On the convergence of the HMC algorithm and NUTS sampler.

Samuel Gruffaz

Introduction to the recent paper :
« On the convergence of dynamic implementations of Hamiltonian Monte Carlo and No U-Turn Samplers »,
Alain Durmus, Samuel Gruffaz, Miika Kailas, Eero Saksman, Matti Vihola, July- 2023

18/09

Introduction

1. Motivations and contributions.
2. Hamiltonian Monte Carlo (HMC) and the NUTS intuition.
3. Dynamic HMC algorithms.

Introduction

Bayesian Computational goals :

Compute $\int f \, \mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) :

Markov Chain Monte Carlo (MCMC) :

Metropolis Hasting algorithm (MH) :

Hamiltonian Monte Carlo (HMC) :
[Duane and al, 1987]

State of the art

No U-Turn Sampler (NUTS) :
[Hoffman and al, 2011]

Introduction

How ?

Bayesian Computational goals :

Compute $\int f \,\mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) :

Given a *sampling method* for π and $N \in \mathbb{N}^*$, $\int f \,\mathrm{d}\pi \approx \sum_{i=1}^{N} f(X_i)/N, \quad X_i \overset{i.i.d}{\sim} \pi, \quad 1 \leq i \leq N$.

Markov Chain Monte Carlo (MCMC) :

Metropolis Hasting algorithm (MH) :

Hamiltonian Monte Carlo (HMC) :
[Duane and al, 1987]

State of the art
No U-Turn Sampler (NUTS) :
[Hoffman and al, 2011]

Introduction

How ?

Bayesian Computational goals :

Compute $\int f \, \mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) :

Given a *sampling method* for π and $N \in \mathbb{N}^*$, $\int f \, \mathrm{d}\pi \approx \sum_{i=1}^{N} f(X_i)/N, \quad X_i \overset{i.i.d}{\sim} \pi, \quad 1 \leq i \leq N.$

Markov Chain Monte Carlo (MCMC) :

Given a *Markov Kernel* K s.t. $\lim_{n\to\infty} ||\mathrm{K}^n(x,\cdot) - \pi||_{TV} = 0, \quad x \in \mathbb{R}^d,$

and $(n_s, n_b, x_0) \in (\mathbb{N}^*)^2 \times \mathbb{R}^d$, then generate $X_i \sim \mathrm{K}^{i \times n_s + n_b}(x_0, \cdot), \quad 1 \leq i \leq N.$

Metropolis Hasting algorithm (MH) :

Hamiltonian Monte Carlo (HMC) :
[Duane and al, 1987]

State of the art

No U-Turn Sampler (NUTS) :
[Hoffman and al, 2011]

Introduction

How ?

Bayesian Computational goals :

Compute $\int f \, \mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) :

Given a *sampling method* for π and $N \in \mathbb{N}^*$, $\int f \, \mathrm{d}\pi \approx \sum_{i=1}^{N} f(X_i)/N, \quad X_i \overset{i.i.d}{\sim} \pi, \quad 1 \leq i \leq N$.

Markov Chain Monte Carlo (MCMC) :

Given a *Markov Kernel* K s.t. $\lim_{n\to\infty} ||\mathrm{K}^n(x, \cdot) - \pi||_{TV} = 0, \quad x \in \mathbb{R}^d$,
and $(n_s, n_b, x_0) \in (\mathbb{N}^*)^2 \times \mathbb{R}^d$, then generate $X_i \sim \mathrm{K}^{i \times n_s + n_b}(x_0, \cdot), \quad 1 \leq i \leq N$.

Metropolis Hasting algorithm (MH) :

Given a *proposition kernel* $\tilde{\mathrm{K}}$, at iteration n :
1. Generate $\tilde{X}_{n+1} \sim \tilde{\mathrm{K}}(\cdot \,|\, X_n)$, $U_n \sim U([0,1])$, compute $\alpha_n(\tilde{X}_{n+1}, X_n)$.
2. Accept or Reject, set $X_{n+1} = \tilde{X}_{n+1}$ if $U_n \leq \alpha_n$, otherwise set $X_{n+1} = X_n$.

Hamiltonian Monte Carlo (HMC) :
[Duane and al, 1987]

State of the art

No U-Turn Sampler (NUTS) :
[Hoffman and al, 2011]

Introduction

How ?

Bayesian Computational goals : Compute $\int f \, \mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) : Given a *sampling method* for π and $N \in \mathbb{N}^*$, $\int f \, \mathrm{d}\pi \approx \sum_{i=1}^{N} f(X_i)/N, \quad X_i \overset{i.i.d}{\sim} \pi, \quad 1 \leq i \leq N$.

Markov Chain Monte Carlo (MCMC) : Given a *Markov Kernel* K s.t. $\lim_{n\to\infty} || \mathrm{K}^n(x, \cdot) - \pi ||_{TV} = 0, \quad x \in \mathbb{R}^d$, and $(n_s, n_b, x_0) \in (\mathbb{N}^*)^2 \times \mathbb{R}^d$, then generate $X_i \sim \mathrm{K}^{i \times n_s + n_b}(x_0, \cdot), \quad 1 \leq i \leq N$.

Metropolis Hasting algorithm (MH) : Given a *proposition kernel* $\tilde{\mathrm{K}}$, at iteration n :

1. Generate $\tilde{X}_{n+1} \sim \tilde{\mathrm{K}}(\cdot \mid X_n)$, $U_n \sim U([0,1])$, compute $\alpha_n(\tilde{X}_{n+1}, X_n)$.
2. Accept or Reject, set $X_{n+1} = \tilde{X}_{n+1}$ if $U_n \leq \alpha_n$, otherwise set $X_{n+1} = X_n$.

Hamiltonian Monte Carlo (HMC) : [Duane and al, 1987]

Given the the target **potential** $U = -\log \pi$ and its gradient ∇U,
a *stepsize* h and a *number of steps* T,
define $\tilde{\mathrm{K}}_{h,T}$ by integrating a system of
Hamiltonian equations using the **leapfrog** integrator.

State of the art

No U-Turn Sampler (NUTS) : [Hoffman and al, 2011]

Introduction

How ?

Bayesian Computational goals :

Compute $\int f \, \mathrm{d}\pi$, denoting by f a test function and π a target distribution.

