Scaling Limits of Markov Chains/Processes in Monte Carlo Methods

Kengo Kamatani (Institute of Statistical Mathematics, JST CREST)

In this presentation, we will explore the recent results of scaling limit of piecewise deterministic Markov processes for anisotropic targets. Suppose we wish to sample from

$$\Pi(\mathrm{d}x) = \exp(-\mathcal{H}(x))\mathrm{d}x$$

where $\mathcal{H} : \mathbb{R}^d \to \mathbb{R}$ is a continuously differentiable function. For the Bayesian context, this probability distribution is the posterior distribution of interest. If we have an i.i.d. sample from Π, we can approximate Π-integral of any function $f(x)$ by the law of large numbers. In most of the cases, direct i.i.d. sampling is impossible or computationally very expensive. For these cases, the Markov chain Monte Carlo method is useful which originated with the classic paper by Metropolis et al. (1953) almost 70 years ago. The Markov chain Monte Carlo method is designed to construct an ergodic Markov kernel P which is Π-invariant. If a Markov chain $X_1, X_2, \dots$ is generaetd from the Markov kernel P then the law of large numbers is satisfied. The Markov chain Monte Carlo is now a gold standard for Bayesian inference.

Recently, its continuous process version, the Markov **process** Monte Carlo method is of substantial interest for Monte Carlo analysis. Known Markov process Monte Carlo methods rely on an auxiliary variable trick which uses an auxiliary variable v with a probability density ν on Ξ and considers the joint probability distribution $\mu := \Pi(\mathrm{d}x) \otimes \nu(\mathrm{d}v)$ as an extended target distribution on $\mathcal{Z} = \mathbb{R}^d \times \Xi$. The original target distribution is a marginal distribution of the extended target distribution. Since Brownian motion does not have an absolutely continuous path, we can not simulate processes driven by Brownian motion exactly. For our Monte Carlo analysis, exact sampling is necessary. Therefore, the Markov processes of interest should not have a Brownian part. Known processes consist of a deterministic part and a pure jump part. These processes are known as the **piecewise deterministic Markov processes**.

Here we follow Azaïs et al. (2014) for the expression of the piecewise deterministic Markov processes. The processes are constructed by characteristics $(\phi, \lambda_k, Q_k : l = 1, \dots, K)$. The flow $\phi : \mathcal{Z} \times \mathbb{R} \to \mathcal{Z}$ is continuous, $\phi(\cdot, t)$ is a homeomorphism for each $t \in \mathbb{R}$ and $\phi(\phi(\cdot, s), t) = \phi(\cdot, s + t)$. For each $k = 1, \dots, K$, the jump rate $\lambda_k : \mathcal{Z} \to \mathbb{R}_+$ determines the jump time of pure jump processes, and Q_k is a Markov kernel on $\mathcal{Z}$. Let $\Lambda_k(z, t) = \int_0^t \lambda_k(\phi(z, s))\mathrm{d}s$.

The Markov process is defined by the following way. Suppose $z(0) = (x(0), t(0)) \in \mathcal{Z}$. Let $T_1, \dots, T_K$ be independent processes with $\mathbb{P}(T_k \geqslant t) = \exp(-\Lambda_k(z, t))$. Let $T_* = \min_{k=1,\dots,K} T_k$. If $T_k = T_*$, then Z is generated from $Q_k(\phi(z, T_*), \cdot)$ and set

$$X(t) = \begin{cases} \phi(z(0), t) & \text{for } t < T_* \\ Z & \text{for } t = T_*. \end{cases}$$

After T_*, the process evolves in the same way with starting value Z. There are several choices of characteristics. Two popular piecewise deterministic Markov processes use the same flow ϕ defined by $x'(t) = v(t)$ and $v'(t) = 0$. The **Zig-Zag sampler** proposed by Bierkens et al. (2019)

uses d Markov kernels $Q_1, \ldots, Q_d$ with d jump rates $\lambda_1, \ldots, \lambda_d$. For each $i = 1, \ldots, d$, the Markov kernel is a deterministic kernel Q_i defined by a map $(x, v) \mapsto (x, F_i(v))$ where F_i is an operator that flips the i-th coordinate of x. The jump rate is defined by $\lambda_i((x, v)) = \max\{0, \partial_i \mathcal{H}(x) v_i\}$.

The **bouncy particle sampler** proposed by Peters and de With (2012), Bouchard-Côté et al. (2018) uses two Markov kernels Q_{bounce} and Q_{ref} with corresponding jump rates λ_{bounce} and λ_{ref}. The kernel Q_{bounce} is a deterministic kernel defined by a map $(x, v) \mapsto (x, \kappa(x, v))$:

$$\kappa(x, v) = v - 2\frac{\langle \nabla \mathcal{H}(x), v \rangle}{\|\nabla \mathcal{H}(x)\|^2} \nabla \mathcal{H}(x)$$

and $\lambda_{\text{bounce}}(x, v) = \max\{0, \langle \nabla \mathcal{H}(x), v \rangle\}$. The jump rate λ_{ref} is a positive constant, and Q_{ref} is a μ-invariant Markov kernel. For our analysis, for simplicity, we assume $Q_{\text{ref}}((x, v), \mathrm{d}(y, w)) = \nu(\mathrm{d}w)$.

We have several critical findings. For the Zig-Zag algorithm, its performance is intricately linked to the orientation of the target's anisotropy; specific alignments with the algorithm's operational axes lead to enhanced efficiency, while others can hinder its effectiveness. The BPS algorithm, on the other hand, exhibits a deterministic dynamical behaviour in its limiting form with a better rate of convergence.

This is joint work with Joris Bierkens (TU Delft) and Gareth O. Roberts (Warwick). See our paper on arxiv `https://arxiv.org/abs/2305.00694` for the detail.

References

Christophe Andrieu, Alain Durmus, Nikolas Nüsken, and Julien Roussel. Hypocoercivity of piecewise deterministic markov process-monte carlo. *The Annals of Applied Probability*, 31, 10 2021. ISSN 1050-5164. doi: 10.1214/20-AAP1653.

Romain Azaïs, Jean-Baptiste Bardet, Alexandre Génadot, Nathalie Krell, and Pierre-André Zitt. Piecewise deterministic Markov process—recent results. In *Journées MAS 2012*, volume 44 of *ESAIM Proc.*, pages 276–290. EDP Sci., Les Ulis, 2014. doi: 10.1051/proc/201444017.

Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1715.

Joris Bierkens, Kengo Kamatani, and Gareth O. Roberts. High-dimensional scaling limits of piecewise deterministic sampling algorithms. *The Annals of Applied Probability*, 32, 10 2022. ISSN 1050-5164. doi: 10.1214/21-AAP1762.

Alexandre Bouchard-Côté, Sebastian J. Vollmer, and Arnaud Doucet. The bouncy particle sampler: A nonreversible rejection-free markov chain monte carlo method. *Journal of the American Statistical Association*, 113(522):855–867, 2018. doi: 10.1080/01621459.2017.1294075.

George Deligiannidis, Daniel Paulin, Alexandre Bouchard-Côté, and Arnaud Doucet. Randomized hamiltonian monte carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *The Annals of Applied Probability*, 31, 12 2021. ISSN 1050-5164. doi: 10.1214/20-AAP1659.

N. Metropolis, W. A. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

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