BPS and Randomised HMC

George Deligiannidis,

With D. Paulin, A. Bouchard-Côté and A. Doucet

Department of Statistics, Oxford University

September 21, 2018

Set up

- Consider a probability distribution on $\mathbb{R}^d$ of density

$$\pi(x) = \frac{\exp(-U(x))}{Z}$$

where $U : \mathbb{R}^d \to \mathbb{R}$ and $Z = \int \exp(-U(x))\,\mathrm{d}x$ cannot be computed in closed form.

- We are interested in computing expectations w.r.t. to π.

- Generic problem appearing in statistics, machine learning, computational physics and chemistry and theoretical computer science.

- MCMC is one of the most successful methods for doing this and comes with many guarantees.

Piecewise Deterministic Markov Processes

Piecewise deterministic Markov processes(PDMPs) allow for the construction of generic **non-reversible** MCMC algorithms:

- these emerged recently in physics (Peters & De With, 2012; Krauth et al., 2009, 2015, 2016, 2017; Hukushima et al., 2016):
- **state-of-the-art performance for a wide range of large scale physical models**.
- they are generic enough to sample from any target in continuous space.
- they go back to the *telegraph process* of the 50's.

Intro to PDMPs

A piecewise deterministic Markov process is defined from the following ingredients:

(1) **Deterministic dynamics**: an ODE such that

$$\frac{\mathrm{d}z_t}{\mathrm{d}t} = \phi(z_t),\ z_t = \begin{pmatrix} x_t \\ v_t \end{pmatrix} = \mathbf{\Phi}_t\left(z_0\right).$$

(2) **Event rate** $\lambda : \mathcal{Z} \to \mathbb{R}^+$ with $\lambda\left(z_t\right)\epsilon + o\left(\epsilon\right)$ being the probability of an event in $\left[t, t+\epsilon\right]$.

(3) **Markov kernel** Q such that at event time $z_t \sim Q\left(z_{t^-}, \cdot\right)$.

Examples of PDMPs

- The *Bouncy Particle Sampler*(BPS) introduced in Physics by Peters and de With [2012] and in comp. statistics by Bouchard-Côté, Vollmer, and Doucet [2015].
- The *Zig-Zag Sampler*: Fontbona et al. [2012]; Monmarché [2016]; Fontbona et al. [2016], Bierkens and Roberts [2017] and a general version for MCMC purposes appeared in Bierkens, Fearnhead, and Roberts [2016].
- Also more recently
  - ♣ the Stochastic BPS Pakman et al. [2016],
  - ♣ Generalized BPS Wu and Robert [2017],
  - ♣ Binary BPS Pakman [2017].

Bouncy particle sampler: ingredients I

Consider an extended target on $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}^d$

$$\rho(z) = \pi(x)\,\psi(v)\,,\ \ \psi(v) = \mathcal{N}(v; 0, I_d)$$

where v is the *velocity*.

a) **Deterministic dynamics**: an ODE such that

$$\frac{\mathrm{d}z_t}{\mathrm{d}t} = \begin{pmatrix} v_t \\ 0 \end{pmatrix},\ \ z_t = \begin{pmatrix} x_t \\ v_t \end{pmatrix} = \begin{pmatrix} x_0 + v_0 t \\ v_0 \end{pmatrix}.$$

b) **Event rate** $\lambda : \mathcal{Z} \to \mathbb{R}^+$ such that

$$\lambda(z) = \max\left(0, \langle \nabla U(x), v \rangle\right) := \langle \nabla U(x), v \rangle_+ \,.$$

Bouncy particle sampler: ingredients II

c) **Markov kernel** Q such that at event time $z_t \sim Q\left(z_{t^-}, \cdot\right)$:

$$Q\left(z, \mathrm{d}z'\right) = \delta_x\left(\mathrm{d}x'\right)\delta_{R_U(x)v}\left(\mathrm{d}v'\right),$$
$$R_U\left(x\right)v = v - 2\frac{\left\langle \nabla U\left(x\right), v\right\rangle}{\left\|\nabla U\left(x\right)\right\|^2}\nabla U\left(x\right).$$

$R_U\left(x\right)v$ corresponds to a reflection on the hyperplane orthogonal to the gradient of the potential.