Monte Carlo (MC) :

Given a *sampling method* for π and $N \in \mathbb{N}^*$, $\int f \, \mathrm{d}\pi \approx \sum_{i=1}^{N} f(X_i)/N, \quad X_i \overset{i.i.d}{\sim} \pi, \quad 1 \leq i \leq N.$

Markov Chain Monte Carlo (MCMC) :

Given a *Markov Kernel* K s.t. $\lim_{n\to\infty} ||\mathrm{K}^n(x, \cdot) - \pi||_{TV} = 0, \quad x \in \mathbb{R}^d,$
and $(n_s, n_b, x_0) \in (\mathbb{N}^*)^2 \times \mathbb{R}^d$, then generate $X_i \sim \mathrm{K}^{i \times n_s + n_b}(x_0, \cdot), \quad 1 \leq i \leq N.$

Metropolis Hasting algorithm (MH) :

Given a *proposition kernel* $\tilde{\mathrm{K}}$, at iteration n :
1. Generate $\tilde{X}_{n+1} \sim \tilde{\mathrm{K}}(\cdot \,|\, X_n)$, $U_n \sim U([0,1])$, compute $\alpha_n(\tilde{X}_{n+1}, X_n)$.
2. Accept or Reject, set $X_{n+1} = \tilde{X}_{n+1}$ if $U_n \leq \alpha_n$, otherwise set $X_{n+1} = X_n$.

Hamiltonian Monte Carlo (HMC) :
[Duane and al, 1987]

Given the the target **potential** $U = -\log \pi$ and its gradient ∇U,
a *stepsize* h and a *number of steps* T,
define $\tilde{\mathrm{K}}_{h,T}$ by integrating a system of
Hamiltonian equations using the **leapfrog** integrator.

State of the art
No U-Turn Sampler (NUTS) :
[Hoffman and al, 2011]

Given $(U, \nabla U)$ and a stepsize h, define $\mathrm{K}_h^{\mathrm{U}}$ the NUTS kernel .

Introduction

NUTS is used in **PyMC3**, **Stan** and **Turing**, widely used software for Bayesian computational statistics.

Other libraries use Gibbs sampling for its flexibility. (BUGS, JAGS)

Introduction

NUTS is used in **PyMC3**, **Stan** and **Turing**, widely used software for Bayesian computational statistics.

Other libraries use Gibbs sampling for its flexibility. (BUGS, JAGS)

OK, in practice, everyone use **NUTS** when it is possible, but why does it work well ?

Litterature on the qualitative properties of HMC and NUTS.

Qualitative property	π-invariance $\int K(x,.)d\pi = \pi$	Ergodicity $\lim_{n\to\infty} \|\|K^n(x,\cdot) - \pi\|\|_{TV} = 0,$	Geometric ergodicity $\|\|K^n(x,\cdot) - \pi\|\|_{\mathscr{V}} = O(\rho^n), \quad \rho < 1$
HMC T is fixed	✓ [Duane and al, 1987]	✓ [Durmus and al, 2017]	✓ [Durmus and al, 2017]
NUTS T varies	✓ ↻ Appendix of [Betancourt, 2017] Not reviewed	?	?

Our contributions on the qualitative properties of HMC and NUTS.

Qualitative property	π-invariance $\int \mathrm{K}(x,.)\mathrm{d}\pi = \pi$	Ergodicity $\lim_{n\to\infty} \|\|\mathrm{K}^n(x,\cdot) - \pi\|\|_{TV} = 0,$	Geometric ergodicity $\|\|\mathrm{K}^n(x,\cdot) - \pi\|\|_{\mathscr{V}} = O(\rho^n), \quad \rho < 1$
HMC T is fixed	✓ [Duane and al, 1987]	NEW ! ✓ **Without bounding the stepsize** [Durmus and al, 2023]	✓ [Durmus and al, 2017]
NUTS T varies	✓ Appendix of [Betancourt, 2017] Not reviewed NEW ! + **give a proof with a general formalism** [Durmus and al, 2023]	NEW ! ✓ **By bounding the stepsize h and the HMC's assumptions on U, or without bounding the stepsize with more stringent regularity conditions on U.** [Durmus and al, 2023]	NEW ! ✓ **Conditions of the ergodicity + Conditions of HMC geometric ergodicity** [Durmus and al, 2023]

Why this paper was not done before ?

1

The study of NUTS is highly technical.
We introduce a general formalism and explicit expressions.

2

NUTS relies on a stopping time $(q_0, p_0) \in (\mathbb{R}^d)^2 \mapsto S(a, q_0, p_0)$.
Its regularity is hard to analyze.

3

Theoretical properties are not very attractive.
We try to stick to the practical situation framework.

