

# Introduction aux méthodes de Monte Carlo par dynamique Hamiltonienne

Shmuel RAKOTONIRINA-RICQUEBOURG, Amaury DURAND

14 décembre 2017

# Plan

- 1 Introduction
- 2 Dynamique hamiltonienne
- 3 Hamiltonian Monte-Carlo
- 4 Simulations
- 5 Conclusion

## 1 Introduction

- Algorithmes MCMC
- Algorithme de Metropolis (Random Walk Metropolis)

## 2 Dynamique hamiltonienne

## 3 Hamiltonian Monte-Carlo

## 4 Simulations

## 5 Conclusion

# Principe des MCMC

- Objectif : pour  $\pi$  à densité  $h_\pi$  simuler  $\pi$  ou approcher  $\pi f$
- Idée : trouver une chaîne de Markov  $X$  admettant  $\pi$  comme loi invariante et convergeant vers  $\pi$

## Théorème (Théorème ergodique)

Soit  $(X_k)_{k \in \mathbb{N}}$  une chaîne de Markov de noyau  $P$  sur  $(\mathbb{X}, \mathcal{X})$  admettant une unique loi invariante  $\pi$ . Alors pour tout  $f \in \mathbb{F}_+(\mathbb{X}, \mathcal{X}) \cup \mathbb{F}_b(\mathbb{X}, \mathcal{X})$  et pour  $\pi$ -presque tout  $x \in \mathbb{X}$ ,

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_x - p.s.} \pi f$$

# Algorithme Random-walk Metropolis

---

## Algorithme 1.1 : Random Walk Metropolis

---

**Données :**  $h_\pi$  proportionnel à la densité cible,  $Q$  loi simulable

$X_0 \leftarrow x \in \mathbb{X}$  arbitraire;

$(U_k)_{k \in \mathbb{N}} \stackrel{\text{iid}}{\sim} Q$  ;

**répéter**

$Y_{k+1} \leftarrow X_k + U_{k+1}$  ; // Proposer un mouvement

$\alpha_{k+1} \leftarrow \alpha(X_k, Y_{k+1})$  où  $\alpha(x, y) = 1 \wedge \frac{h_\pi(y)}{h_\pi(x)}$ ;

$X_{k+1} \leftarrow \begin{cases} Y_{k+1} & \text{avec probabilité } \alpha_{k+1} \\ X_k & \text{avec probabilité } 1 - \alpha_{k+1} \end{cases}$  ; // Accepter ou rejeter

**jusqu'à une condition d'arrêt;**

**retourner**  $(X_k)_k$

---

## 1 Introduction

## 2 Dynamique hamiltonienne

- Définition
- Propriétés
- Discrétisation

## 3 Hamiltonian Monte-Carlo

## 4 Simulations

## 5 Conclusion

# Dynamique hamiltonienne

## Définition

(Dynamique hamiltonienne) Soit  $H : \begin{matrix} \mathbb{R}^d \times \mathbb{R}^d & \rightarrow & \mathbb{R} \\ (x, p) & \mapsto & H(x, p) \end{matrix}$ . Deux fonctions de position  $x : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  et de quantité de mouvement  $p : \mathbb{R}_+ \rightarrow \mathbb{R}^d$  sont dites solutions du hamiltonien  $H$  (ou suivant la dynamique hamiltonienne de  $H$ ) si

$$\begin{aligned} x'(t) &= \frac{\partial H}{\partial p}(x(t), p(t)) \\ p'(t) &= -\frac{\partial H}{\partial x}(x(t), p(t)) \end{aligned}$$

Pour  $T > 0$ ,  $H$  peut être associée à la densité (sur  $\mathbb{R}^{2d}$ )

$$h(z) \propto \exp\left(-\frac{H(z)}{T}\right)$$

# Hamiltonien pour l'algorithme HMC

Pour avoir  $X \sim \pi$ , on cherche à simuler  $(X, P) \sim \tilde{\pi} = \pi \otimes \nu$  pour une loi  $\nu$  choisie. Dans toute la suite, on fera les hypothèses suivantes :

