Table 1: Model parameters. n_0 determines tempering by $T = \frac{n_0}{n}$. For missing n_0, $T = 1$. Start deviation is the standard deviation of the random starting point in the DP experiments. a is the parameter determining how curved the banana distribution is, with $a = 0$ corresponding to a Gaussian distribution.

Name	Dimension	n	n_0	Start Deviation	a
Easy banana, d = 2	2	100000		0.020	20.0
Easy banana, d = 10	10	200000		0.020	20.0
Tempered banana, d = 2	2	100000	1000.0	0.020	20.0
Tempered banana, d = 10	10	200000	1000.0	0.020	20.0
High dimensional Gauss	30	200000		0.020	0.0
Hard banana	2	150000		0.020	350.0
Correlated Gauss	2	200000		0.005	0.0
Circle	2	100000		0.100	

1 DP MCMC Experiments Summary

1.1 MMD

The main metric I used to evaluate MCMC algorithms is maximum mean discrepancy (MMD). MMD is a statistic that measures the difference between two distributions, and can be estimated from samples of both distributions [1]. Using MMD requires choosing a kernel for the estimation, and the properties of MMD depend on the chosen kernel. I used the Gaussian kernel, as it requires the distributions to exactly match reach 0 MMD. The Gaussian kernel requires choosing a kernel width, which affects how differences in different moments affect the MMD. I chose the width with the same method as Gretton et al. [1], taking the median between the differences between the two samples, with the exception that I took a subsample of both samples, and computed the median between differences of the subsamples. This is necessary as the MCMC algorithms produce chains of variable lengths.

1.2 Posteriors

Figure 1 shows contour plots of the 2 dimensional posteriors used for the evaluations. Table 1 contains the parameters of the different models. The variances of the models did not fit well in the table, so I'll describe them here. For all banana experiments and 30-dimensional Gauss, the first component has variance 20, the second has variance 2.5, and the rest have variance 1. The correlated Gauss has covariance

$$\begin{bmatrix} 1 & 0.999 \\ 0.999 & 1 \end{bmatrix}.$$

Note that these are the know variances of the likelihood, not posterior variances.

1.3 Clipping

Before running experiments on DP MCMC I evaluated the performance of non-DP random walk Metropolis Hastings and HMC with log likelihood ratio clipping on the easy banana distribution in both 2 and 10 dimensions. The purpose

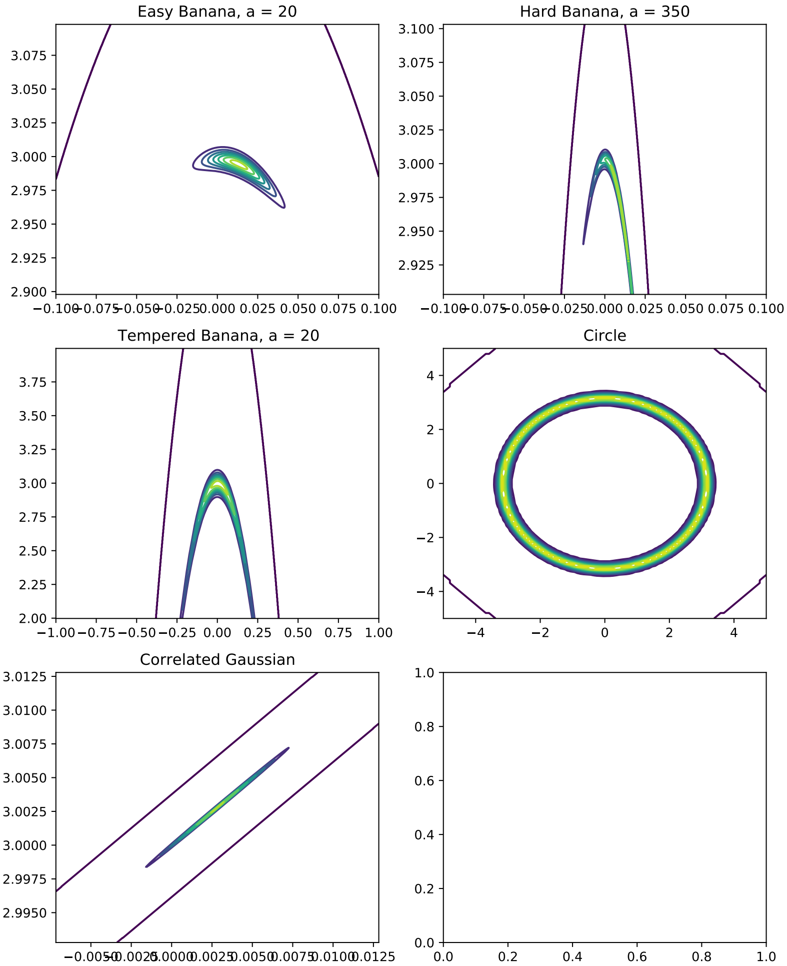

Figure 1: Contour plots of the 2-dimensional posteriors

of this evaluation is to validate the hypothesis that a small amount of clipping does not significantly affect the convergence of an MCMC algorithm, and to find out what that small amount is.

