Discrétisation de la dynamique de Langevin pour l'échantillonnage de mesures de probabilité en grande dimension

Régis SANTET

(CERMICS, École des Ponts & Équipe MATHERIALS, Inria Paris)

Encadrement : G.Stoltz, T.Lelièvre, J.Reygner

Soutenance PFE (ENPC, 3 Mars 2022)

Plan

Algorithmes MCMC

But : Évaluer

$$\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x$$

avec l'estimateur

$$\hat{I}_N := \frac{1}{N}\sum_{i=1}^{N} f(X^i), \qquad X^i \sim \pi$$

- Algorithme de Metropolis-Hastings [1] : noyau de proposition T + acception/rejection avec probabilité

$$\min\left(1, \frac{\pi(x')T(x', \mathrm{d}x)}{\pi(x)T(x, \mathrm{d}x')}\right)$$

- Metropolis Adjusted Langevin Algorithm [2] : T engendré par discrétisation Euler-Maruyama de la dynamique de Langevin suramortie

$$\mathrm{d}X_t = \nabla \log \pi(X_t)\ \mathrm{d}t + \sqrt{2}\ \mathrm{d}W_t$$

1. Metropolis/Rosenbluth A./Rosenbluth M./Teller A./Teller E. (1953)
2. Rossky/Doll/Friedman (1978)

Hamiltonian Monte Carlo [3, 4]

Hamiltonien du système $H(q,p) = V(q) + \frac{1}{2}p^\mathsf{T} M^{-1} p$.

Pour un pas de temps $\Delta t > 0$ et l'état de la chaîne $q^0 \in \mathbb{R}^d$,

(HMC.i) Générer $p^0 \sim \mathcal{N}_d(0, M)$.

(HMC.ii) Intégrer la dynamique Hamiltonienne avec un flot numérique $\varphi_{\Delta t}$

$$\begin{cases} \dot{q} = \nabla_p H(q,p) = M^{-1}p \\ \dot{p} = -\nabla_q H(q,p) = -\nabla V(q) \end{cases}$$

sur un pas de temps avec conditions initiales $(q(0), p(0)) = (q^0, p^0)$.

(HMC.iii) Calculer le ratio de Metropolis

$$r_{\Delta t}(q,p) = \min\left\{1, \mathrm{e}^{-\beta[(H\circ\varphi_{\Delta t})(q,p) - H(q,p)]}\right\}$$

(HMC.iv) Simuler $U \sim \mathcal{U}(0,1)$. Si $U \leqslant r_{\Delta t}$, renvoyer $q^1 = q(\Delta t)$. Sinon, renvoyer $q^1 = q^0$.

3. Duane/Kennedy/Pendleton/Roweth (1987)
4. Neal (1993)

HMC : conservation de la mesure canonique

Théorème

L'algorithme HMC préserve la mesure de probabilité

$$\mu = \exp\left(-H(q,p)\right)/Z_\mu \,\mathrm{d}q\,\mathrm{d}p$$

Preuve

$T_{\Delta t}((q,p), \mathrm{d}q'\,\mathrm{d}p') = r_{\Delta t}\delta_{\varphi_{\Delta t}(q,p)}(\mathrm{d}q'\,\mathrm{d}p') + (1 - r_{\Delta t}(q,p))\delta_{(q,p)}(\mathrm{d}q'\,\mathrm{d}p')$

Si $f : \mathbb{R}^d \to \mathbb{R}^d \to \mathbb{R}$ mesurable bornée, [x=(q,p), S(q,p)=(q,-p)]

$$\begin{aligned}
\int r_{\Delta t}(x) f(\varphi_{\Delta t}(x))\mu(\mathrm{d}x) &= \int r_{\Delta t}(\varphi_{\Delta t}^{-1}(y)) f(y) \frac{\mathrm{e}^{-\beta\left[H\circ\varphi_{\Delta t}^{-1}\right](y)}}{Z_\mu}\mathrm{d}y \\
[|\nabla\varphi_{\Delta t}| = 1] \quad &= \int r_{\Delta t}((S\circ\varphi_{\Delta t})(z)) f(z) \frac{\mathrm{e}^{-\beta[H\circ S\circ\varphi_{\Delta t}](z)}}{Z_\mu}\mathrm{d}z \\
[S\circ\varphi_{\Delta t}\circ S = \varphi_{\Delta t}^{-1}] \quad &= \int r_{\Delta t}(z) f(z)\mu(\mathrm{d}z)
\end{aligned}$$