## Hypothèses (H)

- $\pi$  et  $\nu$  sont à densité  $h_\pi$  et  $h_\nu$  strictement positives sur  $\mathbb{R}^d$
- $\exists k \geq 1$ ,  $\ln(h_\pi)$  et  $\ln(h_\nu)$  sont de classe  $\mathcal{C}^k$  sur  $\mathbb{R}^d$
- $h_\nu$  est paire
- $H : (x, p) \mapsto U(x) + K(p)$  avec  $U = -T \ln(h_\pi)$ ,  $K = -T \ln(h_\nu)$

Ainsi, la densité  $h \propto e^{H/T}$  est la densité jointe  $h = h_\pi \otimes h_\nu = h_{\tilde{\pi}}$ .

# Flot de l'équation différentielle

En notant  $z = (x, p)$ , les équations  $x' = \frac{\partial H}{\partial p}$ ,  $p' = -\frac{\partial H}{\partial x}$  se réécrivent  $z' = F(z)$  où  $F = J\nabla H$  et  $J = \begin{bmatrix} 0_d & I_d \\ -I_d & 0_d \end{bmatrix}$ .

## Définition (flot hamiltonien)

Pour  $t \in \mathbb{R}$ , on définit le flot  $\phi_t$  de sorte que pour tout  $z_0 \in \mathbb{R}^{2d}$ ,  $t \mapsto \phi_t(z_0)$  soit l'unique solution du hamiltonien avec condition initiale  $z(0) = z_0$ .

# Propriétés des solutions

## Proposition (conservation du hamiltonien)

*Le hamiltonien est conservé le long des trajectoires : si  $z' = F(z)$ , alors  $H \circ z$  est constant.*

## Proposition (conservation du volume)

*La solution du hamiltonien conserve le volume :  $\det\left(\frac{d\phi_t}{dz}(z)\right) = 1$ .*

# Réversibilité du flot

## Lemme (réversibilité du temps)

Soit  $z = (x, p)$  une solution du hamiltonien  $H$ . On définit  $\bar{z} = (\bar{x}, \bar{p})$  par  $\bar{z}(t) = (x(-t), -p(-t))$ . Alors

- ①  $\bar{z}$  est solution du hamiltonien  $H$  (avec d'autres conditions initiales).
- ②  $\forall t \in \mathbb{R}_+, \phi_t(\bar{z}(-t)) = \bar{z}(0)$

## Proposition (réversibilité du flot)

$\phi_t$  est un  $\mathcal{C}^k$ -difféomorphisme d'inverse

$\phi_t^{-1} = s \circ \phi_t \circ s : (x, p) \mapsto (\phi_t^{(1)}(x, -p), -\phi_t^{(2)}(x, -p))$  avec  
 $s : (x, p) \mapsto (x, -p)$ .

# Preuve de la réversibilité du temps

Par définition,  $\bar{z}(t) = s \circ z(-t)$  et par hypothèse de parité de  $h_\nu$ ,  
 $H = H \circ s$ .

Notons  $S \doteq \frac{ds}{dz}(z) = \begin{bmatrix} I_d & 0_d \\ 0_d & -I_d \end{bmatrix}$  (et ce pour tout  $z$ ). Ainsi,  
 $\nabla H = S \nabla H \circ s$ . On remarque que  $SJ = -JS$  et  $s^{-1} = s$ .

$$\bar{z}'(t) = -Sz'(-t) = -SJ\nabla H(z(-t)) = JS\nabla H(z(-t))$$

donc

$$\bar{z}'(t) = J\nabla H(s^{-1}(z(-t))) = J\nabla H(\bar{z}(t))$$



# Preuve de la réversibilité du flot

Il suffit de prouver la formule de l'inverse. On pose  $\bar{\phi}_t = s \circ \phi_t \circ s$ . On fixe  $z_0 = (x_0, p_0)$  et on note  $z$  la solution du hamiltonien avec  $z(0) = z_0$ .

