

# Antithetic Sampling with Hamiltonian Monte Carlo

Dan Piponi  
 Matthew D. Hoffman  
*Google AI*

DPIPONI@GOOGLE.COM  
 MHOFFMAN@GOOGLE.COM

## 1. Introduction

It is useful to generate pairs of samples with the same marginal distribution but which are negatively correlated when considered jointly. For example, when we have a probability distribution with reflection symmetry we can use antithetic variates to lower the variance of estimates of an expectation. Here we show how we can generate pairs of negatively correlated samples from a Hamiltonian Monte Carlo (HMC, Neal (2012); Duane et al. (1987)) chain when the center of reflection is unknown and the distribution is only approximately symmetric. In earlier work, Craiu and Meng (Craiu and Meng (2005)) showed how to sample using antithetically coupled Markov Chain Monte Carlo in some other special cases.

First we consider the usual Hamiltonian Monte Carlo algorithm to generate samples from the probability measure  $\mathcal{Z}^{-1} \exp(-U(x))dx$  on  $\mathbb{R}^d$  with unknown normalization constant  $\mathcal{Z}$ . Let  $\Phi$  be the normal distribution on  $\mathbb{R}^d$  with mean 0 and covariance  $I_d$ . For  $q, p \in \mathbb{R}^n$  define the Hamiltonian  $H(q, p) = U(q) + \frac{1}{2}p^2$ . Also suppose that  $\psi_T(q, p)$  is a symplectic method Hairer et al. (2006) that numerically integrates the physical system described by the Hamiltonian  $H$  for time  $T$ , returning  $(q', p')$ . The standard Metropolis-adjusted Hamiltonian Monte Carlo algorithm is given in Algorithm 1. The output is a sequence of samples,

---

### Algorithm 1 Metropolis-adjusted HMC

---

**Input:** Initial state  $X_0$ , number of samples  $N$ .  
**for**  $i = 0$  to  $n - 1$  **do**  
  Sample  $p_i \sim \Phi$ . {Sample momentum from normal distribution}  
  Set  $(q_{i+1}, p_{i+1}) = \Psi_T(X_i, p_i)$ . {Symplectic integration}  
  Sample  $b_i \sim \text{Unif}([0, 1])$ . {Metropolis adjustment}  
  **if**  $b_i < \min(1, \exp(H(q_i, p_i) - H(q_{i+1}, p_{i+1})))$  **then**  
    Set  $X_{i+1} = q_{i+1}$ .  
  **else**  
    Set  $X_{i+1} = X_i$ .  
  **end if**  
**end for**

---

$X_1, \dots, X_N$ , approximately drawn from  $\mathcal{Z}^{-1} \exp(-U(x))dx$ .

In Algorithm 2 we run a pair of coupled HMC chains sharing momenta and using the same random variables in the Metropolis adjustment step: Both  $X_1, \dots, X_N$  and  $Y_1, \dots, Y_N$  are approximate samples from  $\mathcal{Z}^{-1} \exp(-U(x))dx$ . It can be shown, under reasonable

---

**Algorithm 2** Coupled Metropolis-adjusted HMC

---

```

1: Input: Initial states  $X_0, Y_0$ ; number of samples  $N$ .
2: for  $i = 0$  to  $n - 1$  do
3:   Sample  $p_i \sim \Phi$ . {Shared  $p_i$ }
4:   Set  $(q_{i+1}, p_{i+1}) = \psi_T(X_i, p_i)$ .
5:   Set  $(q'_{i+1}, p'_{i+1}) = \psi_T(Y_i, p_i)$ .
6:   Sample  $b_i \sim \text{Unif}([0, 1])$ . {Shared  $b_i$ }
7:   if  $b_i < \min(1, \exp(H(q_i, p_i) - H(q_{i+1}, p_{i+1})))$  then
8:     Set  $X_{i+1} = q_{i+1}$ .
9:   else
10:    Set  $X_{i+1} = X_i$ .
11:   end if
12:   if  $b_i < \min(1, \exp(H(q'_i, p_i) - H(q'_{i+1}, p'_{i+1})))$  then
13:     Set  $Y_{i+1} = q'_{i+1}$ .
14:   else
15:     Set  $Y_{i+1} = Y_i$ .
16:   end if
17: end for

```

---

conditions on  $U$  and  $T$ , that the steps are contractive in the sense that

$$\mathbb{E}[|X_{i+1} - Y_{i+1}|] < c\mathbb{E}[|X_i - Y_i|]$$

for some  $c$  depending on  $U$  and  $T$ . (See [Bou-Rabee et al. \(2018\)](#), [Mangoubi and Smith \(2017\)](#).)