Plan

Diffusion non constante

Dynamique de Langevin suramortie avec *bruit multiplicatif*[5]
$D\colon \mathbb{R}^d \ni q \to D(q) \in \mathcal{S}_d^{++}(\mathbb{R})$

$$\mathrm{d}q_t = \left(-D(q_t)\nabla V(q_t) + \frac{1}{\beta}\mathrm{div}\ D(q_t)\right)\mathrm{d}t + \sqrt{\frac{2}{\beta}D(q_t)}\mathrm{d}W_t$$

$\rightarrow$ Promesse de proposer un échantillonnage plus rapide (métastabilité, anisotropie)

Extension à la **dynamique de Langevin** (ré-échantillonnage partiel du moment)

$$\begin{cases} \mathrm{d}q_t = \nabla_p H(q_t, p_t)\ \mathrm{d}t \\ \mathrm{d}p_t = -\nabla_q H(q_t, p_t)\ \mathrm{d}t - \gamma \nabla_p H(q_t, p_t)\ \mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\ \mathrm{d}W_t \end{cases}$$

5. Roberts/Stramer (2002)

En pratique : utiliser des schémas numériques symplectiques et réversibles en temps[6,7]

Generalized Störmer–Verlet

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2} \nabla_q H(q^n, p^{n+1/2}), \\ q^{n+1} = q^n + \dfrac{\Delta t}{2} \left(\nabla_p H(q^n, p^{n+1/2}) + \nabla_p H(q^{n+1}, p^{n+1/2}) \right) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2} \nabla_q H(q^{n+1}, p^{n+1/2}) \end{cases}$$

Implicit Midpoint-Rule

$$\begin{cases} q^{n+1} = q^n + \Delta t \ \nabla_p H \left(\dfrac{y^{n+1} + y^n}{2} \right) \\ p^{n+1} = p^n - \Delta t \ \nabla_q H \left(\dfrac{y^{n+1} + y^n}{2} \right) \end{cases}$$

Hamiltonien non séparable $\rightarrow$ **méthodes implicites**

---

6. Leimkuhler/Reich (2005)
7. Hairer/Wanner/Lubich (2006)

Schémas faiblement consistants

Proposition

GSV & IMR sont faiblement consistants avec la dynamique de Langevin suramortie avec bruit multiplicatif si

$$H(q,p) = V(q) - \frac{1}{2}\ln\left(\det D(q)\right) + \frac{1}{2}p^\mathsf{T} D(q) p$$

i.e.

$$\begin{aligned} q_{n+1} = q_n + \Delta t \; D(q_n)^{1/2}\mathrm{G}_n + \frac{\Delta t^2}{2}\mathcal{F}_2(q_n, D(q_n)^{-1/2}\mathrm{G}_n) \\ + \Delta t^3 \mathcal{F}_3(q_n, D(q_n)^{-1/2}\mathrm{G}_n) + \mathcal{O}(\Delta t^4) \end{aligned}$$

avec

$$\begin{aligned} \mathbb{E}_{\mathrm{G}_n}\left[\mathcal{F}_2(q_n, D(q_n)^{-1/2}\mathrm{G}_n)\right] &= -D(q_n)\nabla V(q_n) + \mathrm{div}\; D(q_n) \\ \mathbb{E}_{\mathrm{G}_n}\left[\mathcal{F}_3(q_n, D(q_n)^{-1/2}\mathrm{G}_n)\right] &= 0 \end{aligned}$$

Riemannian Manifold HMC [8]

$$H(q,p) = V(q) - \frac{1}{2}\ln\left(\det D(q)\right) + \frac{1}{2}p^\mathsf{T} D(q) p$$

- $\displaystyle\int \exp\left(-H\right)\mathrm{d}p \propto \exp(-V)$
- $p \mid q \sim \mathcal{N}(0, D(q)^{-1})$

GSV (réversible en temps, symplectique)

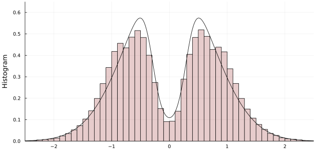
$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla_q H(q^n, p^{n+1/2}) \\ q^{n+1} = q^n + \dfrac{\Delta t}{2}\left(\nabla_p H(q^n, p^{n+1/2}) + \nabla_p H(q^{n+1}, p^{n+1/2})\right) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla_q H(q^{n+1}, p^{n+1/2}) \end{cases}$$