- D'une part,

$$\bar{\phi}_t(\phi_t(x_0, p_0)) = \bar{\phi}_t(z(t)) = (\phi_t^{(1)}(\bar{z}(-t)), -\phi_t^{(2)}(\bar{z}(-t))) = (x_0, p_0)$$

car  $\bar{z}(0) = (x_0, -p_0)$ .

- D'autre part,

$$\bar{\phi}_t(x_0, -p_0) = (\phi_t^{(1)}(x_0, p_0), -\phi_t^{(2)}(x_0, p_0)) = (x(t), -p(t)) = \bar{z}(-t)$$

donc

$$\phi_t(\bar{\phi}_t(x_0, -p_0)) = \phi_t(\bar{z}(-t)) = \bar{z}(0) = (x_0, -p_0).$$



# Algorithme du leapfrog

Variante de la méthode d'Euler : on discrétise  
 $x' = \nabla K(p), p' = -\nabla U(x)$  en

- ➊  $p_{t+\epsilon/2} = p_t - \frac{\epsilon}{2} \nabla U(x_t)$  (demi-pas en  $p$ ).
- ➋  $x_{t+\epsilon} = p_t + \epsilon \nabla K(p_{t+\epsilon/2})$  (pas en  $x$ ).
- ➌  $p_{t+\epsilon} = p_{t+\epsilon/2} - \frac{\epsilon}{2} \nabla U(x_{t+\epsilon})$  (demi-pas en  $p$ ).

---

## Algorithme 2.1 : Discrétisation de l'évolution par saute-mouton (*leapfrog*)

---

**Données** : pas  $\epsilon$ , nombre de pas  $L$ , état initial  $(x_0, p_0)$

**pour**  $k \in \llbracket 0, L - 1 \rrbracket$  **faire** // Saute-mouton

$x_{k+1} \leftarrow x_k + \epsilon \nabla K(p_k - \frac{\epsilon}{2} \nabla U(x_k));$   
 $p_{k+1} \leftarrow p_k - \frac{\epsilon}{2} \nabla U(x_k) - \frac{\epsilon}{2} \nabla U(x_{k+1});$

**retourner**  $(x_L, p_L)$

---

# Flot approché

---

## Algorithme 2.2 : Discrétisation de l'évolution par saute-mouton (*leapfrog*)

---

**Données :** pas  $\epsilon$ , nombre de pas  $L$ , état initial  $(x_0, p_0)$

**pour**  $k \in \llbracket 0, L - 1 \rrbracket$  **faire** // Saute-mouton

$$\left[ \begin{array}{l} x_{k+1} \leftarrow x_k + \epsilon \nabla K \left( p_k - \frac{\epsilon}{2} \nabla U(x_k) \right); \\ p_{k+1} \leftarrow p_k - \frac{\epsilon}{2} \nabla U(x_k) - \frac{\epsilon}{2} \nabla U(x_{k+1}); \end{array} \right.$$

**retourner**  $(x_L, p_L)$

---

### Définition (flot approché)

On fixe  $\epsilon > 0$ . Pour  $L \in \mathbb{N}^*$ , on définit le flot approché du hamiltonien par  $L$  itérations de l'algorithme leapfrog par  $\hat{\phi}_L = \hat{\phi}^L$  où  $\hat{\phi}$  est défini par

$$\hat{\phi}^{(1)} : (x, p) \mapsto x + \epsilon \nabla K \left( p - \frac{\epsilon}{2} \nabla U(x) \right)$$

$$\hat{\phi}^{(2)} : (x, p) \mapsto p - \frac{\epsilon}{2} \nabla U(x) - \frac{\epsilon}{2} \nabla U(\hat{\phi}^{(1)}(x, p))$$

# Propriétés du flot approché

## Proposition (conservation du volume)

*On suppose  $k \geq 2$  ( $U$  et  $K$  de classe  $\mathcal{C}^2$ ). La solution approchée du hamiltonien par le leapfrog conserve le volume :  $\det\left(\frac{d\hat{\phi}}{dz}(z)\right) = 1$ .*

