

Riemann manifold Hamiltonian Monte Carlo methods

Pierre Boyeau Baptiste Kerléguer

École Normale Supérieure Paris-Saclay

10 janvier 2018

① Context

Hamiltonian Monte Carlo

Variété Riemannienne

Riemann Manifold Hamiltonian Monte Carlo

② Application: Régression Logistique Bayésienne

Validation de l'algorithme

Performances sur données simulées

Performances sur données réelles

③ Limites et ouvertures

① Context

Hamiltonian Monte Carlo

Variété Riemannienne

Riemann Manifold Hamiltonian Monte Carlo

② Application: Régression Logistique Bayésienne

Validation de l'algorithme

Performances sur données simulées

Performances sur données réelles

③ Limites et ouvertures

Hamiltonian Monte Carlo Classique

Idée: Introduction d'une variable auxiliaire \mathbf{p} :

Log-proba négative cible: Hamiltonien:

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\mathcal{L}(\boldsymbol{\theta}) + \frac{1}{2} \log (2\pi)^2 |\mathbf{M}| + \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}$$

avec M une matrice de masse

Équation d'évolution :

$$\frac{d\boldsymbol{\theta}}{d\tau} = \frac{\partial H}{\partial \mathbf{p}} \quad (1)$$

$$\frac{d\mathbf{p}}{d\tau} = -\frac{\partial H}{\partial \boldsymbol{\theta}} \quad (2)$$

Variété Riemannienne

Idée: Prendre en compte des arguments de géométrie Riemannienne pour *sampler* $p: M \rightarrow \theta \ G(\theta)$.

Nouvelle équation d'évolution:

$$\frac{d\theta}{d\tau} = G(\theta)^{-1} \mathbf{p} \quad (3)$$

$$\frac{d\mathbf{p}}{d\tau} = \nabla_{\theta} \mathcal{L}(\theta) - \frac{1}{2} \text{tr} \left(G(\theta)^{-1} \frac{\partial G(\theta)}{\partial \theta} \right) + \frac{1}{2} \mathbf{p}^T (\theta)^{-1} \frac{\partial G(\theta)}{\partial \theta} (\theta)^{-1} \mathbf{p} \quad (4)$$

\Rightarrow Calcul de $G(\theta)^{-1}$ nécessaire.

Choix de $G(\theta)$

Prise en compte de la réalité géométrique locale

- Espérance de l'information de Fisher
- Observation de l'information de Fisher observée
- Information de Fisher empirique
- Autres tenseurs

On suppose qu'on a θ_n . Gibbs sampling:

- 1 Sampler $\mathbf{p}_{n+1} | \theta_n \sim \mathcal{N}(0, G(\theta_n))$
- 2 Le *proposal* (θ_*, \mathbf{p}_*) est obtenu à l'aide de l'intégrateur de Verlet.
 θ_* accepté avec probabilité:

$$\min(1, \exp[H(\theta_n, \mathbf{p}_{n+1}) - H(\theta_*, \mathbf{p}_*)])$$

Interprétation et remarques

Intégrateur de Verlet

Pour assurer un MCMC correct, l'intégration des équations de trajectoire doit être

- Réversible
- Préservation du volume

$$\begin{aligned}p\left(\tau + \frac{\epsilon}{2}\right) &= p(\tau) - \frac{\epsilon}{2} \nabla_{\theta} H(\theta(\tau), p\left(\tau + \frac{\epsilon}{2}\right)) \\ \theta(\tau + \epsilon) &= \theta(\tau) \\ &+ \frac{\epsilon}{2} \left[\nabla_p H\left(\theta(\tau), p\left(\tau + \frac{\epsilon}{2}\right)\right) + \nabla_p H\left(\theta(\tau + \epsilon), p\left(\tau + \frac{\epsilon}{2}\right)\right) \right] \\ p(\tau + \epsilon) &= p\left(\tau + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2} \nabla_{\theta} H\left(\theta\left(\tau + \frac{\epsilon}{2}\right), p\left(\tau + \frac{\epsilon}{2}\right)\right)\end{aligned}$$

Algorithme

```
Require:  $G(0) - \mathcal{L}(\theta) \theta^0$   
for  $i = 0 : N - 1$  do  
  sample  $p^{i+1}$  selon  $\mathcal{N}(O, G^i)$   
  for  $j = 1 : Nb_{leapfrogs}$  do  
     $p_{tempo} = p_\tau$   
    for  $k = 1 : N_{pointfixe}$  do  
       $p\left(\tau + \frac{\epsilon}{2}\right) = p(\tau) - \frac{\epsilon}{2} \nabla_\theta H(\theta(\tau), p_{tempo})$   
    end for  
     $p\left(\tau + \frac{\epsilon}{2}\right) = p_{tempo}$   
     $\theta_{tempo} = \theta(\tau)$   
    for  $k = 1 : N_{pointfixe}$  do  
       $\theta(\tau + \epsilon) = \theta(\tau) + \frac{\epsilon}{2} [G^{-1}(\theta(\tau)) + G^{-1}(\theta_{tempo})]$   
    end for  
     $\theta(\tau + \epsilon) = \theta_{tempo}$   
     $p(\tau + \epsilon) = p\left(\tau + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2} \nabla_\theta H\left(\theta\left(\tau + \frac{\epsilon}{2}\right), p\left(\tau + \frac{\epsilon}{2}\right)\right)$   
  end for  
end for
```