Introduction

1. Motivations and contributions.

2. Hamiltonian Monte Carlo (HMC) and the NUTS intuition.

3. Dynamic HMC algorithms.

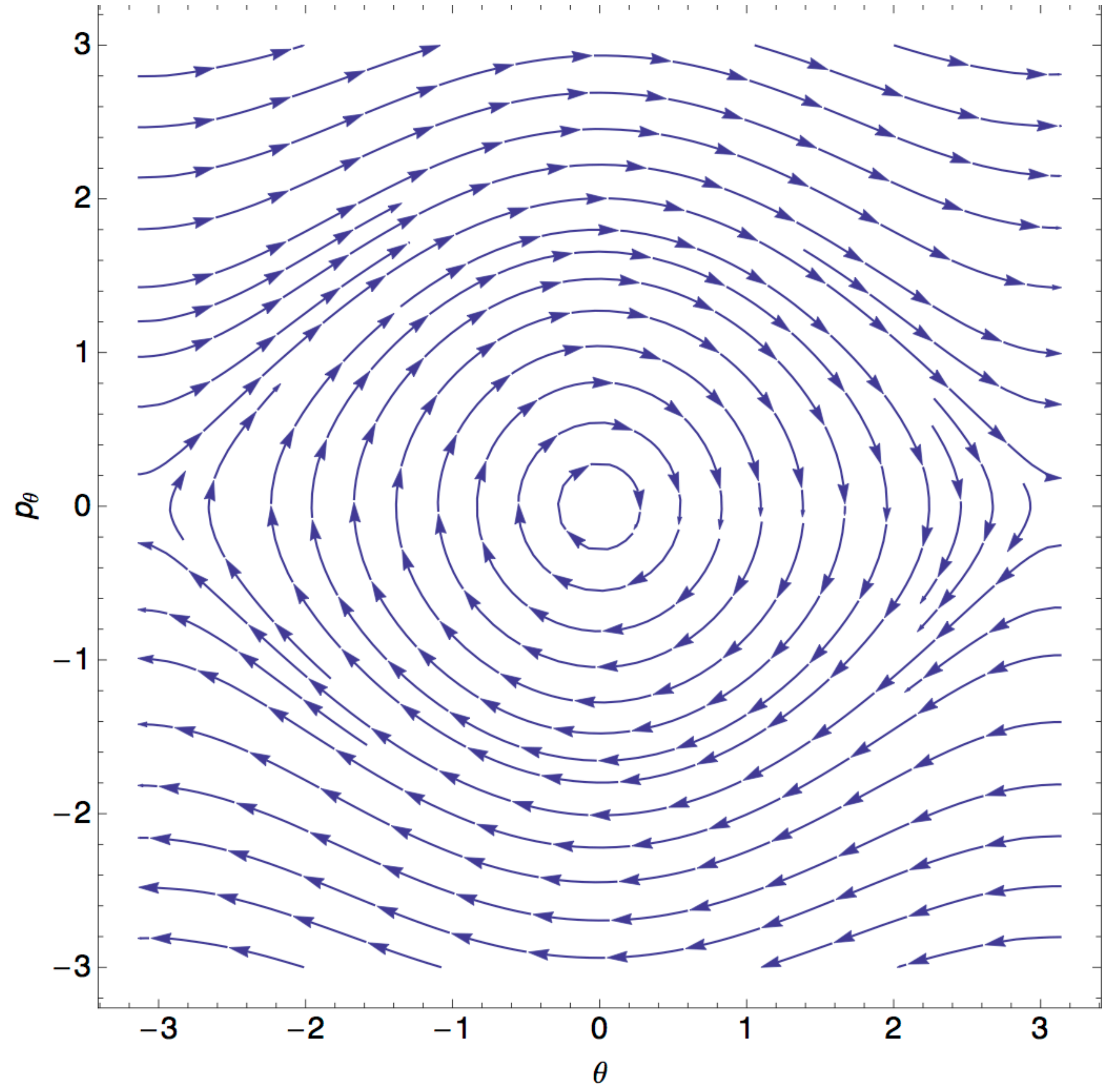
Hamiltonian Monte Carlo

Define **potential energy** $U(q) = -\log(\pi(q))$ and **Hamiltonian** $H(q,p) = U(q) + p^\top p/2$ for any $(q,p) \in (\mathbb{R}^d)^2$.

Hamiltonian dynamics $(q(t), p(t)) \in (\mathbb{R}^d)^2$, for any $t \geq 0$.

$$\frac{dq(t)}{dt} = \nabla_p H(q(t), p(t)) = p(t)$$

$$\frac{dp(t)}{dt} = -\nabla_q H(q(t), p(t)) = -\nabla U(q(t))$$



Hamiltonian dynamics related to a pendulum.

HMC algorithm (h, T)

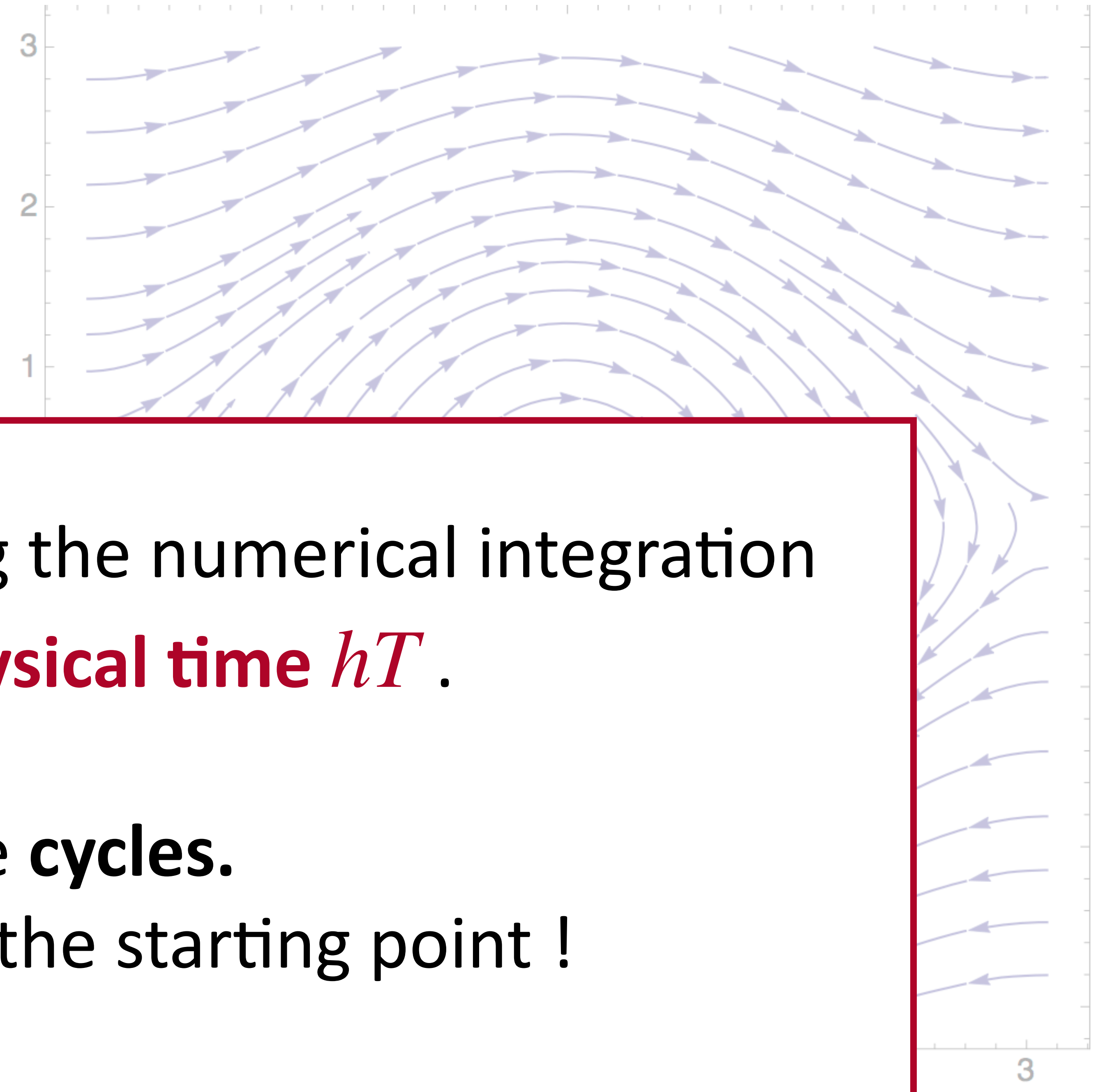
At iteration t, Markov chain at state X_t :

1. Sample $P_0 \sim \mathcal{N}(0_d, I_d)$ and set $\big(q(0), p(0)\big) = \big(X_t, P_0\big)$
2. Solve dynamics over time lengths T with the **leapfrog integrator using** (h, T) to get $\Phi_h^{(T)}(X_t, P_0) = \big(q_T, p_T\big)$.

(MH) 3. Sample $U^* \sim \mathcal{U}([0,1])$. If $U^* \leq \min\left\{1, \exp\left[H(q_0, p_0) - H(q_T, p_T)\right]\right\}$, set $X_{t+1} = q_T$, otherwise set $X_{t+1} = X_t$.

Hamiltonian Monte Carlo

Define **potential energy** $U(q) = -\log(\pi(q))$ and **Hamiltonian** $H(q,p) = U(q) + p^\top p/2$ for any $(q,p) \in (\mathbb{R}^d)^2$.