Based on Figure 2, particularly the bottom row, clipping less than 10% of the log likelihood ratios does not significantly affect convergence. Because of this, I tried to set the clip bounds for the DP experiments to be large enough that clipping is well under 10%, when that was possible.

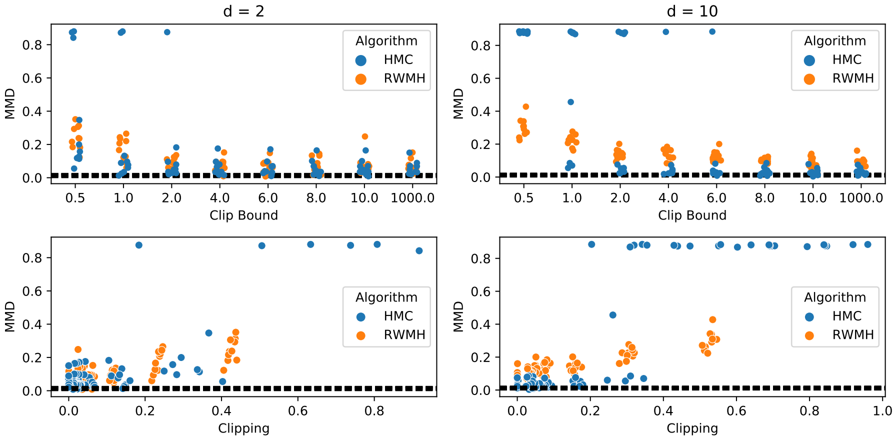

Figure 2: Clipping experiment

1.4 The algorithms

The experiments evaluate 6 MCMC algorithms. DP penalty is the algorithm of Yildirim and Ermis [3], and DP penalty advanced is DP penalty that only updates one component per sample and keeps moving in the same direction for each component until a proposal is rejected, which according to Yildirim and Ermis [3] can improve the performance of the algorithm. Minibatch DP penalty and minibatch DP penalty advanced are variants of DP penalty (advanced) that only consider a subsample of the log likelihood ratios for the accept test. Barker is the algorithm of Heikkilä et al. [2].

1.5 Banana experiments

All of the DP experiments ran the algorithms 20 times with varying values of ϵ. The chains started at a randomly chosen point centered around the true model parameter values, expect for circle where the random starting point is centered around $(0, 1)$. The 20 starting points are the same across algorithms and values of ϵ. Table 1 column Starting Deviation shows the the standard deviation of the random starting points.

Figures 3, 4 and 5 show MMD, clipping and acceptance rate for the easy and tempered 2 and 10-dimensional banana posteriors. In the non-tempered experiments, DP HMC is slightly behind DP penalty, but in the tempered experiment DP HMC is slightly ahead. Clipping is small for all algorithms except Barker, as the clip bound for Barker is not adjustable.

Figures 6, 7 and 8 show MMD, clipping and acceptance for the harder experiments, the 30-dimensional Gaussian, the very curved hard banana, and the very correlated Gaussian. In the 30-dimensional Gaussian, HMC is very slightly better than DP penalty, but loses to DP penalty advanced. With the hard banana, all algorithms perform equally well, until $\epsilon = 4$ where DP penalty advanced gets very high variance for MMD, and larger values of ϵ where DP penalty advanced fails completely. This is likely due to the large amount of clipping for DP penalty advanced. In the correlated Gaussian, HMC outperforms the other two algorithms.

Figure 9 shows the gradient clipping of HMC for all the previous experiments. Gradient clipping does not affect the convergence of HMC, so keeping it small it not as important as with log likelihood clipping. However, gradient clipping does the acceptance rate of HMC, but increasing the clip bound to lower clipping will increase the noise added to gradients, thus also lowering the acceptance rate.

Finally, Figure 10 shows the distance from the posterior mean to the true mean on the left, and clipping on the right, for the circle experiment. Mean error was used instead of MMD, because in the circle model the mean error measures how well the samples cover the entire ring of high probability evenly, and the true mean can easily be shown to be 0.

With $\epsilon = 0.5$ HMC outperforms DP penalty, but with larger epsilons the differences disappear. This is likely due to both algorithms getting close sampling from the true posterior with $\epsilon = 1$.