- Possible to simulate exactly for a wide class of interesting problems.

Reducibility Issue

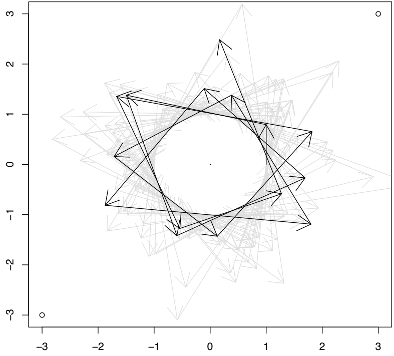


Figure: The BPS trajectory for an isotropic normal, it never enters the region around the mode if started tangent to a level set of the potential. Reproduced from Bouchard-Côté et al. [2015].

Refreshment and ergodicity

- To address reducibility: randomise the trajectory by refreshing the velocity, at a location dependent rate.
- Event rate becomes

$$\overline{\lambda}(z) = \lambda(z) + \lambda_{\text{ref}}(x)$$

- Transition kernel becomes

$$Q(z, \mathrm{d}z') = \delta_x(\mathrm{d}x') \left\{ \frac{\lambda(z)}{\overline{\lambda}(z)} \delta_{R(x)v}(\mathrm{d}v') + \frac{\lambda_{\text{ref}}(x)}{\overline{\lambda}(z)} \psi(\mathrm{d}v') \right\}$$

- Similar to momentum refreshment in HMC. Randomized bounces can alternatively be introduced, see Wu and Robert [2017].

Invariance and Ergodicity

- Generator

$$\mathcal{L}h(z) = \langle \nabla_x h(z), v \rangle + \overline{\lambda}(z) \int \{h(z') - h(z)\} Q(z, \mathrm{d}z')$$

- The BPS process is invariant w.r.t. ρ for any $\lambda_{\mathrm{ref}} \geq 0$ as

$$\int \rho(\mathrm{d}z) \, \mathcal{L}h(z) = 0$$

Proposition (Bouchard-Côté, Vollmer, and Doucet [2015])

BPS is ρ-invariant and ergodic.

- In D., Bouchard-Côté & Doucet '17 this was extended to the case $\lambda_{\mathrm{ref}}(\cdot)$ is a function of location, $\inf \lambda_{\mathrm{ref}}(\cdot) > 0$.

Theory for PDMPs

Is PDMCMC any good?

- We need guarantees of their convergence and CLTs for approximate confidence intervals.
- Central Limit Theorem:
  - ♣ non-reversible case more complicated than reversible;
  - ♣ many results rely on symmetrised generator being ergodic and satisfying some spectral condition: for BPS symmetrised generator is degenerate;
  - ♣ thus no spectral gap;
  - ♣ geometric ergodicity is an “obvious” route to CLT.

Geometric ergodicity

- we say $\{X_t\}$ is said to be V-uniformly ergodic, where $V \geq 1$ is called the *Lyapunov* function, if

$$\left\|P^t\left(x,\cdot\right)-\rho\right\|_V \leq DV\left(x\right)\alpha^t$$

for $D<\infty$, $\alpha<1$, and $t>0$, where $\|\mu\|_V = \sup_{|f|\leq V}\left|\mu\left(f\right)\right|$ is a norm on the space of signed measures.

Geometric Ergodicity

PMDPs have been around for some time but we have only recently started to understand their properties.

Geometric ergodicity results until recently include

- Mesquita and Hespanha [2010] for exponentially decaying targets;
- Monmarché [2016] for compact state spaces,
- Bierkens and Duncan [2017] for $d = 1$ under general conditions.

