Geometric ergodicity in Wasserstein distance of a Metropolis algorithm based on a first-order Euler exponential integrator

Alain Durmus

Joint work with Éric Moulines

Département TSI, Telecom ParisTech

Séminaire d'analyse numérique, Université de Genève, 28 Octobre 2014

1 Introduction to Markov Chain Monte Carlo methods
- Motivations
- Some Markov chain theory
- The Metropolis-Hastings algorithm
- Uniform ergodicity of the independent sampler
- Symmetric Random Walk Metropolis

2 Geometric ergodicity in Wasserstein distance and application
- Geometric ergodicity in Wasserstein distance
- Application to the EI-MALA

Outlines

1 Introduction to Markov Chain Monte Carlo methods
- Motivations
- Some Markov chain theory
- The Metropolis-Hastings algorithm
- Uniform ergodicity of the independent sampler
- Symmetric Random Walk Metropolis

2 Geometric ergodicity in Wasserstein distance and application
- Geometric ergodicity in Wasserstein distance
- Application to the EI-MALA

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Bayesian setting (I)

- Let (E, d) be a Polish space endowed with its σ-field $\mathcal{B}(E)$.
- In a Bayesian setting, a parameter $x \in E$ is embedded with a prior distribution π and the observations are given by a probabilistic model :

$$Y \sim \ell(\cdot|x)$$

The inference is then based on the posterior distribution :

$$\boxed{\pi(\mathrm{d}x|Y) = \frac{\pi(\mathrm{d}x)\ell(Y|x)}{\int \ell(Y|u)\pi(\mathrm{d}u)} .}$$

In most cases the normalizing constant is not tractable :

$$\pi(\mathrm{d}x|Y) \propto \pi(\mathrm{d}x)\ell(Y|x) .$$

Bayesian setting (II)

Bayesian decision theory relies on minimization problems involving expectations :

$$\int_E L(x,\theta)\ell(Y|x)\pi(\mathrm{d}x)$$

Generic problem : estimation of an expectation $\mathbb{E}_\pi[f]$, where

- π is known up to a multiplicative factor ;
- we do not know how to sample from π (no basic Monte Carlo estimator) ;
- π is high dimensional density (usual importance sampling and accept/reject inefficient).

Key tool : the rejection sampling

In the case $E = \mathbb{R}^d$, and π has a density with respect to the Lebesgue measure Leb^d, also denoted π.

Assume we know that $\pi(x) \leq M\nu(x)$ and that we know how to sample from ν.

1. Sample $X \sim \nu$ and $U \sim U([0,1])$.
2. If $U \leq \frac{\pi(X)}{M\nu(X)}$, accept X.
3. Else go to 1.

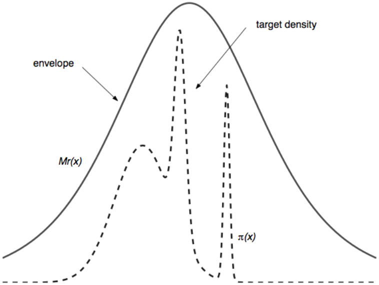


FIGURE: *

Illustration of the Accept-Reject method [Cappé, Moulines, Ryden 2005].

Inefficiency of the rejection sampling

- Hard to find a probability ν such that $\pi \leq M\nu$ (especially for high dimensional settings).
- On one hand M^{-1} is the rate of acceptance so that M has to be as close to 1 as possible.
  But on the other hand, in practice M is exponentially large in the dimension.

Alternative : MCMC method !

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Some Markov chain theory (I)

Definition

Let $P : E \times \mathcal{B}(E) \to \mathbb{R}_+$. P is a Markov kernel if

- for all $x \in E$, $A \mapsto P(x, A)$ is probability measure on E,
- for all $A \in \mathcal{B}(E)$, $x \mapsto P(x, A)$ is measurable from E to $\mathbb{R}$.