References

[1] Arthur Gretton et al. “A Kernel Two-Sample Test”. In: *J. Mach. Learn. Res.* 13 (2012), pp. 723–773. URL: `http://dl.acm.org/citation.cfm?id=2188410`.

[2] Mikko A. Heikkilä et al. “Differentially Private Markov Chain Monte Carlo”. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada.* 2019, pp. 4115–4125. URL: `http://papers.nips.cc/paper/8665-differentially-private-markov-chain-monte-carlo`.

[3] Sinan Yildirim and Beyza Ermis. “Exact MCMC with differentially private moves - Revisiting the penalty algorithm in a data privacy framework”. In: *Statistics and Computing* 29.5 (2019), pp. 947–963. DOI: `10.1007/s11222-018-9847-x`. URL: `https://doi.org/10.1007/s11222-018-9847-x`.

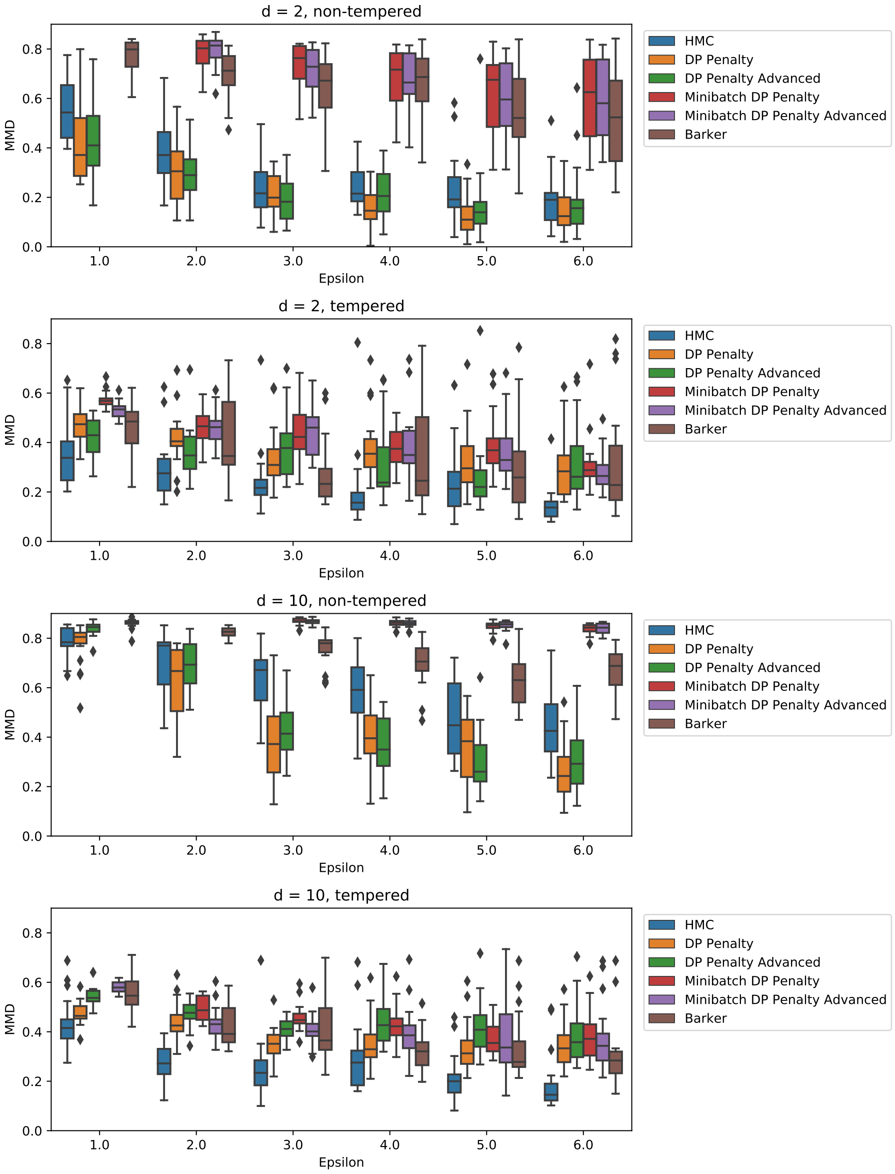

Figure 3: Banana experiment MMD

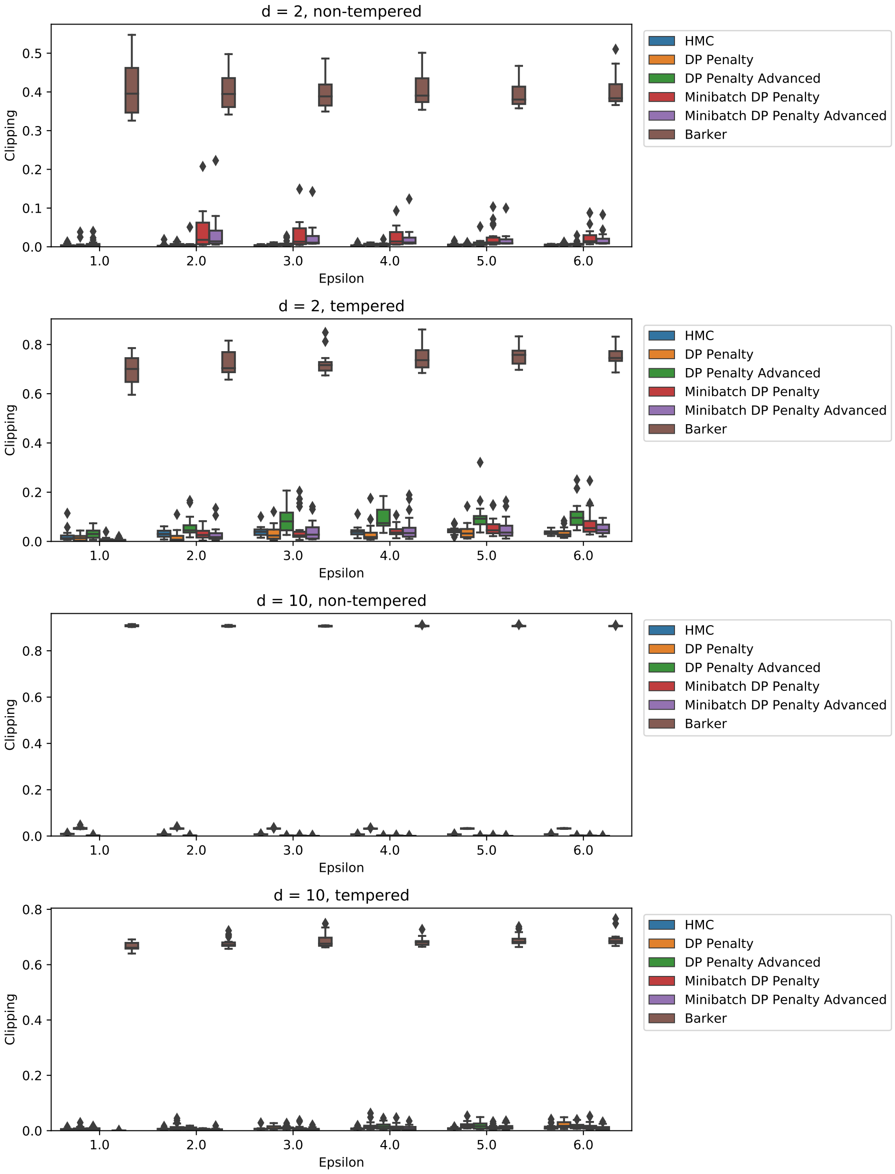

Figure 4: Banana experiment clipping

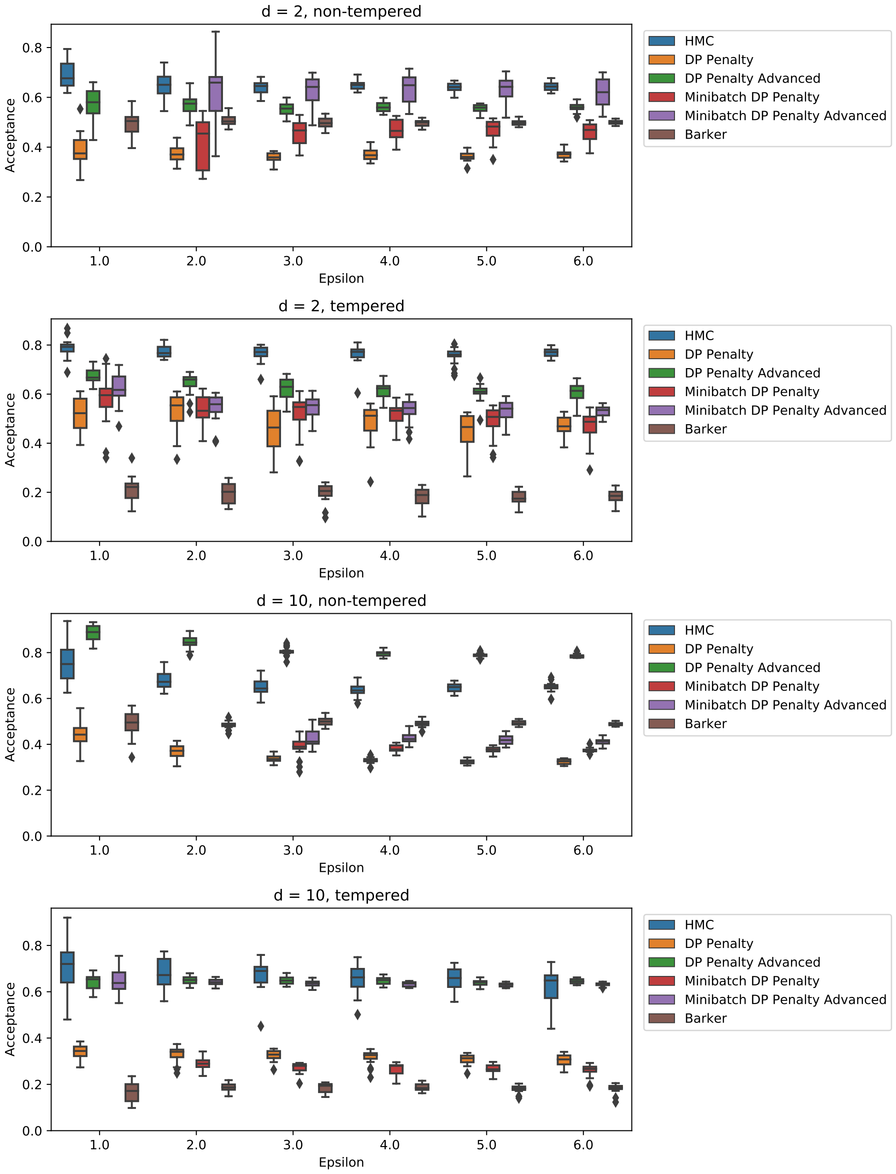