Hamiltonian dy

$\frac{dq(t)}{dt} =$

$\frac{dp(t)}{dt} =$

The **sampler efficiency** related to **the energy loss** during the numerical integration $\left[H(q_0,p_0) - H(q_T,p_T)\right]$ depends on the **physical time** hT .

The Hamiltonian dynamics can create **cycles.**
If T is **too large**, the proposition can be close to the starting point !

How to select h, T ?

HMC algorithm

pendulum.

At iteration t,

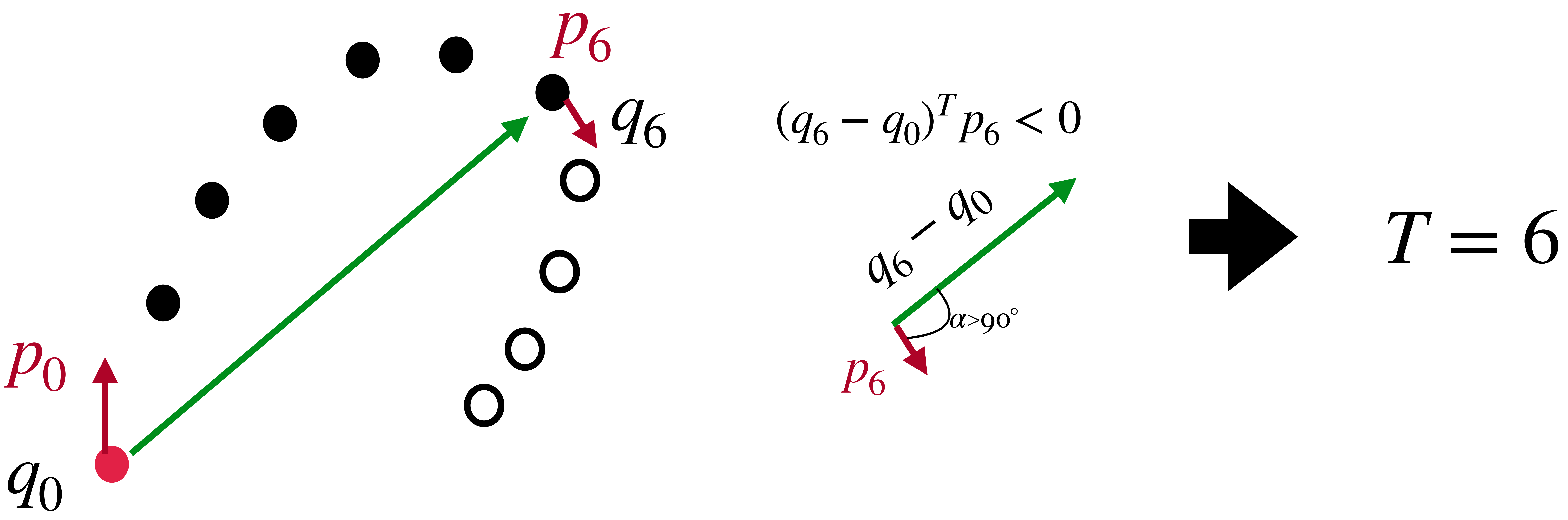
1. Sample $P_0 \sim \mathcal{N}(0_d, I_d)$ and set $\left(q(0), p(0)\right) = \left(X_t, P_0\right)$
2. Solve dynamics over time lengths T with the **leapfrog integrator using** (h, T) to get $\Phi_h^{(T)}(X_t, P_0) = \left(q_T, p_T\right)$.

(MH) 3. Sample $U^* \sim \mathcal{U}([0,1])$. If $U^* \leq \min\left\{1, \exp\left[H(q_0,p_0) - H(q_T,p_T)\right]\right\}$, set $X_{t+1} = q_T$, otherwise set $X_{t+1} = X_t$.

The intuition behind the No U-Turn Sampler.

No U-turn criteria between T_1 and T_2, by denoting $\Phi_h^{(j)}(q_0, p_0) = (q_j, p_j)$ for any $(q_0, p_0) \in (\mathbb{R}^d)^2$,

$F_{q_0}^{T_1,T_2}(p_0) = (q_{T_2} - q_{T_1})^\top p_{T_1} < 0$ or $(q_{T_2} - q_{T_1})^\top p_{T_2} < 0$.

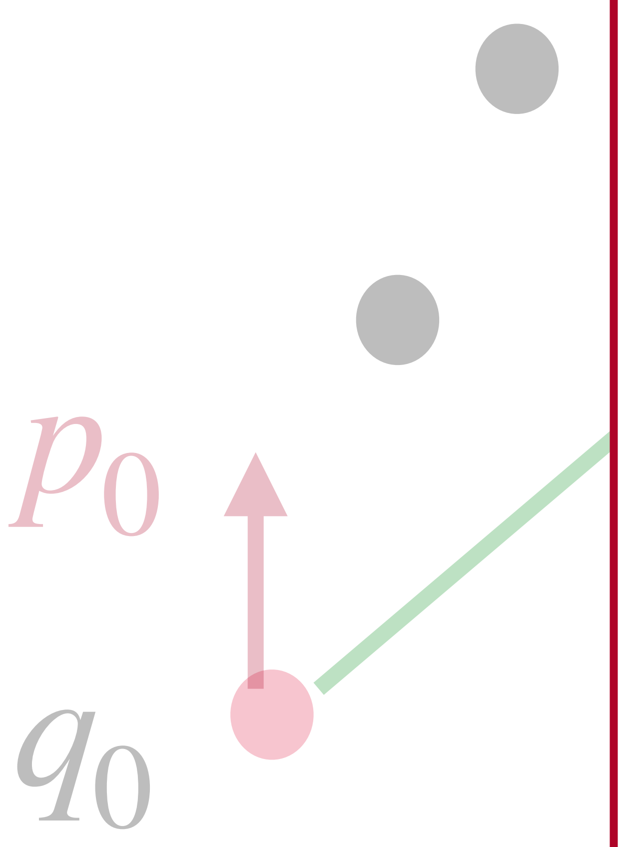


We can not just take the last point p_6 before the U-turn to have the target invariance !

The intuition behind the No U-Turn Sampler.

