned}$$

REMARK.
The assumption (1) implies strict convexity of $U(x) = \frac{1}{2}|x|^2 + V(x)$.

EXAMPLE.
For Transition Path Sampling, (A1) holds for small T with α independent of the dimension.

4 Bounds for MALA rejection probabilities

$\alpha_h(x,y) \; = \; \exp(-G_h(x,y)^+)$ Acceptance probability for semi-implicit MALA,

$$\begin{aligned} G_h(x,y) \;&=\; V(y)-V(x)-(y-x)\cdot(\nabla V(y)+\nabla V(x))/2 \\ &\quad +\frac{h}{8-2h}\left((y+x)\cdot(\nabla V(y)-\nabla V(x))+|\nabla V(y)|^2-|\nabla V(x)|^2\right). \end{aligned}$$

ASSUMPTION. There exist finite constants $C_n, p_n \in [0,\infty)$ such that

$$|(\partial^n_{\xi_1,\dots,\xi_n}V)(x)| \;\le\; C_n \max(1,\|x\|_-)^{p_n}\,\|\xi_1\|_-\cdots\|\xi_n\|_-$$

for any $x\in\mathbb{R}^d$, $\xi_1,\dots,\xi_n\in\mathbb{R}^d$, and $n=2,3,4$.

THEOREM. If the assumption above is satisfied then there exists a polynomial $\mathcal{P} : \mathbb{R}^2 \to \mathbb{R}_+$ of degree $\max(p_3 + 3, 2p_2 + 2)$ such that

$$E[1 - \alpha_h(x, Y_h(x))] \;\leq\; E[G_h(x, Y_h(x))^+] \;\leq\; \mathcal{P}(\|x\|_-, \|\nabla U(x)\|_-) \cdot h^{3/2}$$

for all $x \in \mathbb{R}^d$, $h \in (0, 2)$.

REMARK.

- The polynomial $\mathcal{P}$ is explicit. It depends only on the values C_2, C_3, p_2, p_3 and on the moments

  $$m_k \;=\; E[\|Z\|_-^k], \qquad 0 \leq k \leq \max(p_3 + 3, 2p_2 + 2)$$

  but it does not depend on the dimension d.

- For MALA with explicit Euler proposals, a corresponding estimate holds with m_k replaced by $\tilde{m}_k \;=\; E[|Z|^k]$. Note, however, that $\tilde{m}_k \to \infty$ as $d \to \infty$.

Proof.

$$\begin{aligned}
&|V(y) - V(x) - (y-x)\cdot(\nabla V(y) + \nabla V(x))/2| \\
&\quad = \quad \left| \frac{1}{2} \int_0^1 t(1-t) \partial^3_{y-x} V(x + t(y-x))\, dt \right| \\
&\quad \le \quad \frac{1}{12} \|y-x\|_-^3 \cdot \sup\{\partial^3_\eta V(z) \,:\, z \in [x,y],\ \|\eta\|_- \le 1\}
\end{aligned} \tag{2}$$

$$E\left[\|Y_h(x) - x\|_-^3\right] \;\le\; const. \cdot h^{3/2} \tag{3}$$

$$\begin{aligned}
&|(y+x)\cdot(\nabla V(y) - \nabla V(x))| \\
&\quad \le \quad \|y+x\|_- \cdot \sup_{\|\xi\|_- \le 1} |\partial_\xi V(y) - \partial_\xi V(x)| \\
&\quad \le \quad \|y+x\|_- \cdot \|y-x\|_- \cdot \sup\{\partial^2_{\xi\eta} V(z) \,:\, z \in [x,y],\ \|\xi\|_-, \|\eta\|_- \le 1\}
\end{aligned} \tag{4}$$

5 Dependence of rejection on current state

THEOREM. If the assumption above is satisfied then there exists a polynomial $\mathcal{Q} : \mathbb{R}^2 \to \mathbb{R}_+$ of degree $\max(p_4 + 2, p_3 + p_2 + 2, 3p_2 + 1)$ such that

$$E\left[\|\nabla_x G_h(x, Y_h(x))\|_+\right] \;\leq\; \mathcal{Q}(\|x\|_-, \|\nabla U(x)\|_-) \cdot h^{3/2}$$

for all $x \in \mathbb{R}^d$, $h \in (0, 2)$, where

$$\|\eta\|_+ \;:=\; \sup\{\xi \cdot \eta \,:\, \|\xi\|_- \leq 1\}.$$

CONSEQUENCE.

$$\begin{aligned}
P[\text{Accept}(x) \triangle \text{Accept}(\tilde{x})] \quad &\leq \quad E[|\alpha_h(x, Y_h(x)) - \alpha_h(\tilde{x}, Y_h(\tilde{x}))|] \\
&\leq \quad E[|G_h(x, Y_h(x)) - G_h(\tilde{x}, Y_h(\tilde{x}))|] \\
&\leq \quad \|x - \tilde{x}\|_- \cdot \sup_{z \in [x, \tilde{x}]} \|\nabla_z G_h(z, Y_h(z))\|_+ \\
&\leq \quad h^{3/2} \|x - \tilde{x}\|_- \sup_{z \in [x, \tilde{x}]} Q(\|x\|_-, \|\nabla U(x)\|_-)
\end{aligned}$$