Figure 5: Banana experiment acceptance rate

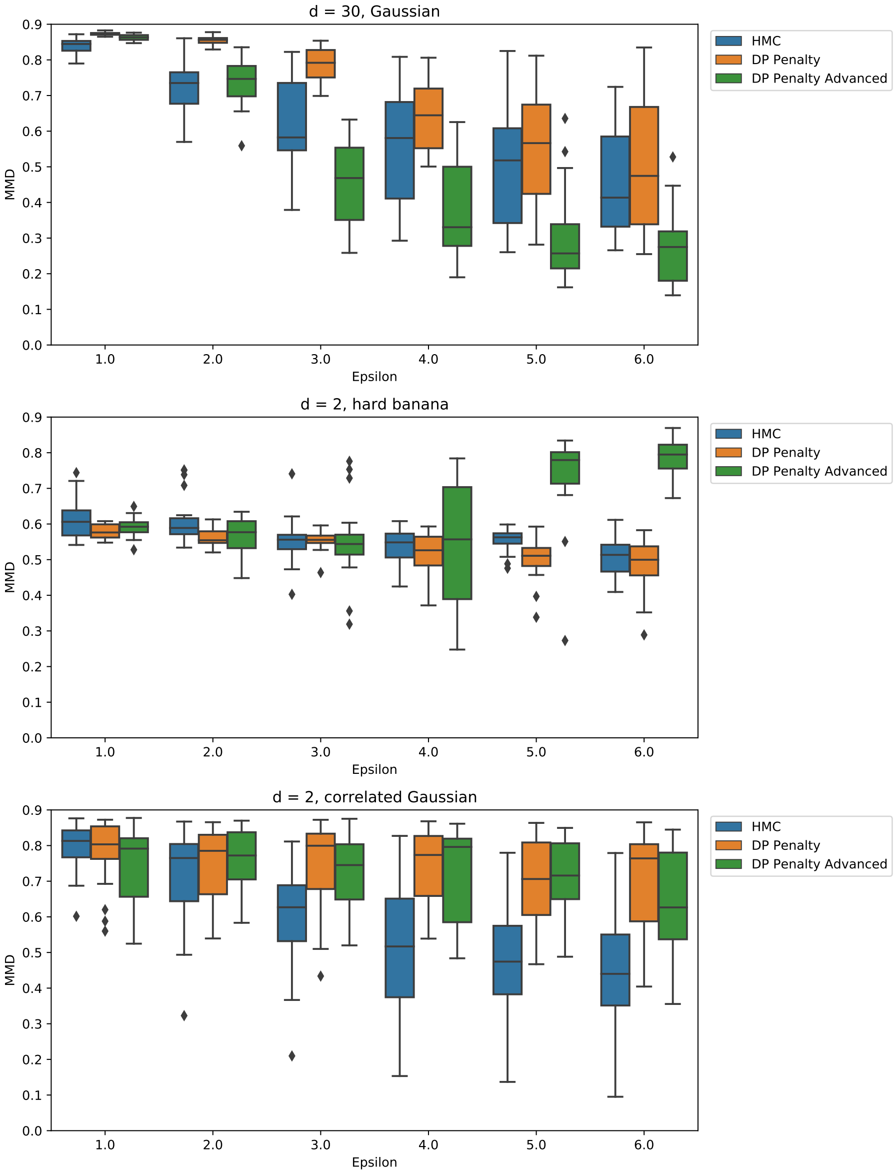

Figure 6: Hard banana and Gaussian MMD

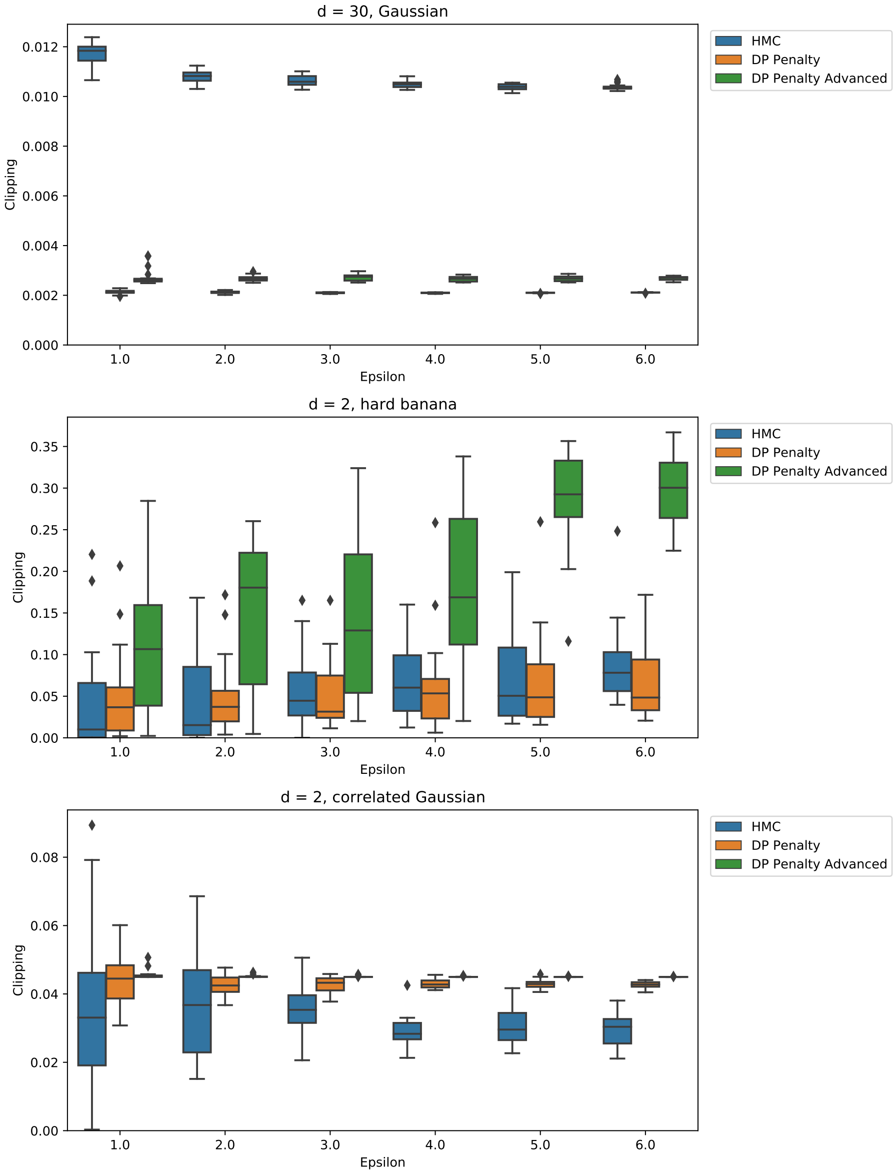

Figure 7: Hard banana and Gaussian clipping