BPS is geometrically ergodic under certain assumptions on the tails and the curvature of the target in any dimension.

Zig-Zag was also recently shown to be irreducible and geometrically ergodic under general conditions by Bierkens, Roberts, and Zitt [2017]. See also Fétique [2017].

Assumptions

We consider BPS with velocities living in $\mathbb{S}^{d-1}$.

This simplifies calculations without modifying dynamics too much.

We henceforth assume that $U : \mathbb{R}^d \to [0, \infty)$ satisfies

$$\frac{\partial^2 U(x)}{\partial x_i \partial x_j} \text{ is locally Lipschitz continuous for all } i, j, \tag{A0}$$

$$\int_{\mathbb{R}^d} \bar{\pi}(\mathrm{d}x) |\nabla U(x)| < \infty, \tag{A1}$$

$$\varliminf_{|x| \to \infty} \frac{\mathrm{e}^{U(x)/2}}{\sqrt{|\nabla U(x)|}} > 0. \tag{A2}$$

Regular tails

For targets with tails that decay at least like an exponential and at most like a Gaussian:

Theorem ("Regular tails")

Let $\overline{\lambda}_{ref}(\cdot) = \lambda_{ref}$, *and assume* $\overline{\lim}_{|x|\to\infty} \|\Delta U(x)\| \leq \alpha_1 < \infty$. *If*

(A) $\underline{\lim}_{|x|\to\infty} |\nabla U(x)| = \infty$ *and* $\lambda_{ref} > (2\alpha_1 + 1)^2$, *OR*

(B) $\underline{\lim}_{|x|\to\infty} |\nabla U(x)| = 2\alpha_2 > 0$ *and* $\lambda_{ref} \leq \alpha_2 / 16\sqrt{d}$

then BPS is V*-uniformly ergodic with Lyapunov function*

$$V(z) = \overline{\lambda}(x, -v)^{-1/2} \exp\left(U(x)/2\right).$$

A few words about the Lyapunov function

The choice

$$V(z) = \overline{\lambda}(x, -v)^{-1/2} \exp(U(x)/2) \tag{1}$$

may appear random at first but is not.

Costa [1990], Costa and Dufour [2008] analyze the PDMP in terms of the embedded Markov chain that tracks the process just after event times.

This chain admits an invariant measure $\mu(\mathrm{d}z)$ with density

$$\mu(x, v) \propto \overline{\lambda}(x, -v)\, \mathrm{e}^{-U(x)}.$$

Thus $V(z) = \mu(z)^{-1/2}$.

"Thin tails"

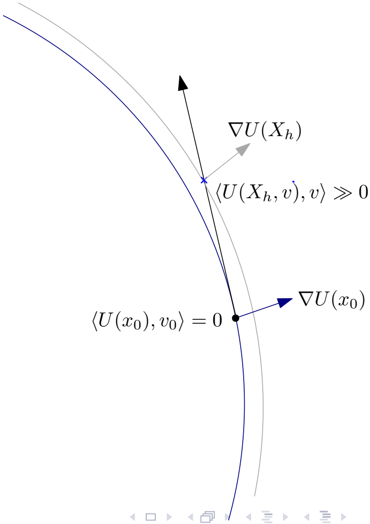
If tails decay faster than a Gaussian: gradient grows too fast in the tails.

Initialise tangent to the level-set.

Chance you refresh over next h time units $\propto \lambda_{\mathrm{ref}} h$, ie small.

But if gradient grows too fast, after h time units $\langle \nabla U(x_0 + vh), v_0) \rangle \gg 0$.

Chance of refreshment is tiny.

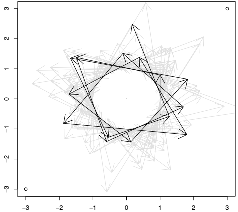


"Thin tails"

If tails decay faster than a Gaussian: gradient grows too fast in the tails.

With non-zero refreshment this is not a problem anymore.