No U-turn criteria between T_1 and T_2, by denoting $\Phi^{(j)}(q_0, p_0) = (q_j, p_j)$ for any $(q_0, p_0) \in (\mathbb{R}^d)^2$,

$F_{q_0}^{T_1,T_2}(p_0) = (q_{T_2} -$



We can not just tak

Stopping time regularity condition :

For any $q \in \mathbb{R}^d$ the following set is dense,

$$\mathsf{F}_{q,-0} = \{p \in \mathbb{R}^d : F_q^{T_1,T_2}(p) \neq 0, T_1, T_2 \in [-2^{K_m} + 1 : 2^{K_m} - 1]^2, T_1 \neq T_2\}$$

A more « human » condition to satisfy the previous one :

∇U is L-lypschitz and the stepsize is bounded by $C/(L2^{K_m})$ with $C > 0$.

or

π is gaussian and the stepsize is in $\mathbb{R}_+^* \backslash \mathscr{H}$ with $\mathscr{H}$ countable.

or

U is real-analytic and $\lim_{|q| \to \infty} |\nabla^2 U(q)| = 0$

Introduction

1. Motivations and contributions.

2. Hamiltonian Monte Carlo (HMC) and the NUTS intuition.

3. Dynamic HMC algorithms.

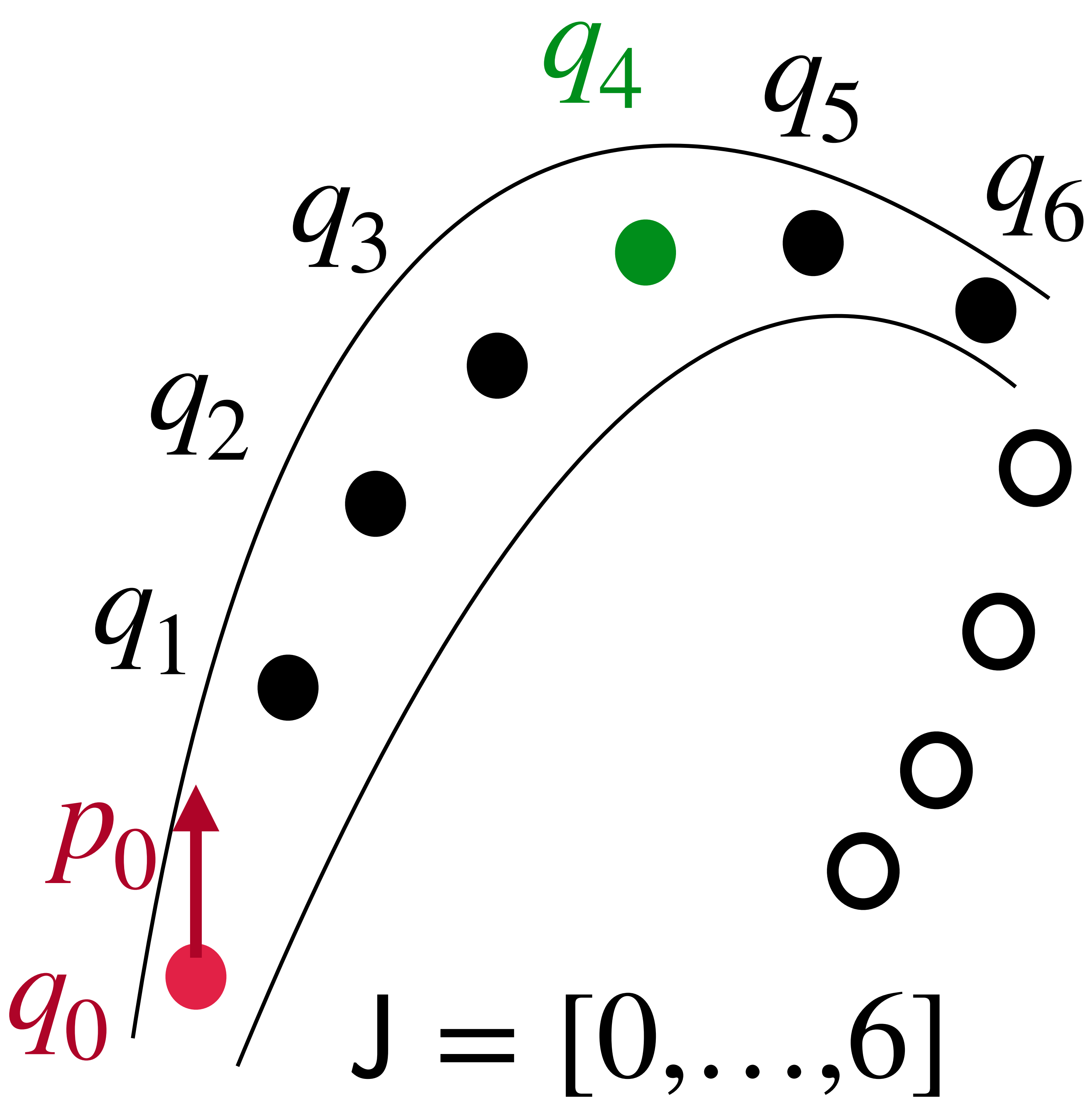
Introduction of Dynamic HMC kernels

1. Let $h > 0$, $K_m \in \mathbb{N}$ and let an **orbit selection** kernel be a family of probability distributions on $\mathscr{P}([-2^{K_m}, \ldots, 2^{K_m}])$

$$\{\mathrm{P}_h(\cdot \mid q_0, p_0) : (q_0, p_0) \in (\mathbb{R}^d)^2\}$$

2. Let an **index selection** kernel be a family of probability distributions on $[-2^{K_m}, \ldots, 2^{K_m}]$

$$\{\mathrm{Q}_h(\cdot \mid \mathrm{J}, q_0, p_0) : \mathrm{J} \subset [-2^{K_m}, 2^{K_m}], (q_0, p_0) \in (\mathbb{R}^d)^2\}$$



Dynamic HMC algorithm

Define the Dynamic HMC $(\mathrm{P}_h, \mathrm{Q}_h)$ *as the Markov chain* $(Q_k)_{k\in\mathbb{N}}$ *defined by the following steps that define* Q_{k+1} *given* Q_k *:*

1. *Sample* $P_{k+1} \sim \mathcal{N}(0_d, I_d)$
2. *Sample* $\mathtt{I}_{k+1}$ *with distribution* $\mathrm{P}_h(\,\cdot\,|\,Q_k, P_{k+1})$
3. *Sample* J_{k+1} *with distribution* $\mathrm{Q}_h(\,\cdot\,|\,\mathtt{I}_{k+1}, Q_k, P_{k+1})$
4. *Set* $Q_{k+1} = \mathrm{proj}_1\left\{\Phi_h^{J_{k+1}}(Q_k, P_{k+1})\right\}$, *where*

$\mathrm{proj}_1 : (x, y) \in (\mathbb{R}^d)^2 \mapsto x \in \mathbb{R}^d$

HMC case :

$$\mathrm{P}_h^{\mathrm{HMC}}(\{0,T\} \mid q_0, p_0) = 1$$

$$\mathrm{Q}_h^{\mathrm{HMC}}(\,\cdot\,|\,\{0,T\}, q_0, p_0) = \left(1 \wedge \frac{\tilde{\pi}(\Phi_h^{(T)}(q_0, p_0))}{\tilde{\pi}(q_0, p_0)}\right)\delta_T(\,\cdot\,) + \left(1 - 1 \wedge \frac{\tilde{\pi}(\Phi_h^{(T)}(q_0, p_0))}{\tilde{\pi}(q_0, p_0)}\right)\delta_0(\,\cdot\,)$$

where we denote by $\tilde{\pi}(q, p) \propto \exp(-H(q, p)) = \pi(q) \times \exp(-p^\top p/2)$