Some Markov chain theory (II)

Some simple properties :

- If P_1 and P_2 is two Markov kernel, we can define a new Markov kernel, denoted P_1P_2, by for all $x \in E$ and $A \in \mathcal{B}(E)$:

$$P_1P_2(x, A) = \int_E P_1(x, \mathrm{d}z)P_2(z, A) \,.$$

- If P is a Markov kernel and ν a probability measure on E, we can define a probability measure, denoted νP, by for all $A \in \mathcal{B}(E)$:

$$\nu P(A) = \int_E \nu(\mathrm{d}z)P(z, A) \,.$$

- Let P be a Markov kernel on E. For $f : E \to \mathbb{R}_+$ measurable, we can define a measurable function $Pf : E \to \bar{\mathbb{R}}_+$ by

$$Pf(x) = \int_E P(x, \mathrm{d}z)f(z) \,.$$

Some Markov chain theory (III)

Invariant probability measure : π is said to be an invariant probability measure for the Markov kernel P if $\pi P = P$.

Theorem (Meyn and Tweedie, 2003, Ergodic theorem)

With some conditions on P, we have for any $f \in \mathrm{L}^1(\pi)$,

$$\hat{\pi}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[\pi\text{-a.s.}]{} \int f(x)\pi(\mathrm{d}x)\,.$$

Conditions of the Theorem

Definition

- Irreducibility : there exists a measure ν such that, for all x and all A such that $\nu(A) > 0$, there exists $n \in \mathbb{N}^*$ s.t. $P^n(x, A) > 0$.

- Harris recurrence : P is Harris recurrent : for all $A \in \mathcal{B}(E)$ satisfying $\pi(A) > 0$, for all x in A

$$\mathbb{P}\left[\sum_{k=1}^{+\infty} \mathbb{1}_A(X_k) = +\infty \,\middle|\, X_0 = x\right] = 1 \,.$$

MCMC : rationale (I)

The Theorem above gives the following idea to approximate $\mathbb{E}_\pi[f]$:

- Find a kernel P with invariant measure π, from which we can efficiently sample.
- Sample a Markov chain $X_1, \ldots, X_n$ with kernel P and compute

$$\hat{\pi}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

  to approximate $\mathbb{E}_\pi[f] = \int f(x) \pi(\mathrm{d}x)$.

$\Rightarrow$ How to find a Markov kernel P with invariant measure π ?

MCMC : rationale (II)

Simple condition to check that π is invariant for P : reversibility.

Definition

P is reversible with respect to π if for all $A_1, A_2 \in \mathcal{B}(E)$:

$$\int_{A_1} \int_{A_2} \pi(\mathrm{d}z_1) P(z_1, \mathrm{d}z_2) = \int_{A_1} \int_{A_2} \pi(\mathrm{d}z_2) P(z_2, \mathrm{d}z_1) \, .$$

- Note the variables z_1 and z_2 are switched.
- For $A_1 = E$ and $A_2 = A$, we get $\pi(A) = \pi P(A)$.

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

The Metropolis-Hastings algorithm (I)

The Metropolis-Hastings algorithm gives a generic method to build Markov kernels P reversible w.r.t. π in the case where :

- $E = \mathbb{R}^d$.
- Objective target probability π has a density w.r.t. Leb^d, also denoted π.

Using of a transition density $q(x, y)$ w.r.t. Leb^d :

- $(x, y) \mapsto q(x, y)$ is measurable,
- For all x, $y \mapsto q(x, y)$ is a density of a probability measure also denoted $q(x, \cdot)$.

The Metropolis-Hastings algorithm (II)

Given X_k,

1. Generate $Y_{k+1} \sim q(\cdot, X_k)$.
2. Set

$$X_{k+1} = \begin{cases} Y_{k+1} & \text{with probability } \alpha(X_k, Y_{k+1})\,, \\ X_k & \text{with probability } 1 - \alpha(X_k, Y_{k+1})\,. \end{cases}$$

   where

$$\boxed{\alpha(x, y) = 1 \wedge \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}}\,.$$

- With this choice of α the algorithm produces a Markov kernel P_{MH} reversible w.r.t. π.
- “No restriction” on π and q.