But if gradient grows too fast, an easy informal calculation shows that with any constant refreshment rate will be dominated in the tails.



Not irreducible but these excursions outside a large ball longer and longer.

Thin tails also problematic for RWM, MALA and HMC.

"Thin tails"

Solution: scale refreshment with potential.

Theorem

Suppose that Assumptions (A0)-(A2) *hold. Let* $\lambda_{ref} > 0$ *and define for some* $\epsilon > 0$

$$\overline{\lambda}_{ref}(x) := \lambda_{ref} + \frac{|\nabla U(x)|}{\max\{1, |x|^{\epsilon}\}}. \tag{2}$$

Suppose that

$$\lim_{|x|\to\infty} \frac{|\nabla U(x)|}{|x|} = \infty, \qquad \lim_{|x|\to\infty} \frac{\|\Delta U(x)\|}{|\nabla U(x)|} |x|^{\epsilon} = 0.$$

Then BPS is V*-uniformly ergodic.*

Thick tails

For thick tailed targets, such as a multivariate t-distribution

$$\pi(x) \propto \left[1 + \frac{|x|^2}{k}\right]^{-\frac{k+d}{2}},$$

or *generalised Gaussians* of the form

$$\pi(x) \propto \exp\left\{-\left(1 + |x|^2\right)^{\frac{\beta}{2}}\right\}, \qquad \beta \in (0,1),$$

the problem is the opposite to that for thin tails.

The vanishing gradient provides no information at the tails allowing essentially random walk behaviour.

Thick tails: transformation

In these cases we follow the approach of Johnson and Geyer (AoS '12):

- Apply a diffeomorphism $h : \mathbb{R}^d \to \mathbb{R}^d$ such that the transformed target π_h satisfies any of the conditions above.
- Use BPS to sample the transformed target.
- Map trajectory back to the original parameterisation: this gives a geometrically ergodic scheme.

CLT

In all of these cases we get a CLT, but hard to estimate asymptotic variance.

More recent progress

More recent results on geometric ergodicity:

- Durmus, Guillin, and Monmarché [2018] extends and generalises geometric ergodicity for BPS; also introduces a coupling for the minorisation.
- Bierkens, Roberts, and Zitt [2017] proves Zig-Zag is ergodic without refreshment and geometrically ergodic under fairly natural set of assumptions.

All of these results are based on drift and minorisation arguments.

- The small set becomes too small in high-dimensions;
- Hard to control contraction in drift;
- Exponential rates deteriorate with dimension.
- Different methods are needed to understand the performance of these methods in high-dimensions.

Scaling with dimension

- Scaling limits introduced in MCMC by Roberts, Gelman, and Gilks [1997] to study RWM and to derive a theoretical "optimal" acceptance probability.
- this provides tuning guidelines.
- Extended in Roberts and Rosenthal [1998] to MALA and in Beskos et al. [2013] to HMC.

Scaling limits

- Let π_d be a sequence of target distributions of increasing dimension n.
- Common over-simplified scenario $\pi_d = \pi^{\otimes d}$.
- Let $\{X_t^{(d)}\}_t$ be the Markov chain(or process) generated by the algorithm of interest.
- Let $f_d : \mathbb{R}^d \to \mathrm{R}^k$ be an observable of interest.
- Very often $f_d(x_1, \ldots, x_d) = x_1$, the first component.

Scaling limits

- We then say $\{Y(t) : t \geq 0\}$ is a scaling limit if

$$\left\{f_d\left(X^{(d)}_{[c_d t]}\right); t \geq 0\right\} \Rightarrow \{Y(t); t \geq 0\},$$

  where time has been rescaled by the sequence c_d.

- For RWM, MALA and HMC $c_d = d$, the proposal variance to get a non-trivial scaling limit is $d^{-1}, d^{-1/3}, d^{-1/4}$ respectively.

- For RWM and MALA the limiting process is a Langevin (overdamped diffusion).