8. Girolami/Calderhead (2011)

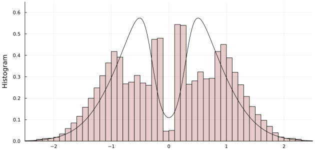
Apparition de biais dans l'échantillonnage (HMC)

Potentiel double puits confinant $V(q) = q^2 - 1 + \frac{h}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{q^2}{2\sigma^2}\right)$

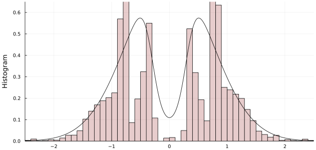
$\Delta t = 0.18$



$\Delta t = 0.69$



$\Delta t = 0.86$



$\Delta t = 1.68$

Méthodes implicites $\Rightarrow$ problèmes de **convergence** et de **réversibilité numérique**[9]

9. Brofos/Lederman (2021)

Plan

Analyse numérique

On se donne un **schéma numérique** $\Phi_{\Delta t}$ (possiblement implicite)

$$\Phi_{\Delta t}(q,p,q',p') = \begin{bmatrix} q' + \text{update of } q \\ p' + \text{update of } p \end{bmatrix}$$

Couples admissibles

$$\mathscr{D} = \left\{ (q,p,q',p') \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \,\middle|\, \Phi_{\Delta t}(q,p,q',p') = 0 \right\}$$

Travail dans un ouvert $\mathcal{D} \subset \mathscr{D}$ (différentiabilité, symplecticité)

$$\mathcal{A} = \left\{ (q,p) \in \mathbb{R}^d \times \mathbb{R}^d \,\middle|\, \exists (q',p') \in \mathbb{R}^d \times \mathbb{R}^d, (q,p,q',p') \in \mathcal{D} \right\}$$

Le **flot numérique** est

$$\varphi_{\Delta t} \colon \mathcal{A} \ni (q,p) \mapsto (p',q') \in \mathbb{R}^d \times \mathbb{R}^d$$

Peut être multivarié en théorie $\to$ problèmes de réversibilité

Bifurcations forward/backward

$$
\begin{tikzcd}
 & (q^n, p^n_+) \arrow[r] & (q^{n+1}_+, p^n_+) \arrow[r] & (q^{n+1}_+, p^{n+1}_+) \\
(q^n, p^n) \arrow[ur] \arrow[dr] & & & \\
 & (q^n, p^n_-) \arrow[r] & (q^{n+1}_-, p^n_-) \arrow[r] & (q^{n+1}_-, p^{n+1}_-)
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & (q^{n+1}_+, -p^{n+1}_{++}) \arrow[r] & (q^{n+1}_{++}, -p^{n+1}_{++}) \arrow[r] & (q^n, -p^n) \\
(q^{n+1}_+, -p^{n+1}_+) \arrow[ur] \arrow[dr] & & & \\
 & (q^{n+1}_+, -p^{n+1}_{+-}) \arrow[r] & (q^{n+1}_{+-}, -p^{n+1}_{+-}) \arrow[r] & (q^{n+1}_{+-}, -\tilde{p}^n)
\end{tikzcd}
$$

Algorithme avec réversibilité numérique

On définit $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ où $S(q,p) = (q,-p)$. On se donne un schéma numérique $\mathcal{C}^1$ réversible et symplectique

$$(i)\ \Phi_{\Delta t}(q,p,\varphi_{\Delta t}(q,p)) = 0 \Longleftrightarrow \Phi_{\Delta t}((S\circ\varphi_{\Delta t})(q,p), S(q,p)) = 0,$$

$$(ii)\ \nabla_{q,p}\varphi_{\Delta t}(q,p)^{\mathsf{T}} J \nabla_{q,p}\varphi_{\Delta t}(q,p) = J, \qquad J = \begin{pmatrix} 0 & \mathrm{I}_d \\ -\mathrm{I}_d & 0 \end{pmatrix}.$$