Plan

① Context

Hamiltonian Monte Carlo

Variété Riemannienne

Riemann Manifold Hamiltonian Monte Carlo

② Application: Régression Logistique Bayésienne

Validation de l'algorithme

Performances sur données simulées

Performances sur données réelles

③ Limites et ouvertures

Régression Logistique Bayésienne

Modèle:

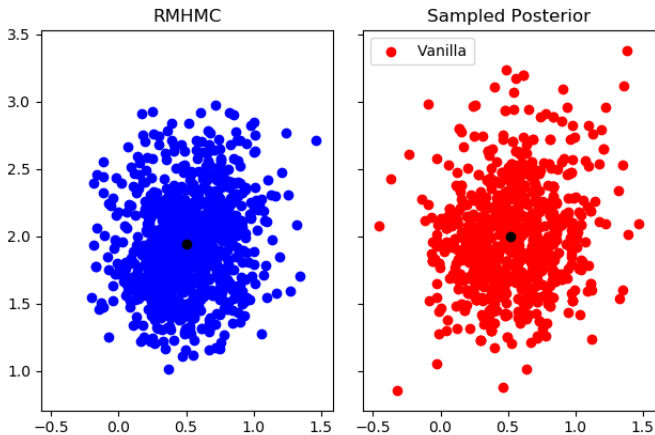
- 1 Prior gaussien: $\beta \sim \mathcal{N}(0, \alpha I)$
- 2 Loi $Y|\beta$: modèle de régression logistique classique

Tenseur métrique choisi: information de Fisher

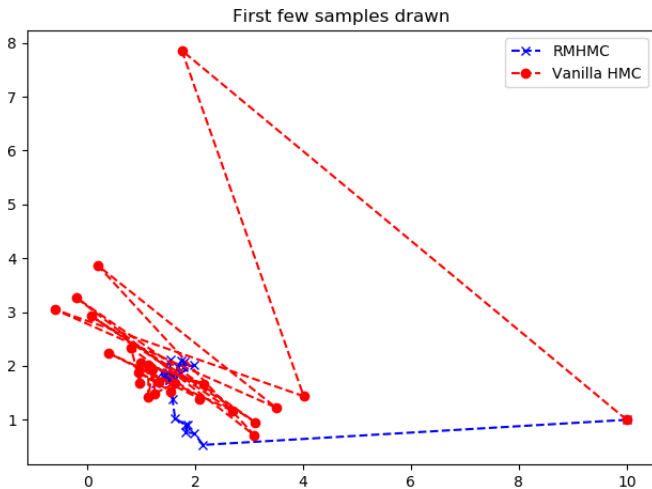
$$G(\beta) = \mathbf{X}^T \mathbf{\Lambda} \mathbf{X} + \alpha^{-1} I$$

Travail réalisé: Etude comparée de RMHMC et de HMC sur cette tâche et sur données simulées.

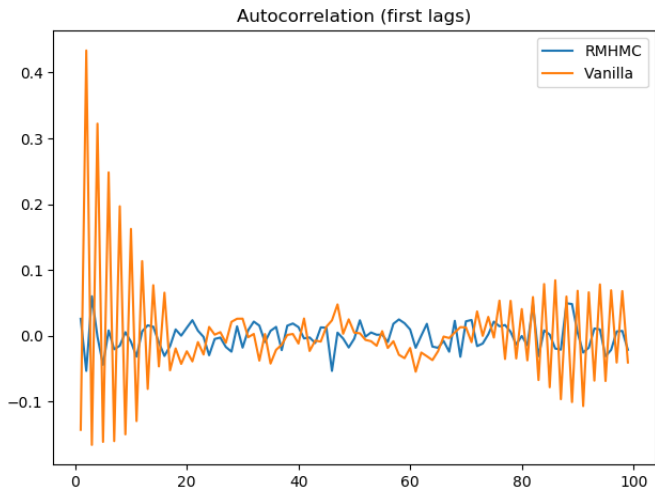
Validation de l'algorithme



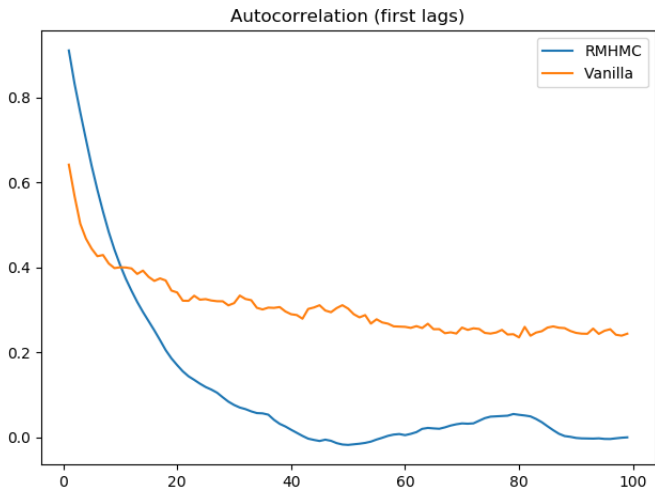
Performances



Performances



Données réelles



Plan

1 Context

Hamiltonian Monte Carlo

Variété Riemannienne

Riemann Manifold Hamiltonian Monte Carlo

2 Application: Régression Logistique Bayésienne

Validation de l'algorithme

Performances sur données simulées

Performances sur données réelles

3 Limites et ouvertures

Limites et ouvertures

Limites:

- Coût computationnel important
- Temps à passer à chercher les hyperparamètres (HMC) devient un temps d'implémentation (RMHMC)

Ouverture

- Acceptance ratio
- Choix de $G(\theta)$