- This suggests that RWM, MALA and HMC scale like $d^2, d^{4/3}$ and $d^{5/4}$ respectively.

Scaling limits

- What can we say about BPS?
- Empirical evidence suggests $d^{3/2}$

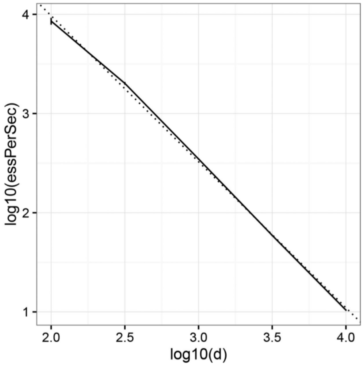


Figure: BPS on isotropic Gaussian, reproduced from Bouchard-Côté et al. [2015].

Scaling limits for PDMPs

- For a standard Gaussian target, several scaling limits for the *first coordinate*, the *angular momentum* and the *log-density* of BPS and Zig-Zag recently appeared in Bierkens, Kamatani, and Roberts [2018].
- The BPS process studied uses velocities in $\mathbb{S}^{d-1}$.
- For the first *location* component, with time rescaling $c_d = d$, scaling limit is Langevin diffusion.
- This suggests a cost of $O(d^2)$, since with unit speed number of events per unit time is $O(1)$.
- Zig-Zag, in terms of first-coordinate, is shown to scale like $O(d)$, which is not surprising as the algorithm factorises.
- We will focus on BPS but we will study a different regime, ultimately getting a different scaling limit.

Randomized HMC

Before stating our scaling limit let us introduce the process known as *Randomized HMC*, first studied in Bou-Rabee and Sanz-Serna [2017].

- Essentially a PDMP version of HMC.
- Let $T_1, T_2, \cdots$ be the arrival times of a homogeneous Poisson process with rate λ.
- For $0 \leq t < T_1$, $Y_t = (X_t, V_t)$ follows Hamiltonian dynamics w.r.t. $H(x, v) = U(x) + |v|^2$.
- Let $\xi_1, \xi_2, \cdots$ be i.i.d. Gaussians. At time T_1 we refresh the velocity by setting $V_{T_1} = \xi_1$ or

$$V_{T_1} = \alpha V_{T_1 -} + \sqrt{1 - \alpha^2}\xi_1.$$

Assumptions

We make the following assumptions to simplify calculations.

Assumption

The potential $U_d : \mathbb{R}^d \mapsto \mathbb{R}_+$ *takes the form*

$$U_d(\boldsymbol{x}) = U_d(x_1, \ldots, x_d) = \sum_{i=1}^{d} U(x_i). \tag{3}$$

Assumption

We have $U \in C^2(\mathbb{R})$, $|U(x)| \to \infty$ *as* $|x| \to \infty$, $\|U''\|_\infty \leq M < \infty$ *and*

$$\int \mathrm{e}^{-U(x)} |U'(x)|^2 \mathrm{d}x < \infty.$$

Scaling limit of BPS

We are looking at BPS with standard **Gaussian** velocities.

Theorem

Let the assumptions hold, $0 < \alpha < 1$ and $\mathbf{Z}_d(0) \sim \pi_d$. Then the process $\{Z_d^{(1)}(t) : t \geq 0\}$ corresponding to the ***first location and velocity components*** *of the BPS process converges weakly to the RHMC process.*

Important differences with Bierkens et al. [2018]:

- We consider standard Gaussian velocity (norm $\sim d^{1/2}$);
- we look at both location and velocity;
- we look at natural time scale.

Overall the effect of this is that the regime in Bierkens et al. [2018] corresponds to the *overdamped* regime of the Langevin equation.