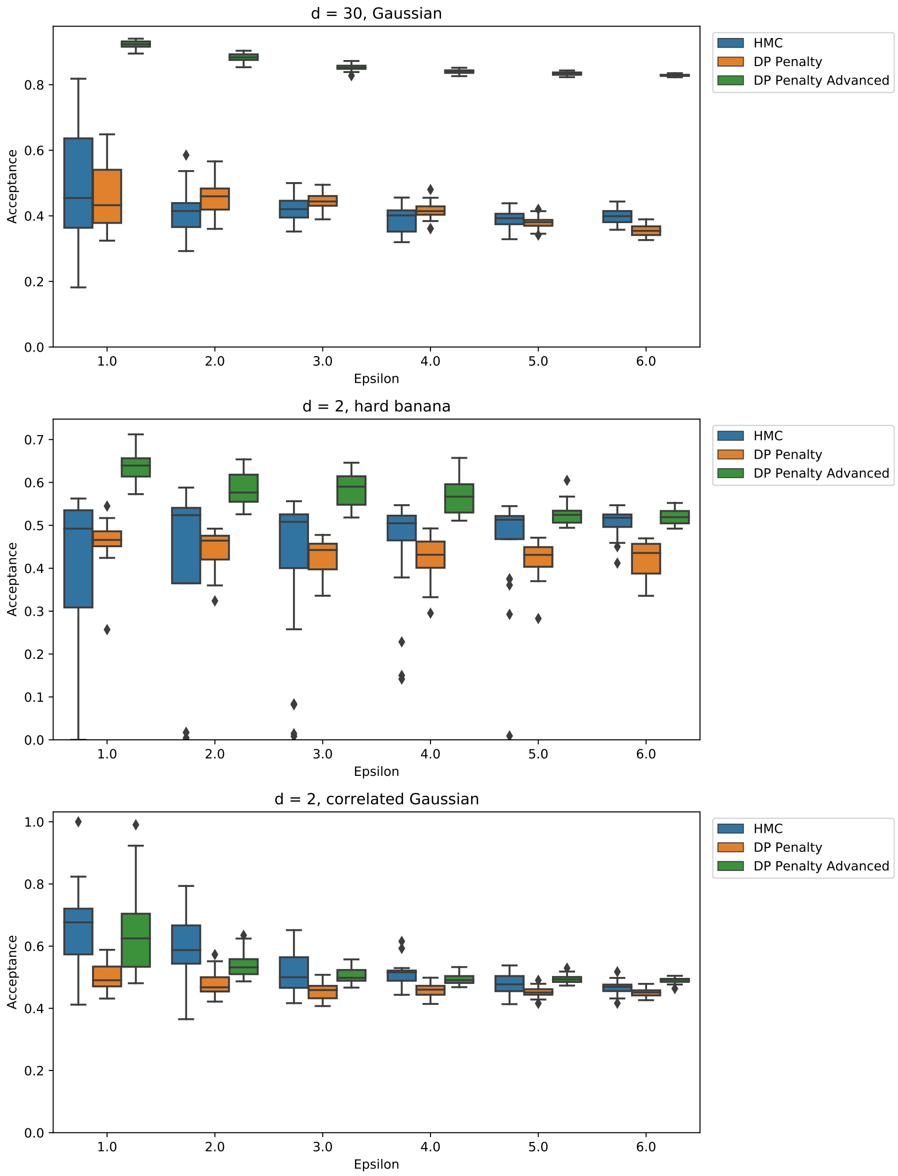

Figure 8: Hard banana and Gaussian acceptance rate

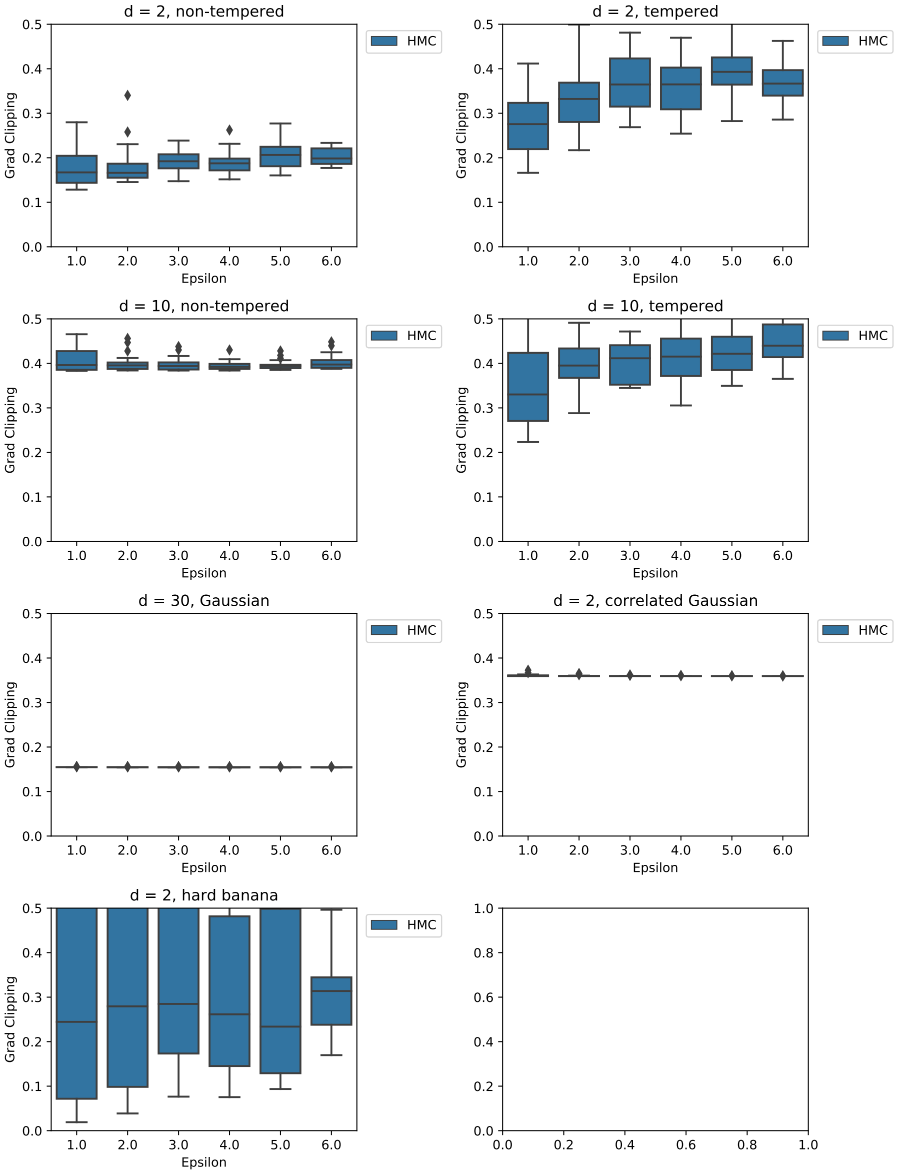

Figure 9: Gradient clipping for HMC

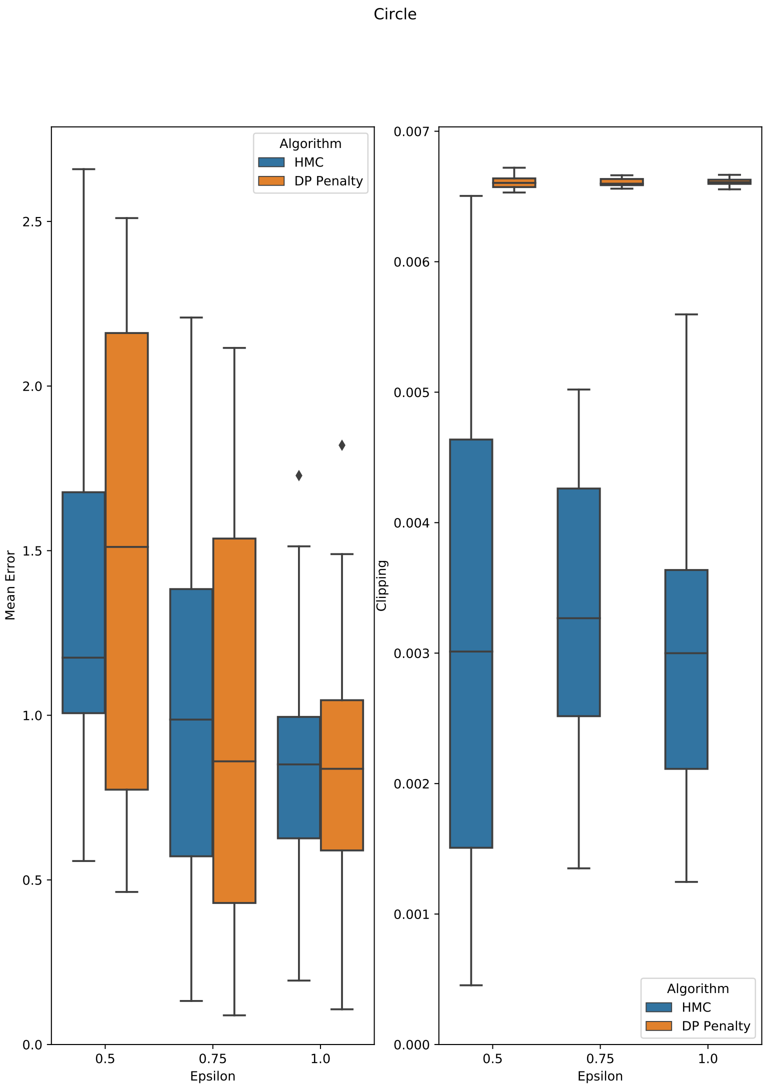

Figure 10: Circle experiment