Dynamic HMC properties

General expression of the Dynamic HMC kernel for any $\mathsf{A} \in \mathscr{B}(\mathbb{R}^d)$:

$$\mathrm{K}_h(q_0, \mathsf{A}) = \int \mathcal{N}(p; 0_d, I_d)(p_0)\tilde{\mathrm{K}}_h((q_0, p_0), \mathsf{A})\mathrm{d}\, p_0$$

$$\tilde{\mathrm{K}}_h((q_0, p_0), \mathsf{A}) = \sum_{\mathsf{J}\subset\mathbb{Z}} \sum_{j\in\mathsf{J}} \mathrm{P}_h(\mathsf{J}\,|\,q_0, p_0)\mathrm{Q}_h(j\,|\,\mathsf{J}, q_0, p_0)\delta_{\mathrm{proj}_1(\Phi_h^{(j)}(q_0,p_0))}(\mathsf{A})$$

It is **not** a trivial extension of the HMC case $\mathrm{K}_h(q_0, \mathsf{A}) \neq \sum_{j\in\mathbb{Z}} \omega_j(q_0)\, \mathrm{K}_{h,j}^{\mathrm{HMC}}(q_0, \mathsf{A})$

Proposition :

Assume that $(\mathrm{P}_h, \mathrm{Q}_h)$ satisfy the **following equation** for any $(q_0, p_0) \in (\mathbb{R}^d)^2, \mathsf{J} \subset \mathbb{Z}$:

$$\tilde{\pi}(q_0, p_0)\mathrm{P}_h\left(\mathsf{J}\,|\,q_0, p_0\right) = \sum_{j\in\mathbb{Z}} 1_{\mathsf{J}}(0)\tilde{\pi}\left(\Phi_h^{(-j)}(q_0, p_0)\right)\mathrm{P}_h\left(\mathsf{J}+j\,|\,\Phi_h^{(-j)}(q_0, p_0)\right)\mathrm{Q}_h\left(j\,|\,\mathsf{J}+j, \Phi_h^{(-j)}(q_0, p_0)\right)$$

Then, K_h leaves the target measure π invariant

Dynamic HMC properties

General expression of the Dynamic HMC kernel for any $A \in \mathscr{B}(\mathbb{R}^d)$:

Assume,

For any $(q_0, p_0) \in (\mathbb{R}^d)^2$, $\mathsf{J} \subset \mathbb{Z}$ and $-j \in \mathsf{J}$, $\mathrm{P}_h\left(\mathsf{J} + j \,|\, \Phi_h^{(-j)}(q_0, p_0)\right) = \mathrm{P}_h\left(\mathsf{J} \,|\, q_0, p_0\right)$,

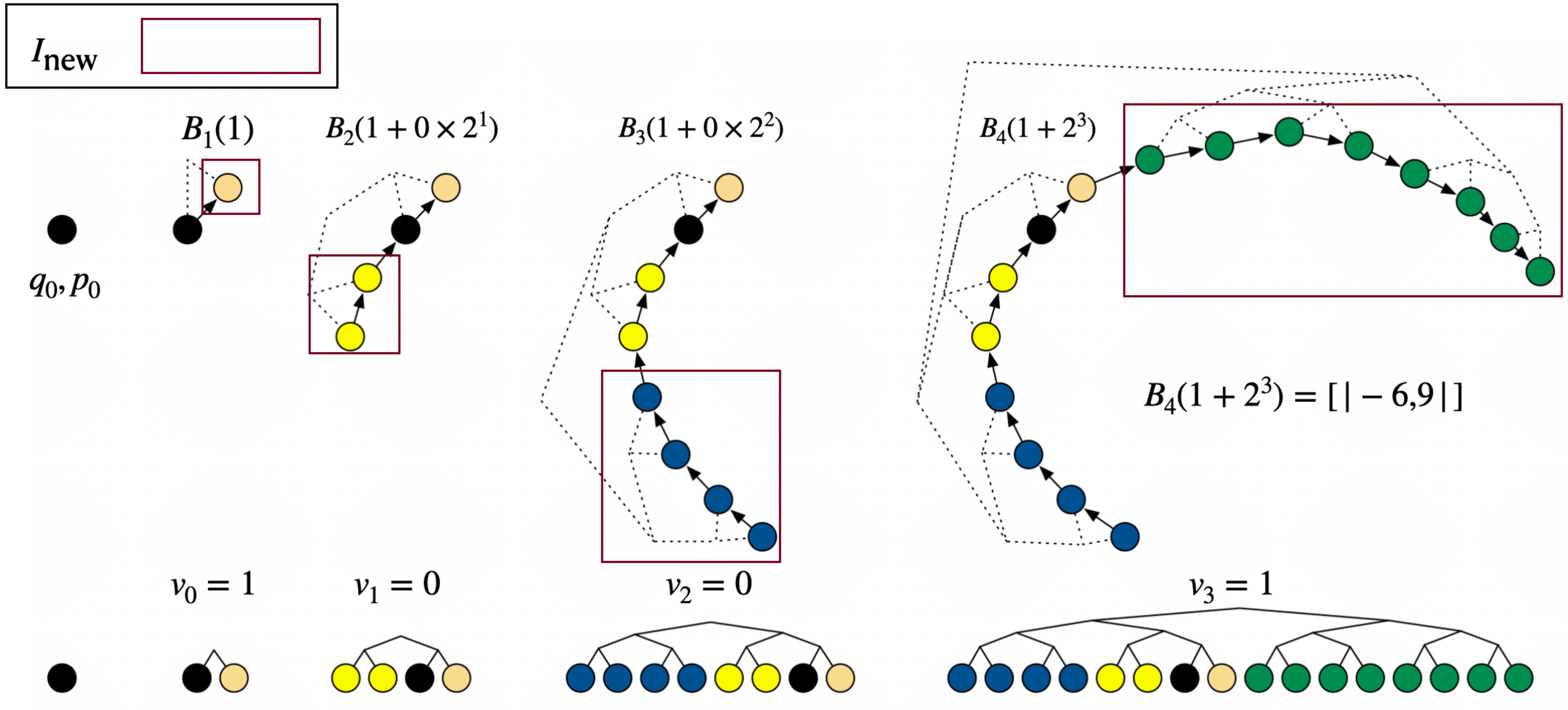
Then, the invariant condition reduces to

$$\tilde{\pi}(q_0, p_0) = \sum_{j \in \mathbb{Z}} 1_{\mathsf{J}}(0) \tilde{\pi}\left(\Phi_h^{(-j)}(q_0, p_0)\right) \mathrm{Q}_h\left(j \,|\, \mathsf{J} + j, \Phi_h^{(-j)}(q_0, p_0)\right)$$

$$\tilde{\pi}(q_0, p_0)\mathrm{P}_h\left(\mathsf{J} \,|\, q_0, p_0\right) = \sum_{j \in \mathbb{Z}} 1_{\mathsf{J}}(0) \tilde{\pi}\left(\Phi_h^{(-j)}(q_0, p_0)\right) \mathrm{P}_h\left(\mathsf{J} + j \,|\, \Phi_h^{(-j)}(q_0, p_0)\right) \mathrm{Q}_h\left(j \,|\, \mathsf{J} + j, \Phi_h^{(-j)}(q_0, p_0)\right)$$

Then, K_h leaves the target measure π invariant

NUTS' orbit selection kernel p_h



Binary tree enable fast practical recursive implementation.