MH : properties

Simple condition to apply the Ergodic theorem :

- q and π are continuous.
- For all x, y such that $\pi(y) > 0$,

$$q(x, y) > 0\,.$$

Consequence :

For any $f \in \mathrm{L}^1(\pi)$,

$$\hat{\pi}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \underset{a.s.}{\longrightarrow} \int f(x)\pi(x)\mathrm{d}x\,.$$

Question : can we have a rate of convergence for some f ?

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Total variation distance

Definition

For μ, ν two probabilities measure on E, the total variation distance between μ and ν is given by

$$W_{d_0}(\mu,\nu) = \inf_{A \in \mathcal{B}(E)} \left| \mu(A) - \nu(A) \right| ,$$
$$= \sup_{|f| \leq 1} \left| \mathbb{E}_\mu[f] - \mathbb{E}_\nu[f] \right| .$$

- Convergence in total variation distance implies the weak convergence.
- Convergence rates in total variation distance imply convergence rates for $\mathbb{E}_{\mu_n}[f]$.

Uniform ergodicity

Definition

Let P be a Markov kernel on E, with invariant measure π. P is uniformly geometrically ergodic is there exists $C < +\infty$, and $\rho \in (0,1)$ such that for all $x \in E$:

$$W_{d_0}(P^n(x,\cdot),\pi) \leq C\rho^n \,.$$

Theorem (Meyn and Tweedie, 2003)

When P satisfy a technical condition (aperiodicity), P is uniformly geometrically ergodic if and only if there exist $\delta, \epsilon \in (0,1)$, $n \in \mathbb{N}^*$ and a probability measure μ such that

$$\forall A \in \mathcal{B}(E),\ \mu(A) > \delta \Rightarrow \inf_{x \in A} P^n(x,A) > \epsilon \,.$$

Uniform ergodicity : the independent case [Hastings 1970] (II)

Theorem ([Roberts, Tweedie 1996], [Mengersen, Tweedie 1996])

If there exists M such that $\pi(z) \leq Mg(z)$ then for all $x \in \mathbb{R}^d$

$$W_{d_0}(P^n_{\mathrm{IMH}}(x,\cdot),\pi) \leq \left(1 - \frac{1}{M}\right)^n .$$

1. Expected acceptance probability still is $1/M$.
2. But no need to know M to run the algorithm !
3. If the majoration condition does not hold, no uniform ergodicity.

Cauchy vs Normal (I) [Cappé, Moulines, Ryden 2005]

- Target distribution : $\pi(x) \propto (1 + x^2)^{-1}$.
- Proposal distribution : $g(y) \sim \mathcal{N}(0, 1)$.

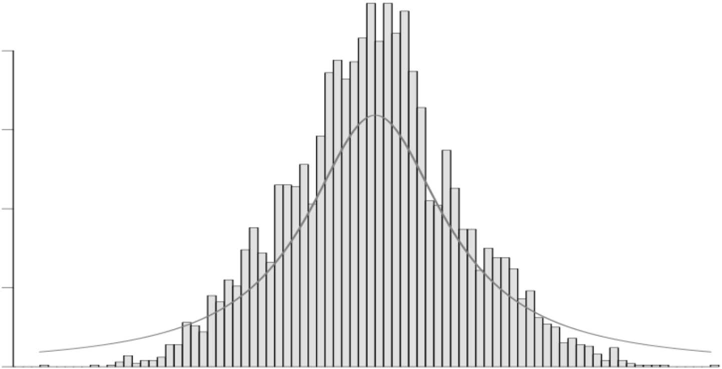


FIGURE: *

Histogram of IMH with 5000 samples.

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

symmetric Random walk Metropolis-Hastings (I)

- The idea in the RWM is to propose local moves around the current states and not moves independent of the position.

- The proposal mechanism is given by

$$Y_{k+1} = X_k + Z_{k+1} ,$$

  where Z_{k+1} is independent of X_k and is distributed according to a probability measure with a symmetric density $\tilde{q}$.

