A Map of Hamiltonian Monte Carlo

竹田航太

2021 年 7 月 26 日

目次

概要

ハミルトニアンモンテカルロ (Hamiltonian Monte Carlo: HMC) についての入門・基礎

的な理論と最近の研究の現状，今後の方向性についてまとめる．

1 Foundations of Hamiltonian Monte Carlo

HMC の直感的なイントロダクション [1, Conceptual introduction of HMC, 2018]
HMC の数学的定式化 [2, Geometric Foundations of HMC, 2017]

1.1 Symplectic Geometory

HMC の定式化に必要な Symplectic 幾何学の本 [3, Lectures on symplectic geometry]

1.2 Conditional Probability

条件付き確率を mfd 上で表現する disintagration と Radon 空間について [4, Regular conditional probability]

1.3 Integrators

Hamilton flow を近似するための ODE の構造保存型アルゴリズムの本 [5, Geometric numerical integration 2006]
geometric integrators の本 [6, Simulating Hailtonian Dynamics]

1.4 Hamilton Dynamics

Hamilton 系の ergod 性など [7, Hamiltonian Chaos]
the standard modern texts on classical mechanics [8, Classical dynamics]

1.5 Markov Chain

Markov Chain の本 [9, Markov Chains and Stochastic Stability, 2012]

2 Latest Results

2.1 Benchmark

2.1.1 Target Distribution

テストとして使われる的分布は Gaussian, Cauchy, Hierarchical などである．特に banana Gaussian[10] Hoffman and Gelman で'Sigma Wishart' D-dimensional Gaussian.[11, o-U-Turn sampler, 2014]

2.1.2 Metrics

[1, Conceptual introduction of HMC, 2018] や [12, Mix and Match HMC] では以下のような metrics が挙げられている．

- acceptance rate
- autocorrelation
- Effective Sample Size
- ESS/T
- BFMI

2.2 Time Step Size

Hamiltonian Monte Carlo では Hamilton flow を近似するために数値的 integrator を使う．integrator に由来する誤差を補正するため Metroplolis の acceptance ステップの段階を設ける．このため integrator の time step size が HMC のパフォーマンスに影響を与える．

HMC のコストを step size の関数として評価し，average acceptance probability も同様に step size の関数の関数として表す．step size を介してコストと acceptance probability の関係を与えることができる．このような方法で optimal step size について考える．

optimal integrator step size の最新評価の一つ: Beskos らは i.i.d な目的分布と leapfrog に対して HMC の step size の関数としての HMC のコストを下から評価した．[13, Optimal Tuning of Hamiltonian Monte Carlo, 2013]

Betancout らはさらに目的分布と integrator の条件を緩和し，上からの評価を行った．[14, Optimizing integrator step size, 2015]

2.3 Integration Time

integrator に関するパラメータで step size とともに重要なのが integration time である．各 Hamiltonian level set の探索において，integration time が短すぎると level set を十分カバーできず，長すぎると無駄な計算をすることになってしまう．また，各 level set で十分な探索に必要な itegration time は異なる．

itegration time は Hamilton flow の dynamic ergodicity(力学系の意味でのエルゴード性) とも深く関係している．

2.3.1 No-U-Turn Sampler(NUTS)

現在，Stan で実装されている integration time を動的に最適化するためのアルゴリズムが No-U-Turn Sampler(NUTS) である．NUTS は huristic な方法で，Euclidian Gaussian Kinetic に対してのみ well-difined である．[11, No-U-Turn sampler, 2014], [15, No U Turn Riemannian Manifolds, 2013]

2.3.2 Exhaustions

各 initial point に対して十分な探索時間として Exhaustions を定義する．これは Hamiltonian が適切な条件を満たせば構成できる．'Provided that the Hamiltonian is proper a valid exhaustion can always be constructed.' [16, optimal integration time, 2016]

2.4 Kinetic Energy

使用者に自由度がある Hamiltonian の Kinetic Energy 選択の最適化について考える．Kinetic Energy から誘導される contangent bundle 上の測度である Cotangent Disintegrations で呼ばれることもある．これは HMC において各点での momentum resampling の効率に影響を与える．各点 q での momentum resampling 後の Energy(Hamiltonian) の分布と Energy の周辺分布が近いほど効率よく目的の分布を探索できる．

目的の分布の形に依存して Kinetic Energy を適切に選ぶ必要がある．評価の指標としてエネルギーのヒストグラム, BFMI, ESS/T(E) などがある．