## Proposition (réversibilité du flot approché)

$\hat{\phi}$  est inversible d'inverse

$$\hat{\phi}^{-1} = s \circ \hat{\phi} \circ s : (x, p) \mapsto (\hat{\phi}^{(1)}(x, -p), -\hat{\phi}^{(2)}(x, -p)).$$

## 1 Introduction

## 2 Dynamique hamiltonienne

## 3 Hamiltonian Monte-Carlo

- Cas idéal
- Cas réel

## 4 Simulations

## 5 Conclusion

# Algorithme HMC idéal

---

## Algorithme 3.1 : Hamiltonian Monte-Carlo, cas idéal

---

**Données :**  $h_\pi$  proportionnel à la densité cible,  $t$  une durée sur laquelle suivre la dynamique

$X_0 \leftarrow x \in \mathbb{X}$  arbitraire;

**répéter**

$\tilde{P}_k \sim \nu$  et  $\tilde{P}_k \perp\!\!\!\perp (Z_0, \dots, Z_k)$ ; // Tirer la quantité de mouvement

$\tilde{Z}_k \leftarrow (X_k, \tilde{P}_k)$ ;

$Z_{k+1} = (X_{k+1}, P_{k+1}) \leftarrow \phi_t(\tilde{Z}_k)$ ; // Suivre la dynamique

**jusqu'à une condition d'arrêt;**

**retourner**  $(X_k)_k$

---

# Invariance de $\tilde{\pi}$

## Proposition

Pour  $t \in \mathbb{R}_+$ , le processus  $(Z_k)_{k \in \mathbb{N}}$  défini par l'algorithme 3.1 est une chaîne de Markov homogène de noyau  $QR_t$  sur  $\mathbb{R}^{2d}$  avec

$$Q((x, p), \cdot) = \delta_x \otimes \nu \text{ et } R_t((x, p), \cdot) = \delta_{\phi_t(z)}.$$

De plus,  $\tilde{\pi}$  est  $Q$  et  $R_t$ -invariante.

**Preuve de la  $R_t$ -invariance :**

$$\begin{aligned}\tilde{\pi}R_t(A) &= \int \tilde{\pi}(dz)R_t(z, A) = c \int \exp\left(-\frac{H(z)}{T}\right) \mathbb{1}_A(\phi_t(z)) dz \\ &= c \int \exp\left(-\frac{H \circ \phi_t^{-1}(y)}{T}\right) \mathbb{1}_A(y) dy = c \int \exp\left(-\frac{H(y)}{T}\right) \mathbb{1}_A(y) dy \\ &= \tilde{\pi}(A).\end{aligned}$$

# Modification du HMC pour la discrétisation

Contrairement à  $\phi_t$ ,  $\hat{\phi}_L$  ne conserve pas le hamiltonien.

Idée : remplacer  $\phi_t$  par  $g \circ \hat{\phi}_L$  pour  $g$  une fonction à déterminer, i.e.  
remplacer  $R_t$  par

$$\begin{aligned}\hat{R}_L : \quad & \mathbb{R}^{2d} \times \mathcal{B}(\mathbb{R}^{2d}) &\rightarrow [0, 1] \\ & (z, A) &\mapsto \delta_{g \circ \hat{\phi}_L(z)}(A)\end{aligned}$$

# Modification du HMC pour la discrétisation

Contrairement à  $\phi_t$ ,  $\hat{\phi}_L$  ne conserve pas le hamiltonien.