On définit

$$\mathcal{B} = \left\{ (q,p) \in \mathcal{A} \,\middle|\, \begin{array}{c} \psi_{\Delta t}(q,p) \in \mathcal{A}, (\psi_{\Delta t}\circ\psi_{\Delta t})(q,p) = (q,p) \\ \text{et } \nabla_{q',p'}\Phi_{\Delta t}(q,p,\varphi_{\Delta t}(q,p)) \text{ est inversible} \end{array} \right\},$$

et

$$\forall (q,p) \in \mathbb{R}^d \times \mathbb{R}^d, \qquad \psi^{\text{rev}}_{\Delta t}(q,p) = \psi_{\Delta t}(q,p)\mathbf{1}_{(q,p)\in\mathcal{B}} + (q,p)\mathbf{1}_{(q,p)\notin\mathcal{B}}.$$

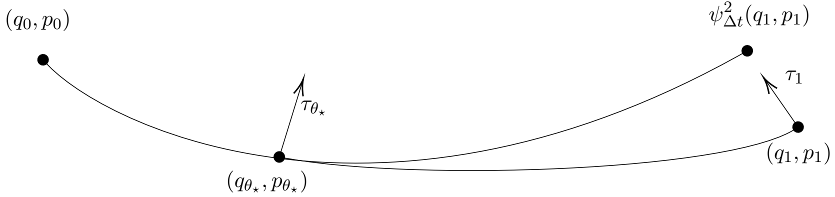
Proposition

La fonction $\psi^{\text{rev}}_{\Delta t}$ est une involution et préserve la mesure de Lebesgue sur $\mathbb{R}^d \times \mathbb{R}^d$.

Structure de la preuve[10]

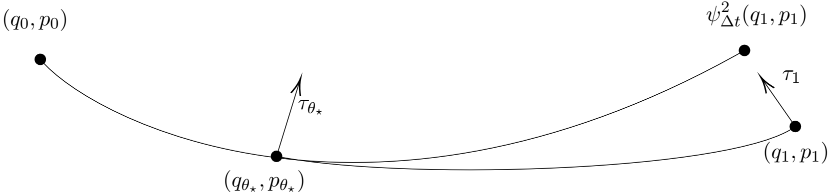
- On montre que B est ouvert comme réunion de composantes connexes ouvertes de $\mathcal{A} \cap \psi_{\Delta t}^{-1}(\mathcal{A})$.
- Raisonnement par contradiction : on a $(q_0, p_0) = \psi_{\Delta t}^2(q_0, p_0)$ et on suppose que $\psi_{\Delta t}^2(q_1, p_1) \neq (q_1, p_1)$ pour un certain (q_1, p_1).
- Relier (q_0, p_0) à (q_1, p_1) et écrire

$$\psi_{\Delta t}^2(q, p) = (q, p) + \varepsilon\tau + \mathcal{O}(\varepsilon), \qquad \|\tau\| = 1$$



10. Lelièvre/Rousset/Stoltz (2019)

Structure de la preuve



- Utiliser la réversibilité en temps du schéma [+ caractère $\mathcal{C}^1$, compacité, passage à la limite] pour obtenir

$$\nabla_{q',p'}\Phi_{\Delta t}(q_{\theta_\star}, p_{\theta_\star}, \varphi_{\Delta t}(q_{\theta_\star}, p_{\theta_\star})) \cdot \tau_{\theta_\star} = 0$$

  contradiction via la définition de $\mathcal{B}$ (dernière propriété)
- On montre que $\mathcal{B}$ et $\mathcal{B}^c$ sont stables par $\psi^{\mathsf{rev}}_{\Delta t}$ + symplecticité et involution sur chaque composante

Algorithme GHMC avec réversibilité numérique

L'état de la chaîne est $X^0 = (q^0, p^0) \in \mathbb{R}^d \times \mathbb{R}^d$

(GHMC Rev.i) Intégrer le processus d'Ornstein-Uhlenbeck avec un pas de temps $\Delta t/2$ et conditions initiales $(q^0, p^0) \to (q^0, p^{1/4})$.

(GHMC Rev.ii) Intégrer sur un pas de temps la dynamique Hamiltonienne **avec vérification de réversibilité numérique et renversement du moment** :

$$(\widetilde{q}^1, \widetilde{p}^{3/4}) = \psi_{\Delta t}^{\mathsf{rev}}(q^0, p^{1/4}),$$

(GHMC Rev.iii) Simuler $U^0 \sim \mathcal{U}(0,1)$:

- si $U^0 \leqslant \exp(-\beta \left(H(\widetilde{q}^1, \widetilde{p}^{3/4}) - H(q^0, p^{1/4})\right))$, accepter la transition : $(q^1, p^{3/4}) = (\widetilde{q}^1, -\widetilde{p}^{3/4})$
- sinon la rejeter : $(q^1, p^{3/4}) = (q^0, p^{1/4})$.