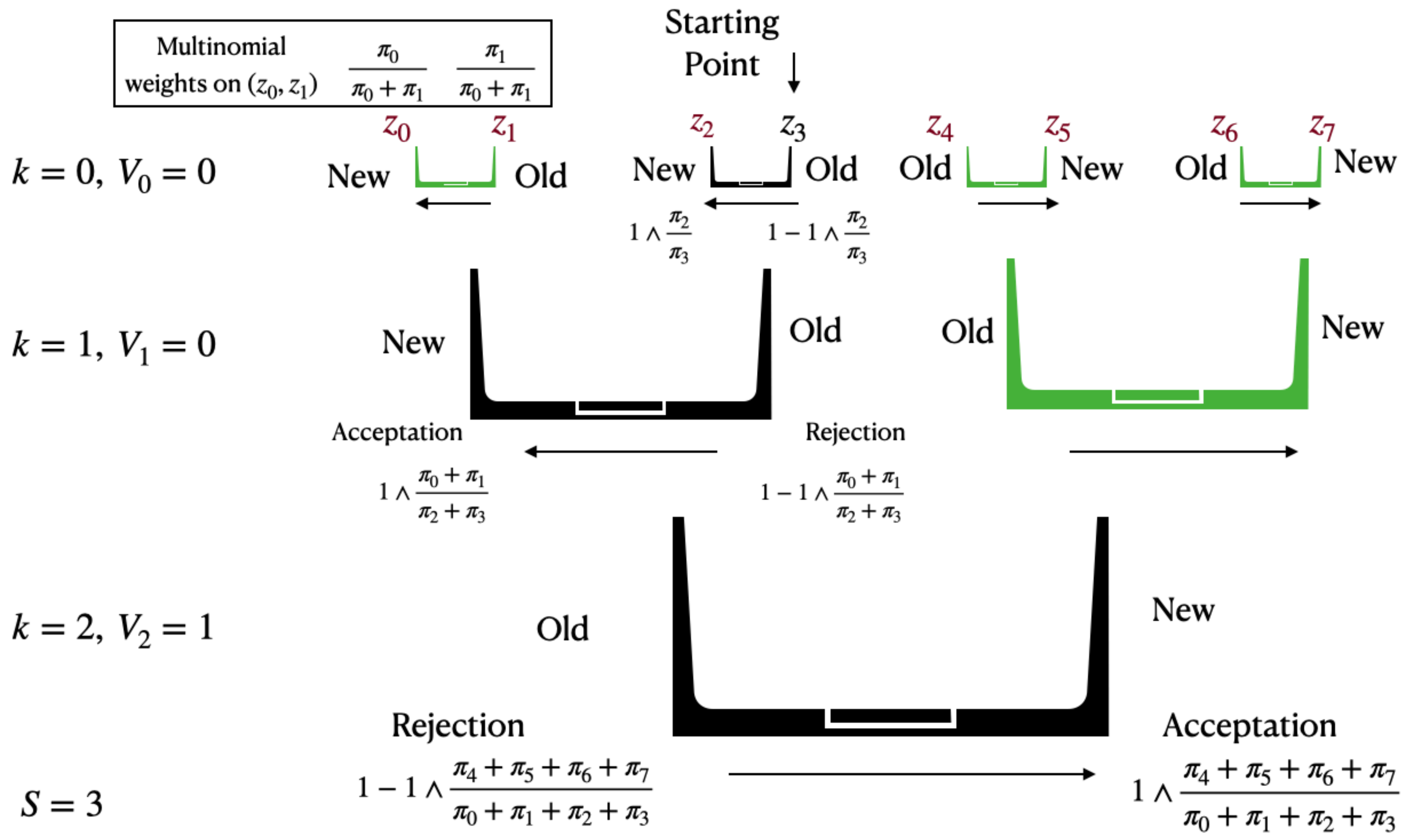
Scheme of the construction of the index set I_f in the Algorithm 1 presented in [Durmus and al, 2023].

Symmetry property : For any $(q_0, p_0) \in (\mathbb{R}^d)^2$, $\mathsf{J} \subset \mathbb{Z}$ and $-j \in \mathsf{J}$, $\mathrm{P}_h\left(\mathsf{J} + j \,|\, \Phi_h^{(-j)}(q_0, p_0)\right) = \mathrm{P}_h\left(\mathsf{J} \,|\, q_0, p_0\right)$

Explicit expression of p_h in the paper.

Thank you again !

NUTS selection kernel q_h .



$$\mathrm{I} = [|0, 2^3-1|] \quad z_0 = q_0, p_0\ ,\ z_i = \Phi_h^i(z_0), i \in \mathrm{I},\ \mathcal{O}_{\mathrm{I}}(z_3) = \{z_0, \ldots, z_7\},\ \pi_i = \tilde{\pi}(z_i)$$

$$\bar{q}_h(3,0\,|\,\mathrm{I}, z_0) = \left(1-1 \wedge \frac{\pi_4+\pi_5+\pi_6+\pi_7}{\pi_0+\pi_1+\pi_2+\pi_3}\right)\left(1 \wedge \frac{\pi_0+\pi_1}{\pi_2+\pi_3}\right)\frac{\pi_0}{\pi_0+\pi_1} \qquad \bar{q}_h(3,3\,|\,\mathrm{I}, z_0) = \left(1-1 \wedge \frac{\pi_4+\pi_5+\pi_6+\pi_7}{\pi_0+\pi_1+\pi_2+\pi_3}\right)\left(1-1 \wedge \frac{\pi_0+\pi_1}{\pi_2+\pi_3}\right)\left(1-1 \wedge \frac{\pi_2}{\pi_3}\right)$$

Litterature on the qualitative properties of HMC and NUTS.

Qualitative property	π-invariance $\int \mathrm{K}(x,.)\mathrm{d}\pi = \pi$	Ergodicity $\lim_{n\to\infty} \|\|\mathrm{K}^n(x,\cdot) - \pi\|\|_{TV} = 0,$	Geometric ergodicity $\|\|\mathrm{K}^n(x,\cdot) - \pi\|\|_{\mathscr{V}} = O(\rho^n), \quad \rho < 1$
HMC	By construction, with the MH mechanism. [Duane and al, 1987]	By assuming U continuously differentiable , ∇U lipschitz and by bounding the stepsize h. [Durmus and al, 2017]	By assuming U to be a gaussian perturbed outside of a compact and by bounding the stepsize h. [Durmus and al, 2017]
NUTS	Less trivial to check, performed in the Appendix of [Betancourt, 2017] Not reviewed	?	?

