Alexander Hamiltonian Monte Carlo

(Anonymized for Review)

Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

A curious defect of the Hamiltonian Monte Carlo algorithm, an otherwise state-of-the-art Markov chain Monte Carlo scheme for simulating from an unnormalized density, is that despite its name, it has very little to do with Lin-Manuel Miranda's hit musical *Hamilton*. To remedy this flaw, we propose *Alexander* Hamiltonian Monte Carlo, a revolutionary new variant of HMC in which the Metropolis-Hastings accept/reject step is replaced by an elaborate *duel* between the proposed position and the previous state of the chain. The result is a loud (in the sense of noise), diverse (in the sense of variance) celebration of both Bayes and Broadway, which is guaranteed to converge to the room where it happens, if the user is willing to wait for it.

1. Introduction

Probabilistic modeling and inference are central to many fields, from statistics to robotics to the natural sciences. In the past half century, the increasing availability of cheap computing power has revolutionized the practice of modeling and inference. Particularly important has been the development of general-purpose methods for probabilistic computation, e.g. Markov chain Monte Carlo.

Despite their widespread adoption, however, MCMC methods are often misunderstood; it is easy for practitioners to make unjustified assumptions about the properties of the MCMC algorithms they use, making them dangerous to apply to real-world problems. One particularly misleading case is the popular Hamiltonian Monte Carlo algorithm, which, despite its name, *has almost nothing to do with Lin-Manuel Miranda's hit Broadway musical Hamilton.*

To remedy this flaw, this paper proposes *Alexander* Hamiltonian Monte Carlo, a revolutionary new variant of HMC in which the usual Metropolis-Hastings accept/reject step is replaced by a novel *Hamilton-Burr* mechanism, involving the simulation of an elaborate *duel* between the proposed position and the previous state of the chain. As one might expect, this change leads to louder (in the sense of higher noise), more diverse (in the sense of higher variance) estimates. Crucially, our algorithm retains the property of ordinary HMC that for users who are willing to wait for it, the samples eventually find their way to the room where it happens.

Algorithm 1 Hamilton-Burr Mechanism

Number 1! With chance $\frac{1}{10}$ as the fraction, x' apologizes; no need for further action. (Repeat x as next state in chain.)
Number 2! If it doesn't, choose your standard normal seconds: your Lieutenants, u, u', when there's reckoning to be reckoned.
Number 3! Have u, u' meet face-to-face. They negotiate this truce with probability $\frac{1}{8}$:
  if a coin with weight $\min(1, \frac{\pi(x+u+u')}{\pi(x)})$ flips heads **then**
    accept $x' = x + u + u'$ as the next state
  else
    say no to this (repeat x in the chain)
Number 4! If they don't reach a peace, that's all right. Time to get Dr. Momentum on site. Draw p in advance, with Gaussian variability. After, have him turn around, to have reversibility.
Number 5! Prepare to duel before the CPU fries; x' tries to fly to where π is high.
  (Using a symplectic integrator, compute a Hamiltonian trajectory starting at (x, p) to obtain proposed (x', p'))
Number 6! x' leaves a note for next of kin, tells 'em H(where it's been). Prays Metropolis will let it in.
Number 7! Time for x to confess its sins. Measure $H(x, p)$, its adrenaline.
Number 8! The last chance to negotiate. If x''s score is biggest, it can set the record straight. (This is commonplace, for x *would surely* lose disputes; this way x goes home and no one shoots.)
Number 9! Counting paces to the point of No-U-Turn, x and x' take their aim, draw their fates from an urn... 1,2,3,4,5,6,7,8,9...
Number 10! Fire!
  Launch two bullets with exponentially-distributed hit times:
    $B_{x'} \sim \text{Exp}(H(x', p'))$, $B_x \sim \text{Exp}(H(x, p))$
  if $B_{x'}$ hits first **then**
    only x' lives to tell its story
  else
    only x lives to tell its story
[N.B.: After each transition, auxiliary variables should be dropped from memory, by insisting, politely but firmly, "I'm sure I don't know what u means; u, forget yourself."]