- The proposal distribution is of the form $q(x, y) = \tilde{q}(y - x)$.

symmetric Random walk Metropolis-Hastings (II)

1. Generate Z_{k+1} from $\tilde{q}$ and set $Y_{k+1} = X_k + Z_{k+1}$.
2. Set

$$X_{k+1} = \begin{cases} Y_{k+1} & \text{with probability } \alpha(X_k, Y_{k+1})\,, \\ X_k & \text{with probability } 1 - \alpha(X_k, Y_{k+1})\,. \end{cases}$$

   where

$$\boxed{\alpha(x,y) = 1 \wedge \frac{\pi(y)}{\pi(x)}}\,.$$

Cauchy vs Normal (II) [Cappé, Moulines, Ryden 2005]

- Target distribution : $\pi(x) \propto (1 + x^2)^{-1}$.
- Proposal distribution : $\mathcal{N}(0, 1)$.

$$\boxed{\alpha(x, y) = 1 \wedge \frac{1 + x^2}{1 + y^2}}$$

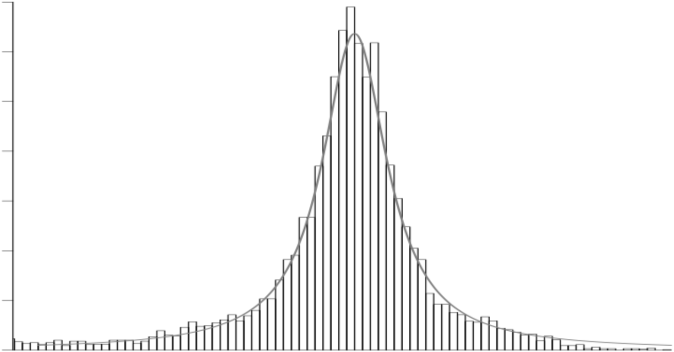


FIGURE: *

Histogram of RMW with 10000 samples

Another kind of convergence : the geometric ergodicity

1. Using random walk moves prevents from being uniformly geometrically ergodic [Robert, Casella 2004]. But still, we can have geometric ergodicity.
2. The condition $W_{d_0}(P^n(x,\cdot),\pi)$ was controlled uniformly in x is relaxed.

Definition

Let P be a Markov kernel with invariant probability measure π. P is geometrically ergodic if there exists $C < +\infty$, $\rho \in (0,1)$ and a measurable function $V : E \to [1,+\infty)$ such that :

$$W_{d_0}(P^n(x,\cdot),\pi) \leq C\rho^n V(x), \qquad \forall x \in E\,.$$

Conditions to get geometric ergodicity (I)

Definition

A set $\mathcal{C} \in \mathcal{B}(E)$ is said to be m-small for P if there exists $\epsilon > 0$ and a probability measure μ such that :

$$\forall A \in \mathcal{B}(E), \quad \inf_{x \in \mathcal{C}} P(x, A) \geq \epsilon \mu(A) .$$

Theorem (Meyn and Tweedie 2003)

Let P an irreducible Markov kernel satisfying a technical condition (aperiodicity).
P is geometrically ergodic if and only if there exists $b < +\infty$, $\lambda \in (0, 1)$ and a measurable function $V : E \to [1, +\infty)$ such that for all $x \in E$

$$PV(x) \leq \lambda V(x) + b \mathbb{1}_{\mathcal{C}}(x) ,$$

where $\mathcal{C}$ is a m-small set for P.

Geometric ergodicity of the RMH on $\mathbb{R}$

Theorem (Mengersen and Tweedie, 1994)

Assume that

- π *is continuous and symmetric on* $\mathbb{R}$, *and log-concave in the tail, ie there exists* $M, a \in> 0$ *such that for all* $|y| \geq |x| \geq M$,

$$\log(\pi(x)) - \log(\pi(y)) \geq a\,|x - y| \ ,$$

- *the transition density of the noise* $\tilde{q}$ *is continuous and positive on* $\mathbb{R}$.

Then, P_{RWM} *is geometrically ergodic.*

The proof follows from the previous theorem applied with $V(x) = \mathrm{e}^{s|x|}$.

Issues of the geometrical ergodicity in total variation norm

1. In an infinite dimensional setting, Markov chain will typically be not irreducible. So we cannot apply the previous result to this kind of kernel.