(GHMC Rev.iv) Renverser le moment : $\widetilde{p}^1 = -p^{3/4}$.

(GHMC Rev.v) Intégrer Ornstein-Uhlenbeck sur un demi pas de temps et avec conditions initiales $(q^1, \widetilde{p}^1) \to (q^1, p^1)$.

(GHMC Rev.vi) Renvoyer $X^1 = (q^1, p^1)$.

Exemples numériques

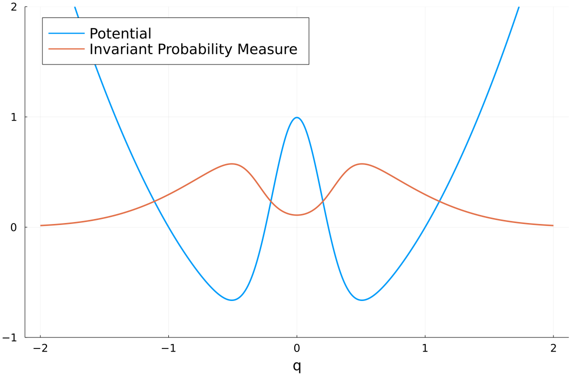


Figure – Double puits confinant $V_{\sigma,h}(q) = q^2 - 1 + \frac{h}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{q^2}{2\sigma^2}\right)$. avec $\sigma = 0.2$, $h = 1$ et la mesure invariante associée $\mathrm{e}^{-\beta V}$ avec $\beta = 1$.

Le coefficient de diffusion choisi est $D(q) = \left(\frac{1.5 + \cos(\pi q)}{2}\right)^2 > 0$

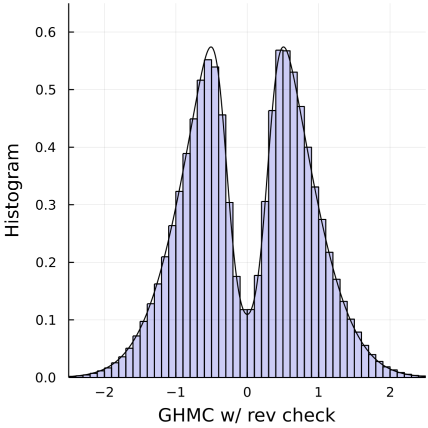


Figure – Échantillonnage avec un pas de temps $\Delta t = 0.28$. L'histogramme à gauche est obtenu grâce à notre algorithme, celui de droite est obtenu sans vérifier la réversibilité numérique.

$$\varphi_{\Delta t}^{\mathrm{rev}} \underset{\Delta t \to 0}{\approx} \varphi_{\Delta t} \Longleftrightarrow \mathcal{B} \approx \mathbb{R} \times \mathbb{R}$$

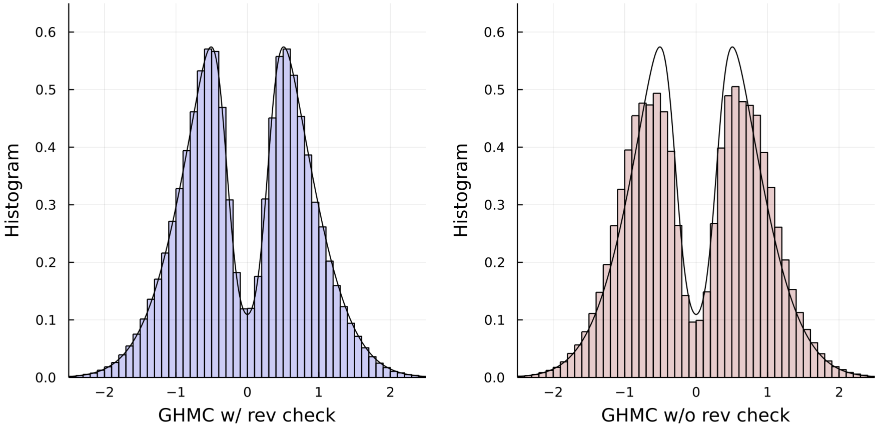


Figure – Échantillonnage avec un pas de temps $\Delta t = 0.44$.