Our contributions on the qualitative properties of HMC and NUTS.

Qualitative property	π-invariance $\int K(x,.)d\pi = \pi$	Ergodicity $\lim_{n\to\infty} \|\|K^n(x,\cdot) - \pi\|\|_{TV} = 0,$	Geometric ergodicity $\|\|K^n(x,\cdot) - \pi\|\|_{\mathscr{V}} = O(\rho^n), \quad \rho < 1$
HMC	By construction, with the MH mechanism. [Duane and al, 1987]	NEW ! **Without bounding the stepsize, by assuming U to be a gaussian perturbated outside of a compact.** [Durmus and al, 2023]	By assuming U to be a gaussian perturbed outside of a compact and by bounding the stepsize h. [Durmus and al, 2017]
NUTS	Less trivial to check, performed in the Appendix of [Betancourt, 2017] Not reviewed NEW ! + **rewrite a proof with a general formalism** [Durmus and al, 2023]	NEW ! **By bounding the stepsize h and the HMC's assumptions on U, or without bounding the stepsize with more stringent regularity conditions on U.** [Durmus and al, 2023]	NEW ! **Conditions of the ergodicity + Conditions of HMC geometric ergodicity** [Durmus and al, 2023]

Hamiltonian Monte Carlo

HMC algorithm (h, T)

At iteration t, Markov chain at state X_t :

1. Sample $p_0 \sim \mathcal{N}(0_d, I_d)$ and set $q_0 = X_t$
2. **Leapfrog integrator**: define for any $l = 0,\ldots, T-1$,

   $$\Phi_h^{(1)} = (\Psi_{h/2}^{(1)} \circ \Psi_h^{(2)} \circ \Psi_{h/2}^{(1)}), \quad \Phi_h^{(l+1)} = \Phi_h^{(1)} \circ \Phi_h^{(l)},$$

   $$\Psi_t^{(1)}(q,p) = (q, p - t\nabla U(q)),$$

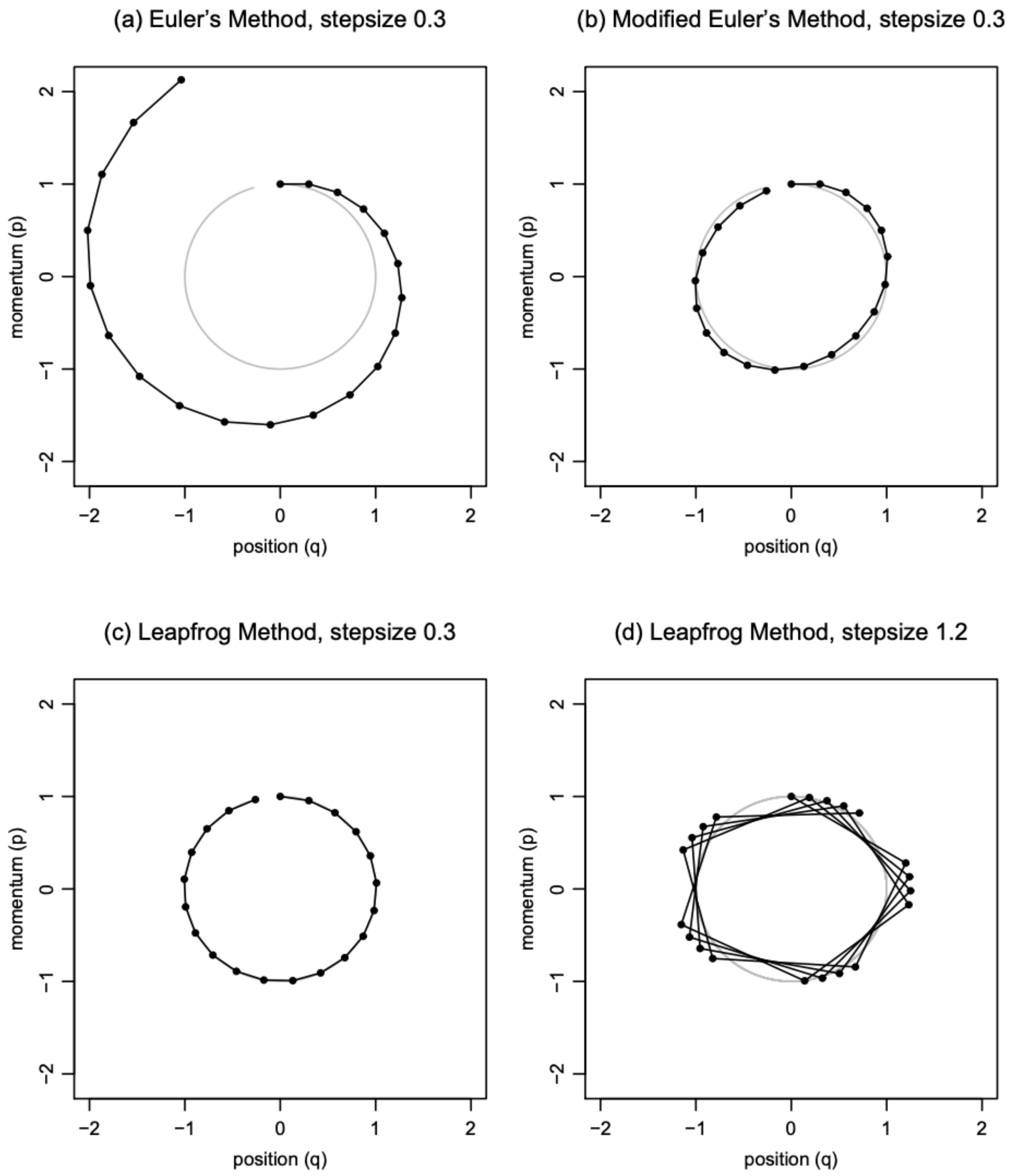
   $$\Psi_t^{(2)}(q,p) = (q + tp, p),$$

   for any $(q,p) \in (\mathbb{R}^d)^2$ and $t \geq 0$.

   Then, set $(q_T, p_T) = \Phi_h^{(T)}(q_0, p_0)$.
3. Sample $U^* \sim \mathcal{U}([0,1])$

   If $U^* \leq \min\left\{1, \exp\left[H(q_0, p_0) - H(q_T, p_T)\right]\right\}$

   Set $X_{t+1} = q_T$, otherwise set $X_{t+1} = X_t$.



Comparison of the Euler types and Leapfrog methods on the Gaussian case.