Assumptions

- The i.i.d. assumption can be weakened significantly.
- What you need is that if $\boldsymbol{X}_d \sim \pi_d \propto \mathrm{e}^{-U(\cdot)}$ and $\boldsymbol{V}_d \sim \mathcal{N}(0, \mathbb{1}_d)$ are independent then

$$\left. \frac{\langle \nabla U(\boldsymbol{X}_d), \boldsymbol{V}_d \rangle}{\| \nabla U(\boldsymbol{X}) \|} \right| X_1, V_1 \Rightarrow \mathcal{N}(0,1),$$

  possible under weak dependence.
- Potential can grow polynomially.
- Can look at any finite number of coordinates.

So what?

- Need to think about what this says about BPS.
- We have **not** rescaled time.
- It can be shown that when the target has weak dependence the number of events for a unit of time scales like $d^{1/2}$, whereas each event roughly costs d to simulate.
- What about the limiting process?
- Usually scaling limit is Langevin diffusion, well understood, and serves as a benchmark.
- RHMC is actually a very natural PDMP, so it could potentially serve as a benchmark, but we need to understand its mixing properties.

Geometric ergodicity of RHMC

- Bou-Rabee and Sanz-Serna [2017] showed that RHMC is geometrically ergodic using drift and minorisation.
- Rates obtained in this fashion are usually not sharp in the dimension.
- Although geometrically ergodic, RHMC has **no spectral gap** as the symmetric part of its generator is degenerate.
- We opt instead for *hypo-coercivity*.

Hypocoercivity I

- Approach originates in Nier and Helffer [2005].
- Work with modified norm, e.g. Sobolev norm

$$\|f\|_{H^1}^2 := \|f\|^2 + \|\nabla_x f\|^2 + \|\nabla_v f\|^2.$$

- One then aims to obtain convergence rates of the form

$$\|P^t f\|_{H^1} \leq C\mathrm{e}^{-\mu t}\|f\|_{H^1}.$$

- Nier and Helffer [2005] and Villani [2009]study the Fokker-Planck equation.
- Regularisation results lead to L^2 bounds

$$\|P^t f\| \leq C\mathrm{e}^{-\mu t}\|f\|. \tag{4}$$

Hypocoercivity II

$$\|P^t f\| \leq C\mathrm{e}^{-\mu t}\|f\|. \tag{5}$$

- Notice that $C > 1$, since $C = 1$ would imply a spectral gap.
- This possibility is unique to non-reversible processes.
- In reversible case (5) with $C > 0$ automatically implies (5) with $C = 1$, see Hairer, Stuart, Vollmer, et al. [2014].

Hypocoercivity for RHMC I

Let $a, c > 0$, $b^2 < ac$ so that

$$\langle\langle h, h \rangle\rangle := a\|\nabla_v h\|^2 - 2b\langle \nabla_x h, \nabla_v h\rangle + c\|\nabla_x h\|^2, \tag{6}$$

defines a norm equivalent to $\|\cdot\|_{H^1}$ and let

$$Bf(x, v) := \langle \nabla_x f, v\rangle - \langle \nabla_v f, \nabla U\rangle.$$

The following is our main result which obtains dimension free convergence rates under the following common assumption

$$m\langle v, v\rangle \leq \langle v, \nabla^2 U(x) v\rangle \leq M\langle v, v\rangle. \tag{7}$$

Hypocoercivity for RHMC

Theorem

Let $f \in \mathcal{D}(B) \subset H^1(\pi) \subset L_0^2(\pi)$ *and*

$$\lambda_{ref} = \frac{1}{1-\alpha^2}\left(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}}\right),$$
$$\mu = \frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)}.$$

Then there are constants a, b, c *such that* $a > 0, c > 0, b^2 < ac$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle P^t f, P^t f\rangle\rangle \leq -\mu\langle\langle P^t f, P^t f\rangle\rangle. \tag{8}$$

In particular we have

$$\|P^t f\|^2 \leq \frac{\left(1+\frac{|b|}{\sqrt{ac}}\right)\max(a,c)}{\left(1-\frac{|b|}{\sqrt{ac}}\right)\min(a,cm)} \cdot \|f\|_{H^1(\pi)}^2 \mathrm{e}^{-\mu t}. \tag{9}$$

- Hypocoercivity has been used in this context before.
- Actually Dolbeault et al. [2015] treats RHMC and related processes using a different metric.
- Very recently the approach in Dolbeault, Mouhot, and Schmeiser [2015] was adapted to cover most known PDMCMC algorithms in Andrieu, Durmus, Nüsken, and Roussel [2018a].