2. Alexander Hamiltonian Monte Carlo

Let π be an unnormalized density with respect to the Lebesgue measure on a state space $X = R^n$. The Alexander Hamiltonian Monte Carlo algorithm specifies a π-invariant transition kernel (Algorithm 1) that can be iterated from any initial distribution to generate samples $x_1, x_2, \ldots, x_n$ whose marginals converge to π/Z. Like HMC, it operates on an extended state space X^2 of position variables x and momentum variables p, with an extended target $H(x, p) = \pi(x)\mathcal{N}(p; 0, I)$. Our key result is as follows:

Figure 1. Out of the box, AHMC does not yield estimates of the MLL (marginal log likelihood). However, it is possible to obtain unbiased estimates of the LMM (Lin-Manuel Miranda). We show seven samples here.

Theorem 2.1. *Alexander Hamiltonian Monte Carlo eventually converges to the room where it happens, if the user is willing to wait for it.*

Proof. The proof requires three fundamental truths, at the exact same time:

Lemma 2.2 (Stationarity)**.** *AHMC won't change the subject, cuz the posterior is its favorite subject, forever, and ever, and ever and ever and ever.*

Proof. The entire duel can be seen as a mixture of three kernels: the identity (if x' apologizes in Step 1), a Gaussian drift kernel (if u, u' reach a truce in Step 3), or an ordinary HMC kernel (otherwise). We declare the mixing probabilities independent of the State (and we hold this truth to be self-evident, i.e., an exercise for the reader). Then since each sub-kernel is stationary for π, so is AHMC. □

Lemma 2.3 (Aperiodicity)**.** *You'll be back—time will tell.*

Proof. Let $A \subset X$ and consider $\Pi(t) = \int_A P^t(x, A)dx$ for $t \in \mathbb{N}$. We show that the chain will be back ($\Pi(t) > 0$ for some $t > 0$), and perhaps more importantly, that time will tell, i.e., that—without waiting and seeing—it is impossible to rule out a particular time t for the homecoming ($\Pi(t) > 0$ for *all* $t > 0$). Because the current state of the chain wins the duel with positive probability from any state, $P^1(x, A) > 0$ whenever $x \in A$, and therefore $\Pi(1) > 0$. For $t > 1$, observe that it is possible to win the duel t times in a row, so that $\Pi(t) > 0$ too. □

Lemma 2.4 (Irreducibility)**.** *Even a bastard, orphan, son of a whore and a Scotsman, dropped in the middle of a forgotten spot in the Carribean by Providence, impoverished in squalor, can grow up to be a hero and a scholar.*

Proof. Let the bastard, orphan, etc. be dropped in the middle of a forgotten spot x in the state space X. The question is whether, for any definition $Y \subset X$ of herodom and scholarship (as clearly there is no single objective definition), it is possible for him to, via a sequence of AHMC transitions, ascend from x to a point in Y. In fact this jump can be achieved in a single AHMC transition, albeit not with very high probability. To see this, note that $Pr(x+u+u' \in Y) = \int_Y \mathcal{N}(y; \mu = x, \sigma^2 = 2)\mathrm{d}y > 0$ whenever Y has positive measure. Because $x + u + u'$ is proposed as x' with positive probability (namely $(1-\frac{1}{10})\cdot\frac{1}{8} = \frac{9}{80}$), and then accepted with non-zero probability, the probability of landing in Y after one transition of AHMC is also positive. □

Because of all three fundamental truths, at the exact same time, the desired convergence property holds. □

3. Experiments

The experiment lasted two minutes, maybe three minutes, and everything it said was in total agreement.

4. Practical Advice on the Application of AHMC

1. It may be tempting to read every sample it writes you, searching and scanning for answers in every line, for some kind of sign. However, we caution users against this behavior ("watching it burn in"); early samples are best erased from the narrative.
2. Users in New Jersey may find it useful to replace the AHMC kernel with any stochastic procedure for proposing and then accepting a new state x'. (Everything is legal in New Jersey.)
3. Unlike sequential Monte Carlo and other algorithms based on importance sampling, Alexander Hamiltonian Monte Carlo unfortunately does not yield estimates of the MLL (marginal log likelihood), at least not out of the box. However, one can use the output of AHMC to obtain unbiased estimates of the LMM (Lin-Manuel Miranda), as depicted in Figure 1. We emphasize the novelty here: most existing estimates of LMM's value are biased.

5. Related Work (by blood)

Unimportant, there's a million things I haven't done, just you wait, just you wait.

6. Discussion

To be honest we are doubtful AHMC will prove useful to practitioners. But you never know. After all, who can say which ideas live? Which die? Who cites our story?