2.4.1 結果

まず，もっとも単純な Euclidian Gaussian cotangent disintegration をいくつかの目的分布に対して適用した結果を示す．Gaussian, non-centered 8 schools target に対して効率がよかった．Cauchy や 8 schools はより heavy-tailed であったため探索効率が落ちた．

ここで 8 schools target とは階層的ベイズモデルの一例で，階層的モデリングの有用性と計算の難しさを示す典型例である．[17, Diagnosing Suboptimal Cotangent Disintegrations, 2016]

2.4.2 Riemannian HMC

Riemann 幾何を Kinetic energy の mass matrix に適用することで曲率の情報を利用して効率的なサンプリングができる．

2.5 Geometric Ergodicity

MCMC などのアルゴリズムにおいて生成される estimator の有効性を考える上で Markov Chain の収束について考えることは重要である．特に MCMC や HMC に対して geometric ergodicity が成り立つかどうかを考える．HMC については Betancourt による結果がある．

2.5.1 results by Betancourt

position independent time:
各位置に依存せず integration time を決めるアルゴリズムで，ポテンシャルによる力が漸近的に中心力であり，gradient が線形増大であるような場合に geometric ergodicity を示した．

position-dependent integration time:
位置に依存して積分時間を決めることができる理想的なアルゴリズムを考える．この方法の利点は candidate map(hamilton flow) における drift 項を位置と積分時間でコントロールできること．Gaussian より heavy-tailed な exponential family について geometric ergodicity を示した．[18, Geometric Ergodicity of HMC, 2018]

2.5.2 証明の仮定

HMC の収束に関する証明にはポテンシャル U に対する log-concave 性, 勾配 ∇U の Lipschitz 性を仮定することが多い．[19, HMC on log-concave, 2017]

本質的な部分は Langevin Monte Carlo に関する議論が参考になる．[20, log-concave ULA] [21, smooth and log-concave]

log-concave 性についての一般的な性質や例は [22, Log Concavitym 2011]

2.5.3 基礎理論

HMC の収束に関する議論の基盤は Markov Chain の理論である．また，離散化した SDE の収束に関する議論も Ergodicity を証明する上で参考になる．[23, Markov Chain 2005] [24, Exponential convergence for LD 1996]

2.6 Advanced Algorithm

HMC から派生したアルゴリズムは数多く存在するらしい．各アルゴリズムで 10003 次元まで実験を行っている．パラメータについても細かく載っている．[12, Mix and Match HMC]

2.6.1 Scheme

高階スキームの有効性についての結果．特に高次元のパラメータ空間の問題に対しては有効かもしれない．[25, multi stage integrator, 2016]

2.7 Stochastic Gradient HMC

データ全てを full-gadient を計算するのを避けるためにランダムなミニバッチのデータを使って gradient を近似する方法が Stochastic Gradient HMC である．この方法はランダム性から local な minimum から抜け出すことができ，non-convex optimization にも効果的である．このアイデアの素は機械学習におけるパラメータ学習において発達した SGD である．[*1]

2.7.1 survey

(見つかる限り) 最初に出たのが [26, SGHMC] [*2]. ここではポテンシャルの微分を確率的に近似しそのノイズを軽減するために摩擦項を導入している．大阪大学でセミナーがあったらしい．[*3]

[27, SRVR-HMC, 2019] は stochastic gradient の variance を減らすアルゴリズム SRVR-HMC を提案し，この non-log-concave target density に対して収束の議論をし，Gaussian-mixture に対する実験を行っている．

その他，[28, finite-time performance of SGHMC, 2020] などがある．

2.7.2 non-log-concave

HMC の収束において density の log-concave 性 (もしくはポテンシャルの convex 性)

3 Future work and Application

3.1 Betancourt's

Betancourt 本人によるこれからの方向性を示したものがある．HMC の動的な step size のチューニングやそれに関する geometric ergodicity など HMC 既存のに関するもの，Bregman Hamiltonian や integrator の検討といった symplectic optimization 関連のもの，さらには Adiabatic Monte Carlo の実装といった話題が挙げられている．https://betanalpha.github.io/projects/index.html

[*1] 最新のものは Adam

[*2] スライド: http://people.ee.duke.edu/~lcarin/Wenzhao2.13.2015.pdf

[*3] http://www-mmds.sigmath.es.osaka-u.ac.jp/structure/activity/seminar.php?id=233

3.2 Association

Molecular Dynamics は natural Hamilton structure をもつ．[29, Understanding Molecular Dynamics, 2001]

3.3 Generalization

熱力学の知見を利用する．逆温度を導入して多峰性を持つ分布に対応する．Adiabatic Monte Carlo [30, Adiabatic Monte Carlo, 2015]

3.4 Complex Targets

Hierarchical Models [31, HMC for Hierarchical Models, 2013]

3.5 Functional Analysis

Hilbert 空間上の HMC*The key idea is to construct an MCMC method which is well defined on the Hilbert space itself* [32, on Hilber space]

PDE の研究，関数解析 [33, MCMC methods for functions]

SDE との関連 [34, Second order SDE]

2021/02/01 発見: 透水係数の逆解析 `http://soil.en.a.u-tokyo.ac.jp/jsidre/search/PDFs/20/%5B1-52%5D.pdf`

参考文献

[1] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo, 2018.