Wasserstein distance I

We also have mixing in terms of the following Wasserstein distance:

$$
d_A^2(Z_1(t), Z_2(t)) := a\|X^{(2)}(t) - X^{(1)}(t)\|^2 \\
+ 2b\left\langle X^{(2)}(t) - X^{(1)}(t), V^{(2)}(t) - V^{(1)}(t)\right\rangle + c\|V^{(2)}(t) - V^{(1)}(t)\|^2
$$

where $Z_i = (X^{(i)}, V^{(i)})$ for $i = 1, 2$ are two copies of RHMC coupled with identical refreshment events.

The parameters a, b, c are chosen to make this a proper distance. Then

Wasserstein distance II

Theorem

Suppose that $0 \le \alpha < 1$, *our assumptions hold and let*

$$\lambda_{ref} = \frac{1}{1-\alpha^2}\left(2\sqrt{M+m} - \frac{(1-\alpha)m}{\sqrt{M+m}}\right),$$

$$\mu = \frac{(1+\alpha)m}{\sqrt{M+m}} - \frac{\alpha m^{3/2}}{2(M+m)}.$$

Then there exist constants $a, c > 0$ *and* $b^2 < ac$ *such that*

$$L_{1,2}\ d_A^2(Z_1(t), Z_2(t)) \le -\mu \cdot d_A^2(Z_1(t), Z_2(t)). \qquad (10)$$

Main References of This Talk I

C/ Andrieu, A. Durmus, N. Nüsken, and J. Roussel. Hypercoercivity of piecewise deterministic markov process-monte carlo. *arXiv preprint arXiv:1808.08592*, 2018a.

Christophe Andrieu, Alain Durmus, Nikolas Nüsken, and Julien Roussel. Hypercoercivity of piecewise deterministic markov process-monte carlo. *arXiv preprint arXiv:1808.08592*, 2018b.

A. Beskos, N. Pillai, G. Roberts, J.M. Sanz-Serna, and A. Stuart. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

J. Bierkens. Non-reversible metropolis-hastings. *Statistics and Computing*, 26(6):1213–1228, 2016.

J. Bierkens and A. Duncan. Limit theorems for the zig-zag process. *Advances in Applied Probability*, 49(3):791–825, 2017.

Main References of This Talk II

J. Bierkens and G. Roberts. A piecewise deterministic scaling limit of lifted metropolis–hastings in the curie–weiss model. *The Annals of Applied Probability*, 27(2):846–882, 2017.

J. Bierkens, P. Fearnhead, and G. Roberts. The zig-zag process and super-efficient sampling for bayesian analysis of big data. *arXiv preprint arXiv:1607.03188*, 2016.

J. Bierkens, G. Roberts, and P.A. Zitt. Ergodicity of the zigzag process. *arXiv preprint arXiv:1712.09875*, 2017.

Joris Bierkens, Kengo Kamatani, and Gareth O Roberts. High-dimensional scaling limits of piecewise deterministic sampling algorithms. *arXiv preprint arXiv:1807.11358*, 2018.

N. Bou-Rabee and J.M. Sanz-Serna. Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.*, 27(4):2159–2194, 2017.

Main References of This Talk III

A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The bouncy particle sampler: A non-reversible rejection-free markov chain monte carlo method. *To appear in JASA, arXiv:1510.02451*, 2015.

OLV Costa. Stationary distributions for piecewise-deterministic markov processes. *Journal of Applied Probability*, 27(1): 60–73, 1990.