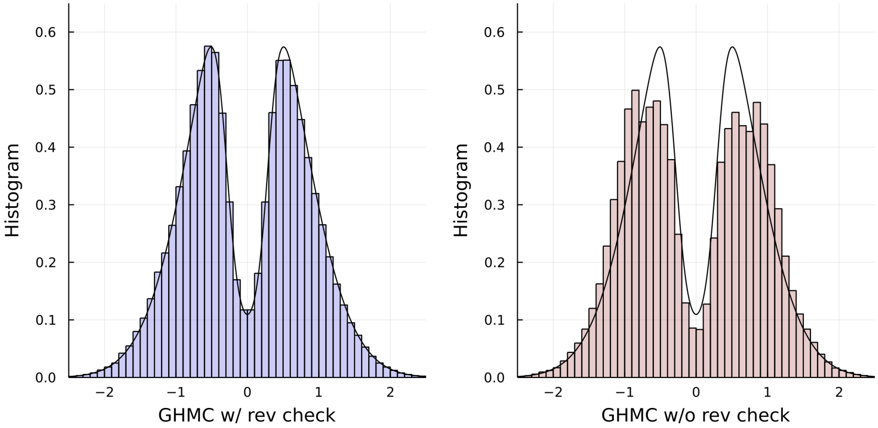
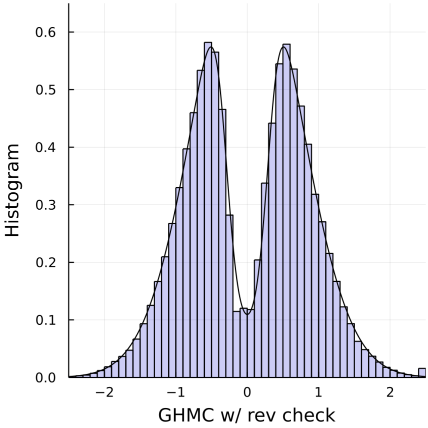


Figure – Échantillonnage avec un pas de temps $\Delta t = 0.69$.



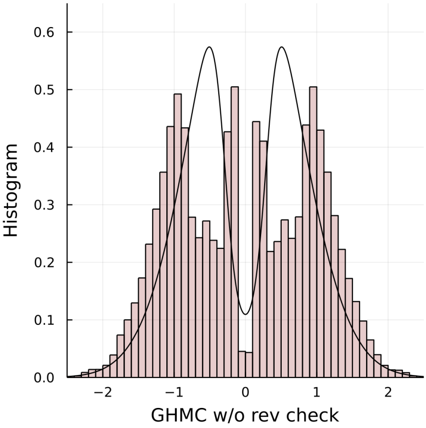


Figure – Échantillonnage avec un pas de temps $\Delta t = 0.86$.

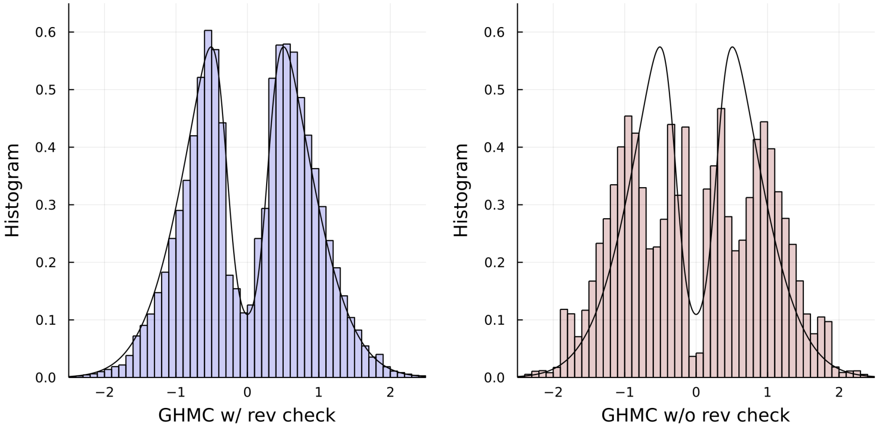


Figure – Échantillonnage avec un pas de temps $\Delta t = 1.08$.