Suppose  $U(x) = U(2\mu - x)$  for some  $\mu$  (i.e. that  $U(\mu + (x - \mu)) = U(\mu - (x - \mu))$ ) so  $U$  is symmetric about  $\mu$ .) and negate the momentum for the  $Y$  chain, so that line 5 becomes

$$\text{Set } (q'_{i+1}, p'_{i+1}) = \psi_T(Y_i, -p_i).$$

Now the sequence  $(X_1, 2\mu - Y_1), (X_2, 2\mu - Y_2), \dots$  is a pair of HMC chains for the distribution given by  $U(x)$ . After all, it is the same as the original coupled algorithm apart from  $Y_i, q'_i$  and  $p_i$  being reflected about  $\mu$ . So both  $X_i$  and  $Y_i$  are sequences of approximate samples from  $\mathcal{Z}^{-1} \exp(-U(x))dx$ . However, we now have

$$\mathbb{E}[|X_{i+1} + Y_{i+1} - 2\mu|] < c\mathbb{E}[|X_i + Y_i - 2\mu|]$$

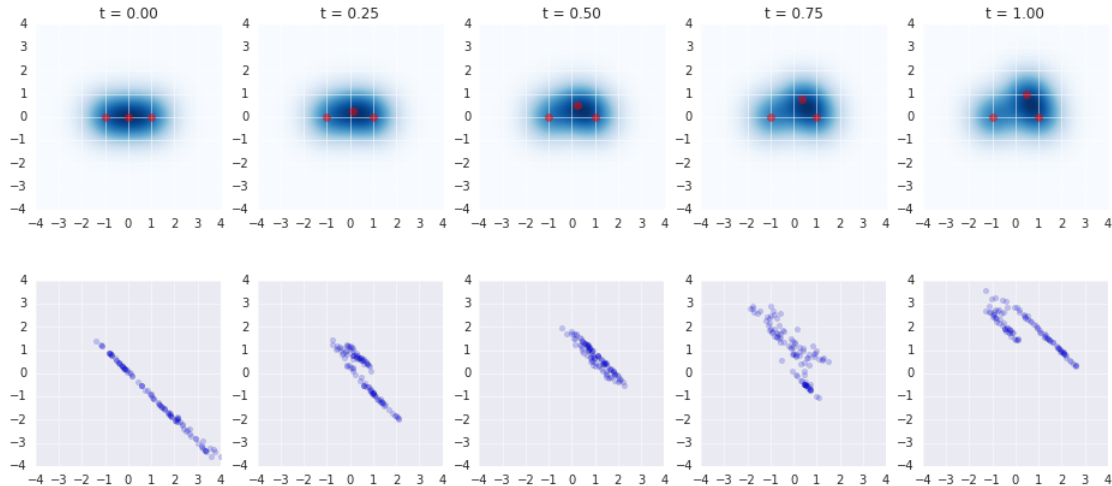
so we expect the  $X_i$  and  $Y_i$  to be anti-correlated. Now suppose that  $U(x) = U(2\mu - x)$  holds only approximately. We can no longer expect perfect anticorrelation between the  $X_i$  and the  $Y_i$ . Nonetheless, if  $U(x)$  and  $U(2\mu - x)$  are similar enough we can expect some degree of anticorrelation and variance reduction when used as antithetic variates.

## 2. Implications

For any real-valued variables  $X$  and  $Y$  we have

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \tag{1}$$

Figure 1: Upper row: Density plots of  $U_t(x, y)$  for various  $t$ . Lower row: scatter plots of the first coordinate of  $Y_i$  against the first coordinate of  $X_i$  for 500 samples.



As a consequence, if  $X$  and  $Y$  are estimators of a parameter, then  $\frac{1}{2}(X + Y)$  provides a much improved estimator if we can ensure that  $X$  and  $Y$  are anti-correlated. This implies that averaging anti-correlated Markov chains can yield low-variance estimators.

Let  $v$  be the variance of the estimator  $\hat{X} = \frac{1}{N} \sum_i X_i$ . Define the effective sample size (ESS) of these  $N$  samples by

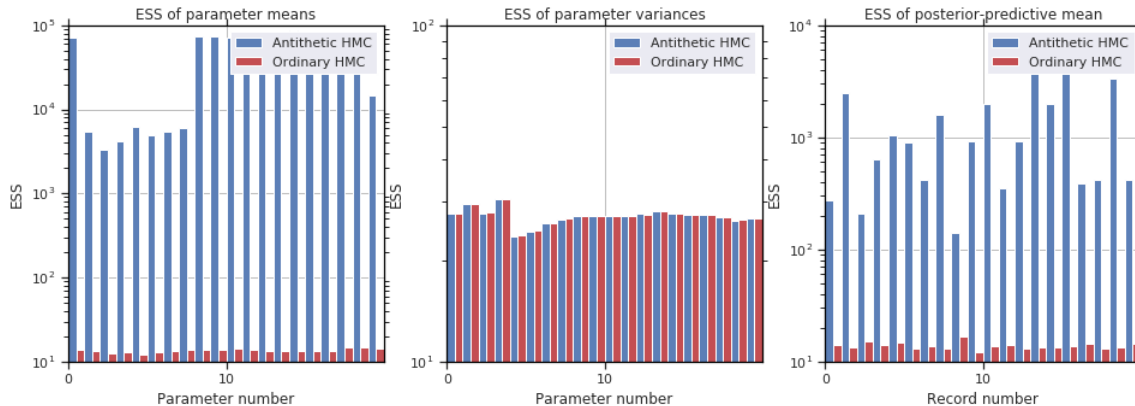
$$v = \frac{\sigma^2}{\text{ESS}} \quad (2)$$