2. The contraction coefficient ρ in the previous theorem is smaller than the constant ϵ which appears in the definition of the a small set. Most of the time, ϵ is exponentially small in the dimension.
   So in high dimensional settings, we can observe poor mixing even if the chain is geometrically ergodic.

In the following, we try to give some solutions to the second point. We consider another kind of convergence, which suggests ways to construct sampler with good mixing rate, even if the dimension is large.

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Wasserstein distance (I)

Recall (E, d) is a Polish space.
We assume in the following that d is bounded by 1.

Definition

1. Let μ and ν two probability measures on E. λ is a coupling of μ and ν if λ is a probability on $E \times E$, such that for all $A \in \mathcal{B}(E)$,

   $$\lambda(A \times E) = \mu(A) \text{ and } \lambda(E \times A) = \nu(A)\,.$$

   The set of the couplings of μ and ν will be denoted $\mathrm{C}(\mu, \nu)$.

2. The Wasserstein metric associated to d, between two probability measures μ, ν is defined by :

   $$W_d(\mu, \nu) = \inf_{\lambda \in \mathrm{C}(\mu,\nu)} \int_{E \times E} d(x, y)\mathrm{d}\lambda(x, y)\,,$$

Wasserstein distance (II)

- We get back the the total variation distance when $d(x,y) = d_0(x,y) = \mathbb{1}_{x \neq y}$
- The convergence in total variation implies the convergence in Wasserstein distance but the converse is false.
- The convergence in W_d is equivalent to the weak convergence ; (see e.g. [Villani, 2009] for details).

So, we generalize the use of the total variation distance.

Coupling set

In the following, we adapt the notion of small set to this setting.

Definition

Let P be a Markov kernel on E. Let $\blacksquare \in \mathcal{B}(E \times E)$, and $\epsilon \in (0,1)$.
We say that $\blacksquare$ is a (ϵ, d)-coupling set for the Markov kernel P if there exists a kernel $\mathbf{K}$ on $(E \times E, \mathcal{B}(E \times E))$ satisfying the following conditions

- for all $x, y \in E$, $\mathbf{K}((x,y), \cdot)$ is a coupling of $(P(x,\cdot), P(y,\cdot))$.
- for all $x, y \in E$,
$$\mathbf{K}d(x,y) \leq d(x,y)\,.$$
- for all $(x,y) \in \blacksquare$,
$$\mathbf{K}d(x,y) \leq (1-\epsilon)d(x,y)\,.$$

If $\mathcal{C}$ is a 1-small set, $\mathcal{C} \times \mathcal{C}$ is an (ϵ, d_0)-coupling set.

Quantitative bound for geometric ergodicity in Wasserstein distance

We have the following theorem which generalizes and precises the constants of the theorem about geometric ergodicity in total variation distance.

Theorem

Let P be Markov kernel on E and assume

- *There exists a measurable function* $V : E \to [1, +\infty)$, $\lambda \in [0, 1)$ *and* $b < +\infty$ *such that for all* $x \in E$,

$$PV(x) \leq \lambda V(x) + b \,.$$

- *For some* $\delta > 0$, *the subset*

$$\blacksquare \overset{\text{def}}{=} \{(x, y) \in E \times E, V(x) + V(y) \leq (2b + \delta)/(1 - \lambda)\} \,,$$

  is an (ϵ, d)*-coupling.*

Then P admits a unique probability measure π *and for all* $x \in E$

$$W_d(P^n(x, \cdot), \pi) \leq C\rho^n V(x)$$

where $C < +\infty$ *and* $\rho \in (0, 1)$, *which can be explicitly calculated in function of* ϵ, λ, b *and* δ.

Outlines

1. Introduction to Markov Chain Monte Carlo methods
   - Motivations
   - Some Markov chain theory
   - The Metropolis-Hastings algorithm
   - Uniform ergodicity of the independent sampler
   - Symmetric Random Walk Metropolis

2. Geometric ergodicity in Wasserstein distance and application
   - Geometric ergodicity in Wasserstein distance
   - Application to the EI-MALA

Presentation of the EI-MALA (I)

EI-MALA is Metropolis Hasting algorithm, based on the following.