Idée : remplacer  $\phi_t$  par  $g \circ \hat{\phi}_L$  pour  $g$  une fonction à déterminer, i.e.  
 remplacer  $R_t$  par

$$\begin{aligned}\hat{R}_L : \quad & \mathbb{R}^{2d} \times \mathcal{B}(\mathbb{R}^{2d}) &\rightarrow [0, 1] \\ & (z, A) &\mapsto \delta_{g \circ \hat{\phi}_L(z)}(A)\end{aligned}$$

## Proposition (réversibilité)

On suppose  $k \geq 2$  ( $U$  et  $K$  de classe  $\mathcal{C}^2$ ). On prend  $g = s$  et

$$\alpha : (z_0, z_1) \mapsto 1 \wedge \frac{h_{\tilde{\pi}}(z_1)}{h_{\tilde{\pi}}(z_0)} = 1 \wedge \exp\left(\frac{-H(z_1) + H(z_0)}{T}\right).$$

Si  $\hat{R}_L^\alpha$  est le noyau de Metropolis-Hastings associé au noyau instrumental  $\hat{R}_L$  et à la fonction de rejet  $\alpha$ , alors  $\tilde{\pi}$  est  $\hat{R}_L^\alpha$ -réversible (et donc  $\hat{R}_L^\alpha$ -invariant).



# Preuve de la réversibilité de la loi cible

Notons  $\psi_L = s \circ \hat{\phi}_L$ , de sorte que  $\hat{R}_L(z, \cdot) = \delta_{\psi_L(z)}$ . Ainsi,

$$\hat{R}_L^\alpha(z, A) = \alpha(z, \psi_L(z)) \mathbb{1}_A(\psi_L(z)) + (1 - \alpha(z, \psi_L(z))) \mathbb{1}_A(z)$$

donc

$$\tilde{\pi} \otimes \hat{R}_L^\alpha(A \times B) = \Lambda_1(A, B) + \Lambda_2(A, B)$$

où

$\Lambda_2(A, B) = \int \mathbb{1}_A(z) \mathbb{1}_B(z) (1 - \alpha(z, \psi_L(z))) e^{-H(z)/T} dz$  est symétrique

et

$$\begin{aligned} \Lambda_1(A, B) &= \int \mathbb{1}_A(z) \mathbb{1}_B(\psi_L(z)) \alpha(z, \psi_L(z)) e^{-H(z)/T} dz \\ &= \int \mathbb{1}_A(z) \mathbb{1}_B(\psi_L(z)) \left( e^{-H(z)/T} \wedge e^{-H(\psi_L(z))/T} \right) dz. \end{aligned}$$

$\psi_L^{-1} = \psi_L$  (réversibilité de  $\hat{\phi}_L$ ) et  $\det\left(\frac{d\psi_L}{dz}(z)\right) = -1$  (conservation du volume)  
 donc le changement de variable  $z \mapsto \psi_L(z)$  donne que  $\Lambda_1$  est symétrique. □

# Algorithme HMC réel

---

## Algorithme 3.2 : Hamiltonian Monte-Carlo, cas concret

---

**Données :**  $h_\pi$  proportionnel à la densité cible,  $\epsilon$  pas du saute-mouton,  $L$  nombre de pas du saute-mouton

$X_0 \leftarrow x \in \mathbb{X}$  arbitraire;

**répéter**

```
 $\tilde{P}_k \sim \nu$  et  $\tilde{P}_k \perp\!\!\!\perp (Z_0, \dots, Z_k)$ ; // Tirer la quantité de mouvement
 $\tilde{Z}_k \leftarrow (X_k, \tilde{P}_k)$ ;
 $Z_{k+1}^{\text{prop}} = (X_{k+1}^{\text{prop}}, P_{k+1}^{\text{prop}}) \leftarrow s \circ \hat{\phi}_L(\tilde{Z}_k)$ ; // Suivre la dynamique approchée
 $\alpha_{k+1} \leftarrow 1 \wedge \exp\left(\frac{-H(Z_{k+1}^{\text{prop}}) + H(\tilde{Z}_k)}{T}\right)$ ;
 $Z_{k+1} \leftarrow \begin{cases} Z_{k+1}^{\text{prop}} & \text{avec probabilité } \alpha_{k+1} \\ \tilde{Z}_k & \text{avec probabilité } 1 - \alpha_{k+1} \end{cases}$ ; // Accepter ou rejeter
```

jusqu'à une condition d'arrêt;