[2] Michael Betancourt, Simon Byrne, Sam Livingstone, and Mark Girolami. The geometric foundations of hamiltonian monte carlo. *Bernoulli*, 23(4A):2257–2298, 11 2017.

[3] Ana Cannas Da Silva and F Takens. *Lectures on symplectic geometry*, volume 3575. Springer, 2001.

[4] D. Leao, Jr., M. Fragoso, and P. Ruffino. Regular conditional probability, disintegration of probability and Radon spaces. *Proyecciones*, 23(1):15–29, 2004.

[5] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.

[6] Benedict Leimkuhler and Sebastian Reich. *Simulating Hamiltonian dynamics*, volume 14 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2004.

[7] M Combescure. Hamiltonian chaos and fractional dynamics. *Journal of Physics A: Mathematical and General*, 38:5380, 05 2005.

[8] Jorge V. José and Eugene J. Saletan. *Classical dynamics*. Cambridge University Press, Cambridge, 1998. A contemporary approach.

[9] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[10] Heikki Haario, Eero Saksman, and Johanna Tamminen. Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14, 1999.

[11] Matthew D. Homan and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, January 2014.

[12] Tijana Radivojevic and E. Akhmatskaya. Mix and match hamiltonian monte carlo. *arXiv: Computation*, 2017.

[13] Alexandros Beskos, Natesh Pillai, Gareth Roberts, Jesus-Maria Sanz-Serna, and Andrew Stuart. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[14] M. J. Betancourt, Simon Byrne, and Mark Girolami. Optimizing the integrator step size for hamiltonian monte carlo, 2015.

[15] M. J. Betancourt. Generalizing the no-u-turn sampler to riemannian manifolds, 2013.

[16] Michael Betancourt. Identifying the optimal integration time in hamiltonian monte carlo, 2016.

[17] Michael Betancourt. Diagnosing suboptimal cotangent disintegrations in hamiltonian monte carlo, 2016.

[18] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of hamiltonian monte carlo, 2018.

[19] Oren Mangoubi and Aaron Smith. Rapid mixing of hamiltonian monte carlo on strongly log-concave distributions, 2017.

[20] Alain Durmus and Éric Moulines. Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm. Preliminary version, April 2016.

[21] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities, 2016.

[22] GR Mohtashami Borzadaran and HA Mohtashami Borzadaran. Log-concavity property for some well-known distributions. *Surveys in Mathematics and its Applications*, 6:203–

219, 2011.

[23] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability.* Springer Science & Business Media, 2005.

[24] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341 – 363, 1996.

[25] Mario Fernández-Pendás, Elena Akhmatskaya, and J.M. Sanz-Serna. Adaptive multi-stage integrators for optimal energy conservation in molecular simulations. *Journal of Computational Physics*, 327:434–449, Dec 2016.

[26] Tianqi Chen, Emily B. Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo, 2014.

[27] Difan Zou, Pan Xu, and Quanquan Gu. Stochastic gradient hamiltonian monte carlo methods with recursive variance reduction. *Advances in neural information processing systems*, 2019.

[28] Xuefeng Gao, Mert Gürbüzbalaban, and Lingjiong Zhu. Global convergence of stochastic gradient hamiltonian monte carlo for non-convex stochastic optimization: Non-asymptotic performance bounds and momentum-based acceleration, 2020.

[29] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation.* Academic Press, Inc., USA, 2nd edition, 2001.

[30] M. J. Betancourt. Adiabatic monte carlo, 2015.

[31] M. J. Betancourt and Mark Girolami. Hamiltonian monte carlo for hierarchical models, 2013.

[32] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart. Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process. Appl.*, 121(10):2201–2230, 2011.

[33] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.

[34] Kevin Burrage, Ian Lenane, and Grant Lythe. Numerical methods for second-order stochastic differential equations. *SIAM J. Sci. Comput.*, 29(1):245–264, 2007.