Oswaldo Luiz V Costa and François Dufour. Stability and ergodicity of piecewise deterministic markov processes. *SIAM Journal on Control and Optimization*, 47(2):1053–1077, 2008.

G. D., A. Bouchard-Côté, and A. Doucet. Exponential ergodicity of the bouncy particle sampler. *arXiv preprint arXiv:1705.04579*, 2017.

Mark HA Davis. *Markov models and optimization*. Routledge, 1993.

Main References of This Talk IV

J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.*, 367(6):3807–3828, 2015.

Douglas Down, Sean P Meyn, and Richard L Tweedie. Exponential and uniform ergodicity of markov processes. *The Annals of Probability*, pages 1671–1691, 1995.

Andrew B Duncan, Tony Lelievre, and GA Pavliotis. Variance reduction using nonreversible langevin samplers. *Journal of statistical physics*, 163(3):457–491, 2016.

Alain Durmus, Arnaud Guillin, and Pierre Monmarché. Geometric ergodicity of the bouncy particle sampler. *arXiv preprint arXiv:1807.05401*, 2018.

N. Fétique. Long-time behaviour of generalised zig-zag process. *arXiv preprint arXiv:1710.01087*, 2017.

Main References of This Talk V

J. Fontbona, H. Guérin, and F. Malrieu. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Advances in Applied Probability*, 44(4): 977–994, 2012.

J. Fontbona, H. Guérin, and F. Malrieu. Long time behavior of telegraph processes under convex potentials. *Stochastic Processes and their Applications*, 126(10):3077–3101, 2016.

P.J. Green and A. Mira. Delayed rejection in reversible jump metropolis–hastings. *Biometrika*, 88(4):1035–1053, 2001.

Martin Hairer, Andrew M Stuart, Sebastian J Vollmer, et al. Spectral gaps for a metropolis–hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.

Leif T Johnson and Charles J Geyer. Variable transformation to obtain geometric ergodicity in the random-walk metropolis algorithm. *The Annals of Statistics*, pages 3050–3076, 2012.

Main References of This Talk VI

A. R. Mesquita and J. P. Hespanha. Construction of lyapunov functions for piecewise-deterministic markov processes. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2408–2413. IEEE, 2010.

P. Monmarché. Piecewise deterministic simulated annealing. *ALEA, Lat. Am. J. Proba. Math. Stat.*, 13(1):357–398, 2016.

R. M Neal. Slice sampling. *Ann. Statist.*, 31(3):705–741, 2003.

F. Nier and B. Helffer. *Hypoelliptic Estimates and Spectral Theory for Fokker-Planck Operators and Witten Laplacians.* Springer, 2005.

A. Pakman. Binary bouncy particle sampler. *arXiv preprint arXiv:1711.00922*, 2017.

A. Pakman, D. Gilboa, D. Carlson, and Liam P. Stochastic bouncy particle sampler. *arXiv preprint arXiv:1609.00770*, 2016.

E.A.J.F. Peters and G. de With. Rejection-free monte carlo sampling for general potentials. *Physical Review E*, 85(2): 026703, 2012.

Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081, 2015.

G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Statist. Soc. B*, 60(1):255–268, 1998.

G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

C. Sherlock and A. H Thiery. A discrete bouncy particle sampler. *arXiv preprint arXiv:1707.05200*, 2017.

Main References of This Talk VIII

Konstantin S Turitsyn, Michael Chertkov, and Marija Vucelja. Irreversible monte carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4-5):410–414, 2011.

P. Vanetti, A. Bouchard-Côté, G. D., and A. Doucet. Piecewise deterministic markov chain monte carlo. *arXiv preprint arXiv:1707.05296*, 2017.

C. Villani. *Hypocoercivity*. Number 949-951. American Mathematical Society, 2009.

C. Wu and C.P. Robert. Generalized bouncy particle sampler. *arXiv preprint arXiv:1706.04781*, 2017.