where  $\sigma^2$  is the variance of the target distribution. When the  $X_i$  are independent,  $\text{ESS} = N$ . Generally for MCMC algorithms each sample is correlated with the previous one resulting in a higher variance and an ESS less than  $N$ . However, if the samples come in negatively-correlated pairs, we can expect a lower variance and possibly an ESS larger than  $N$ . The ESS is often estimated by considering the the integrated autocorrelation spectrum, but we adopt the more direct and robust (but more expensive) approach of estimating the variance of the MCMC estimator as follows: we run  $M$  chains, each with  $N$  samples. We estimate the variance  $\sigma^2$  of the target distribution using all  $NM$  samples  $X_i$ . We estimate  $v$  from the  $N$  estimators  $\frac{1}{N} \sum_{i=1}^N X_i$ , one for each chain, and then estimate the variance from this collection of  $N$  estimators.

### 3. Results

Our first experiment qualitatively validates the anticorrelation effect of anti-coupled HMC on a perfectly symmetric mixture-of-Gaussians target distribution, and examines what happens as the mixture becomes increasingly asymmetric. We use an equal mixture of three two-dimensional Gaussians, each with standard deviation  $\frac{3}{4}$  and with means  $(-1, 0)$ ,  $(1, 0)$  and  $(\frac{t}{2}, t)$  for various values of  $t$ . At  $t = 0$  we have perfect reflection symmetry and we lose this symmetry for  $t \neq 0$ . In Figure 1 we show, for  $t = 0, 0.25, 0.5, 0.75$  and  $1$ , density plots

Figure 2: Effective sample sizes (ESS) of an ordinary HMC chain and for an antithetic HMC chain. Left: ESSs of estimators of the first 20 parameter means (out of 62). Middle: ESSs for the first 20 parameter variances. Right: ESSs for the first 20 posterior-predictive means (out of 1000).



of the probability distribution and a scatter plot of the first coordinates of an anti-coupled pair for 500 steps after a burn-in time of 500 steps. When  $t = 0$  we have near-perfect anti-correlation. For  $t \neq 0$  we lose the perfect line but still have strong negative correlation. Sometimes there are two approximate lines with gradient  $-1$ ; this happens when the distribution is multi-modal and the two chains linger in different potential wells. Nonetheless, even these cases we observe negative correlations.

Our next experiment quantitatively evaluates the efficiency of the anti-coupled HMC estimator on a Bayesian logistic regression model applied to the German credit dataset (Dheeru and Karra Taniskidou (2017)), with 62 parameters. We ran 500 separate pairs of anti-coupled HMC chains for 800 steps each, discarding the first 200 steps as burn-in. In Figure 2 we compare the effective sample sizes (ESSs) for ordinary HMC and for the pair of chains as described in section 2. We consider three summaries of the posterior: means, variances, and predictive distributions (i.e., if in the model  $p(y_n = 1 | x_n, \theta) = \sigma(x_n^\top \theta)$ , our goal is to accurately estimate  $\int_{\theta} p(\theta | x, y) \sigma(x_n^\top \theta) d\theta$ ). The anti-coupled chain requires twice as much computation (which can be parallelized), but as expected yields higher ESSs for estimates of the posterior means. The results for estimating posterior variances are unimpressive, which is to be expected since if  $X$  and  $Y$  have mean  $\mu$  and are perfectly anticorrelated, then  $(X - \mu)^2$  and  $(Y - \mu)^2$  are perfectly correlated. The ESSs for the posterior-predictive distributions are again excellent as they are monotonic functions of the parameters (Crain and Meng, 2005).

## 4. Conclusions

We proposed anti-coupled HMC, a twist on classic parallel-chain HMC that can significantly improve HMC's efficiency. Since anti-coupled HMC is so easy to implement, we hope that it may become a standard part of the posterior-inference toolbox.

**References**

- N. Bou-Rabee, A. Eberle, and R. Zimmer. Coupling and Convergence for Hamiltonian Monte Carlo. *ArXiv e-prints*, May 2018.
- R. V. Craiu and X.-L. Meng. Multiprocess parallel antithetic coupling for backward and forward Markov Chain Monte Carlo. *ArXiv Mathematics e-prints*, May 2005.
- Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017. URL <http://archive.ics.uci.edu/ml>.
- S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett.*, B195:216–222, 1987. doi: 10.1016/0370-2693(87)91197-X.
- Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.* Springer, Dordrecht, 2006. URL <https://cds.cern.ch/record/1250576>.
- O. Mangoubi and A. Smith. Rapid Mixing of Hamiltonian Monte Carlo on Strongly Log-Concave Distributions. *ArXiv e-prints*, August 2017.
- R. M. Neal. MCMC using Hamiltonian dynamics. *ArXiv e-prints*, June 2012.