- Let π be the target density and $\pi_U(x)\mathrm{d}x \propto \mathrm{e}^{-U(x)}\,\mathrm{d}x$ be an auxiliary probability measures on $\mathbb{R}^d$.
- Typically, $-\log(\pi_U)$ will be a convex minorant of $-\log(\pi)$. So, we assume that U is given by

$$U(x) = (1/2)x^T Qx + \blacksquare(x) \text{ with } Q \succ 0\,.$$

- Consider the over-damped Langevin SDE associated with π_U :

$$\mathrm{d}Y_t = -Y_t\mathrm{d}t - Q^{-1}\nabla\blacksquare(Y_t)\mathrm{d}t + \sqrt{2}Q^{-1/2}\mathrm{d}B_t\,.$$

Presentation of the EI-MALA (II)

- A stochastic Euler exponential integrator yields to the following discretization for $h \in (0,2)$:

$$\mathscr{O}_h(x, Z_1) = x - (h/2)Q^{-1}\nabla U(x) + \sqrt{h - h^2/4}\; Q^{-1/2} Z_1 \,,$$

  where $Z_1 \sim \mathcal{N}(0, \mathrm{I}_d)$.

- It yields to a proposal density q_h which can be used in a Metropolis-Hastings algorithm.

- Given $h \in (0,2)$ and X_n,
  - .1 Sample $Z_{k+1} \sim \mathcal{N}(0, \mathrm{I}_d)$ and set $Y_{k+1} = \mathscr{O}(X_k, h)$.
  - .2 Set

$$X_{k+1} = \begin{cases} Y_{k+1} & \text{with probability } \alpha_h(X_k, Y_{k+1}) \,, \\ X_k & \text{with probability } 1 - \alpha_h(X_k, Y_{k+1}) \,. \end{cases}$$

    where

$$\boxed{\alpha_h(x, y) = 1 \wedge \frac{\pi(y)}{\pi(x)} \frac{q_h(y, x)}{q_h(x, y)}} \,.$$

Geometric convergence of the EI-MALA (I)

- Denote by $\|\cdot\|_Q$ the norm on $\mathbb{R}^d$ associated with the positive definite matrix Q.

- Using the previous Theorem, we establish the geometric convergence of the EI-MALA algorithm when π is given by $\pi(x) \propto \mathrm{e}^{-(1/2)x^T Qx - \blacksquare(x) - \blacksquare(x)}$, where $\blacksquare$ satisfies with $\blacksquare$ the following assumptions.

M1

1. The function $\blacksquare$ belongs to $C^1(\mathbb{R}^d)$, is convex and there exists $C_\blacksquare$ such that for all $x, y \in \mathbb{R}^d$, $\left\|Q^{-1}(\nabla\blacksquare(x) - \nabla\blacksquare(y))\right\|_Q \leq C_\blacksquare \|x - y\|_Q$.
2. The function $\blacksquare$ belongs to $C^1(\mathbb{R}^d)$ and there exists $C_\blacksquare$ such that for all $x, y \in \mathbb{R}^d$, $\left\|Q^{-1}(\nabla\blacksquare(x) - \nabla\blacksquare(y))\right\|_Q \leq C_\blacksquare \|x - y\|_Q$.

Geometric convergence of the EI-MALA (I)

We make the following second assumption, which in essence imposes that the acceptance probability is bounded from below by a positive constant.

M2 There exists $h_\ell \in (0,2)$ such that for all $h \in (0, h_\ell]$ there exists three positive real numbers a_h, R_h and r_h such that for all $x \in \mathbb{R}^d$, $\|x\|_Q \geq R_h$,

$$\inf\left\{\alpha_h(x,z), z \in \mathrm{B}_Q\left(\mathscr{O}_h(x,0), r_h\right)\right\} > a_h \,.$$

Theorem

Assume **M1**, **M2** *and let* $h \in \left(0, h_\ell \wedge (4/(C_\blacksquare^2+1))\right)$. *Then, there exist a distance* ℓ *on* $\mathbb{R}^d$, $\rho_{\text{EI-MALA}} \in (0,1)$ *such that for all* $x \in \mathbb{R}^d$ *and* $n \in \mathbb{N}^*$

$$W_\ell(P^n(x,\cdot),\pi) \leq C\rho^n \left\{\mathscr{V}(x) + \mathscr{V}(y)\right\} \;,$$

with $\mathscr{V}(x) = 1 \vee \|x\|_Q$.