**retourner**  $(X_k)_k$

---

## 1 Introduction

## 2 Dynamique hamiltonienne

## 3 Hamiltonian Monte-Carlo

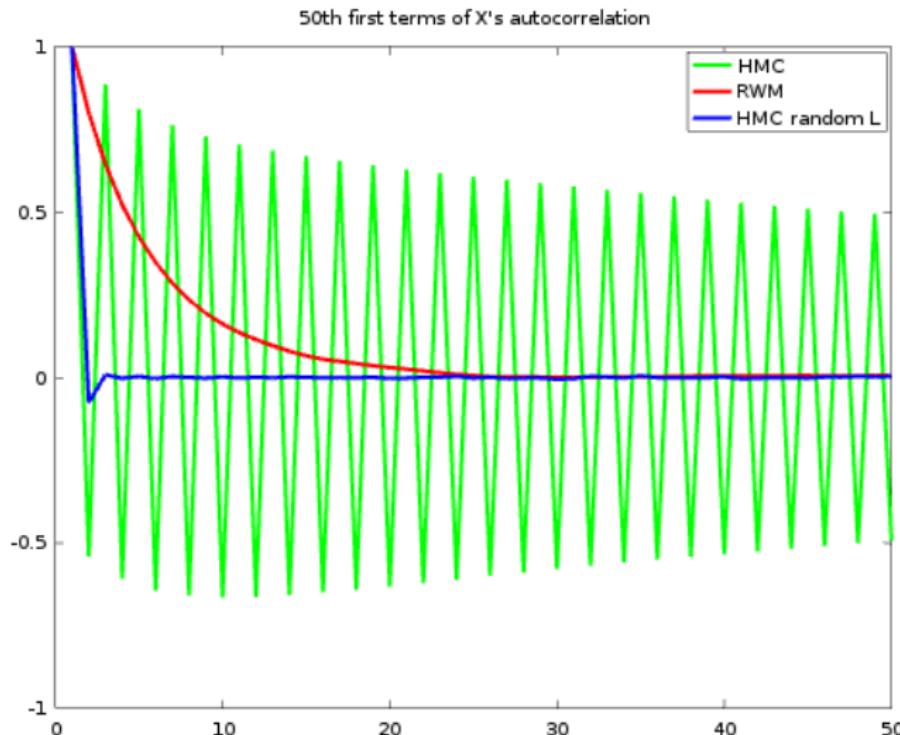
## 4 Simulations

- Loi normale
- Mélange de gaussiennes
- Bayesian Lasso

## 5 Conclusion

# Loi normale : trajectoires

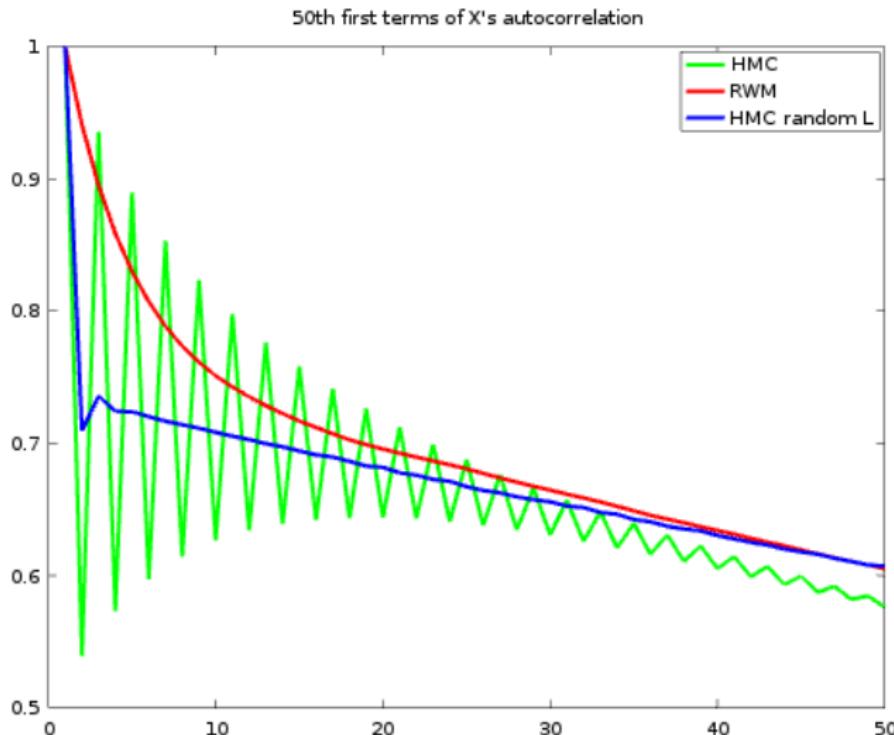
# Loi normale : autocorrelations



# Loi normale : loin du mode

# Mélange de gaussiennes : trajectoires

# Mélange de gaussiennes : autocorrelations

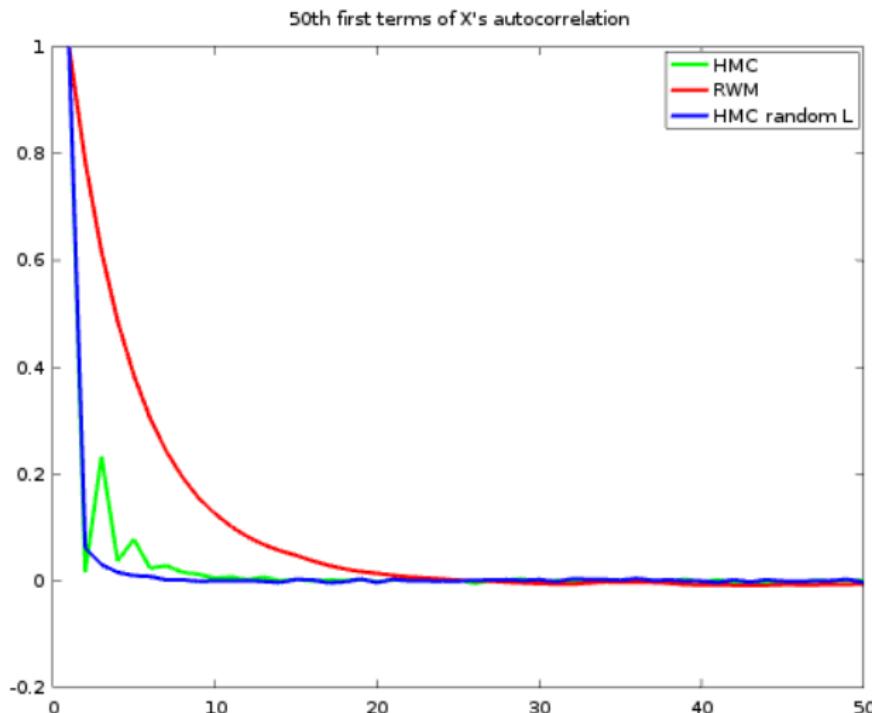


# Bayesian Lasso

- $U(x) = \|Ax - b\|_2^2 + \lambda\|x\|_1$  non différentiable.
- On considère la sous-différentielle  $(A + A^\top)x + \lambda \text{sign}(x)$

# Bayesian Lasso : trajectoires

# Bayesian Lasso : autocorrelations



# Bayesian Lasso : loin du mode

## 1 Introduction

## 2 Dynamique hamiltonienne

## 3 Hamiltonian Monte-Carlo

## 4 Simulations

## 5 Conclusion

# Conclusion

- Le HMC construit une chaîne de Markov de loi invariante la loi cible dans les cas idéal et approché.
- Certaines hypothèses sont nécessaires pour avoir l'invariance (problème dans le Bayesian Lasso).
- Prendre  $L$  aléatoire semble améliorer la stationnarité