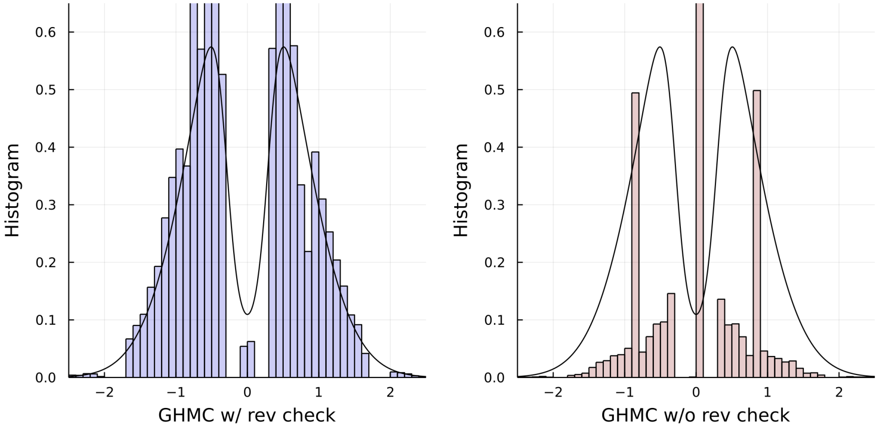
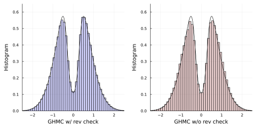
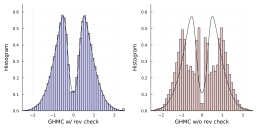
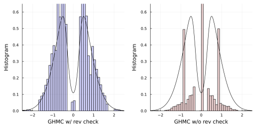


Figure – Échantillonnage avec un pas de temps $\Delta t = 2.1$.







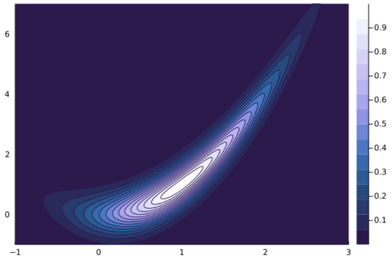
- Notre algorithme échantillonne **sans biais** la mesure de probabilité cible
- Il résiste aux problèmes de convergence/réversibilité plus longtemps (quand $\Delta t \uparrow$)

- Certaines régions de l'espace ne deviennent plus accessibles quand $\Delta t \uparrow$ ($|\mathcal{B}| \downarrow$)
- Les problèmes venant de l'utilisation de méthodes implicites finissent par empêcher l'échantillonnage

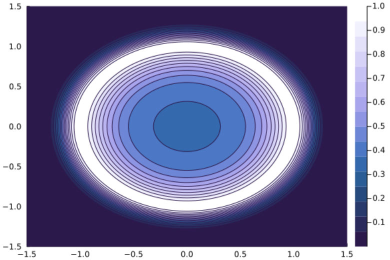
$\Rightarrow$ Trouver le meilleur flot $\varphi_{\Delta t}$ pour que $\mathcal{B}$ soit le plus grand possible ?

Plan

- Exemples numériques (2d, anisotropie, métastabilitié)



$V(x,y) = (a-x)^2 + b(y-x^2)^2.$



$V(x,y) = ((x^2+y^2)^2 - R^2)^2.$

- Utilisation de schémas (semi-)explicites pour l'intégration de la dynamique Hamiltonienne, en étendant l'espace des phases (deux copies du même système (q,p)) (exploratoire)

Travail futur 2 : optimisation de la diffusion

- Trouver la diffusion optimale : quel critère numérique ?

Dynamique de Langevin suramortie : optimisation sur la fonction D [maximalisation du trou spectral[11]]

$$\mathrm{d}q_t = \left(-D(q_t)\nabla V(q_t) + \frac{1}{\beta}\mathrm{div}\ D(q_t)\right)\mathrm{d}t + \sqrt{\frac{2}{\beta}D(q_t)}\mathrm{d}W_t$$

Mesure cible : $\nu(\mathrm{d}q) = Z_\nu^{-1}\mathrm{e}^{-\beta V(q)}$

- Adapter la méthode à des systèmes plus complexes :
  - plus grande dimension
  - Langevin cinétique
  - hors d'équilibre : $-D(q_t)\nabla V(q_t) + \frac{1}{\beta}\mathrm{div}\ D(q_t) + F_\eta$

- Analyser l'efficacité de la méthode numérique : compromis entre convergence rapide vers le régime permanent et un échantillonnage accéléré

11. Lelièvre/Pavliotis/Robin/Stoltz (WIP)