Calculation of the coefficient contraction in a simple case

- To illustrate our bounds, assume that ■ $\equiv 0$ and that ■ is bounded on $\mathbb{R}^d$ and gradient Lipschitz.
- It is easily checked that **M1** and **M2** are satisfied.
- We can compute the dependence of the contraction coefficient in function of the dimension, and see that this dependence is polynomial.

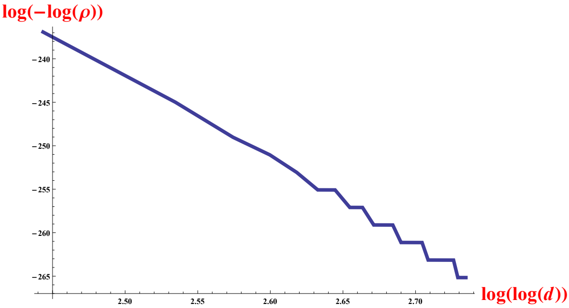


FIGURE: Evolution of the rate of convergence $\rho_{\text{EI-MALA}}$ in function of the dimension d.

Numerical illustrations (I)

We have considered an ill-conditioned Bayesian linear inverse problem.

- It is assumed that the observation $y \in \mathbb{R}^p$ is given by

$$y = Ax + G$$

  with $G \sim \mathcal{N}(0, \mathrm{I}_p)$, $A \in \mathbb{R}^{p \times d}$, and we want to learn x.
- In this problem, the dimension d can be very large and $p \ll d$.
- The prior distribution π_X of x is given to be a small pertubation of a exponential power distribution (see [Box and Tiao ,1992]) :

$$\pi_X(x) \propto \exp\left(-\lambda_1(x^T x + \delta)^\beta - (\lambda_2/2)(x^T x)\right) ,$$

  with $\beta \in (1/2, 1)$, $\lambda_1, \lambda_2, \delta \in \mathbb{R}_+^*$.

Numerical illustrations (II)

- In this setting, the posterior distribution π is proportional to $\exp(-U)$ on $\mathbb{R}^d$, where $\blacksquare = 0$ and the potential U is on the form

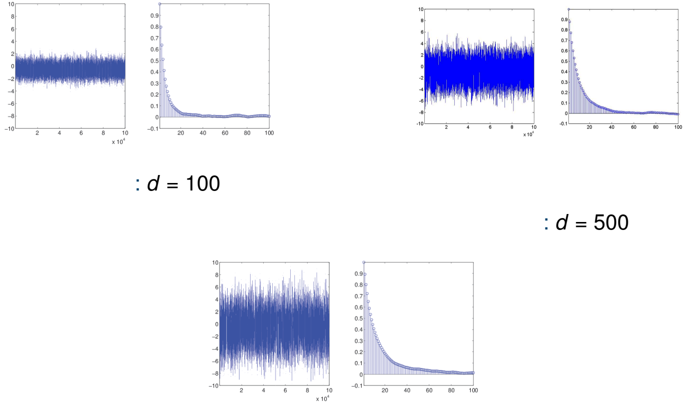
$$U(x) = (1/2)x^T Qx + \blacksquare(x) \text{ with } Q \succ 0\,,$$

where

$$Q = A^T A/2 + \lambda_2 \, \mathrm{I}_d \;\text{ and }\; \blacksquare(x) = \lambda_1 (x^T x + \delta)^\beta - \langle y, Ax\rangle \;.$$

- We can prove that π satisfies **M1** and **M2**.

Numerical illustrations (II)



: d = 100

: d = 500

: d = 1000

FIGURE: Trace plot and auto-correlation in function of the dimension on 10000 iterations with a 10000 burn-in iterations .

The End

Thank you for your attention !