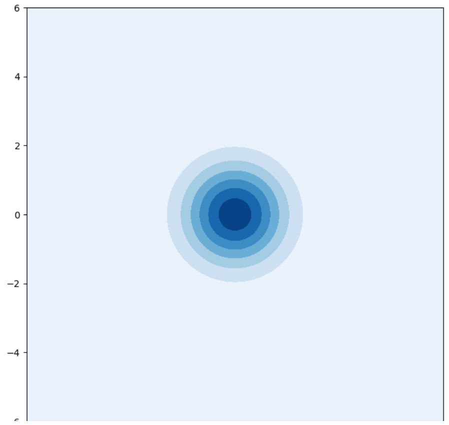
Fixed-distance Hamiltonian Monte Carlo

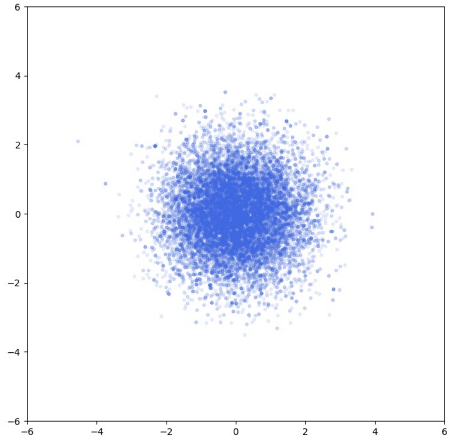
Hadi Mohasel Afshar, Sally Cripps

CSIRO’s Data61 & The University of Sydney

Markov Chain Monte Carlo (MCMC)

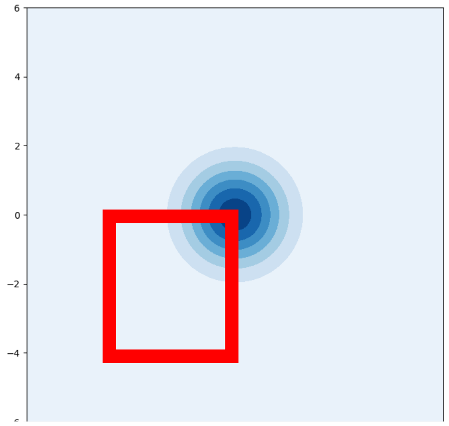
- MCMC are a class methods that approximate a probability distribution with a set of states a.k.a. samples.

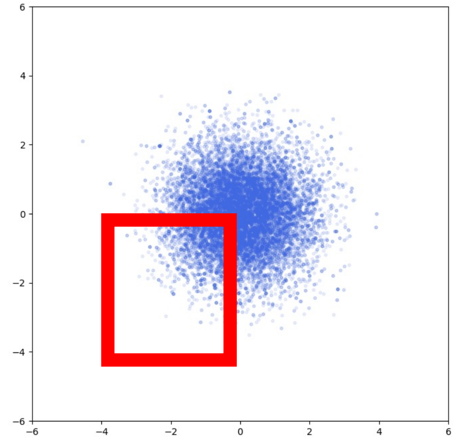




Markov Chain Monte Carlo (MCMC)

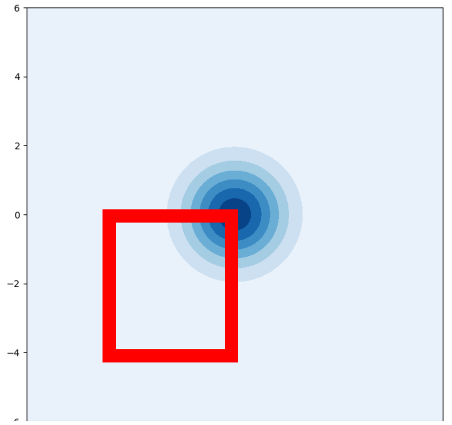
- MCMC are a class methods that approximate a probability distribution with a set of states a.k.a. samples.
- Sampling is unbiased if #draws from each region is proportional to the probability mass of that region.

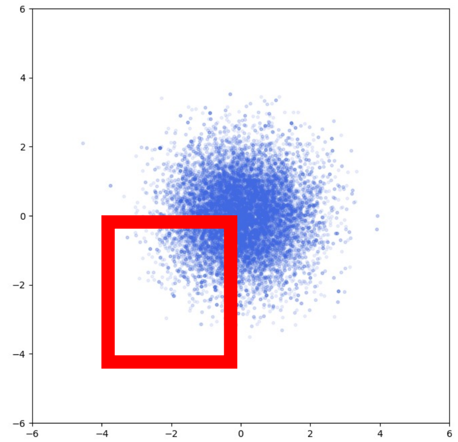




Markov Chain Monte Carlo (MCMC)

- MCMC are a class methods that approximate a probability distribution with a set of states a.k.a. samples.
- Sampling is unbiased if #draws from each region is proportional to the probability mass of that region.
- The state of the art MCMC methods (e.g. NUTS) are variations of the Hamiltonian Monte Carlo (HMC)

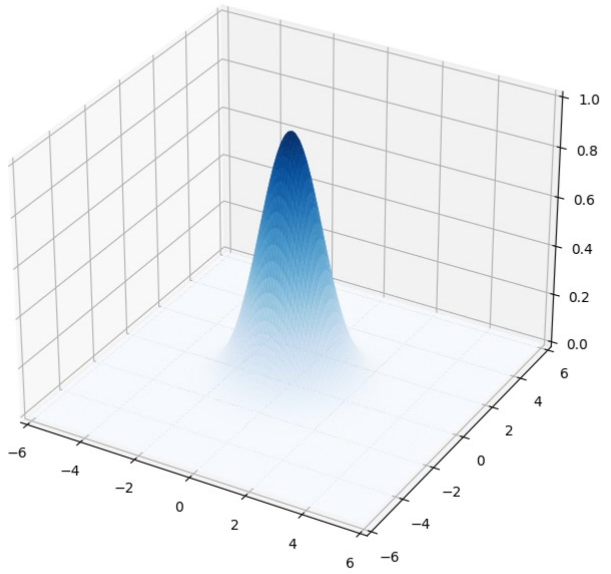




Hamiltonian Monte Carlo (HMC)

- In HMC, the potential energy is defined as the -log of PDF.

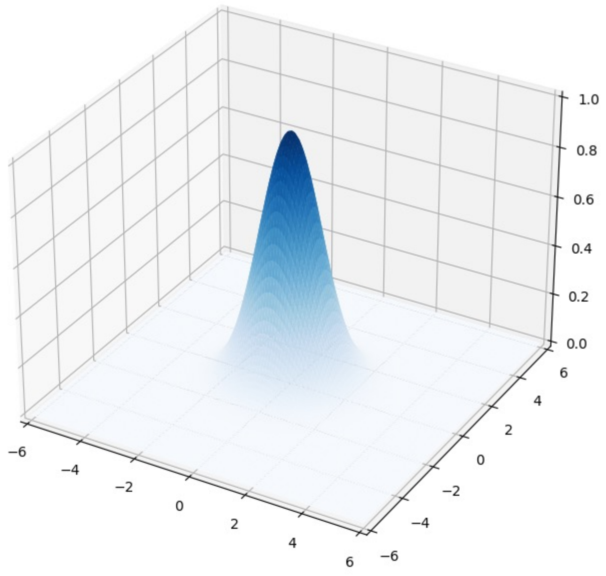
Target probability density: $\pi(\boldsymbol{q})$



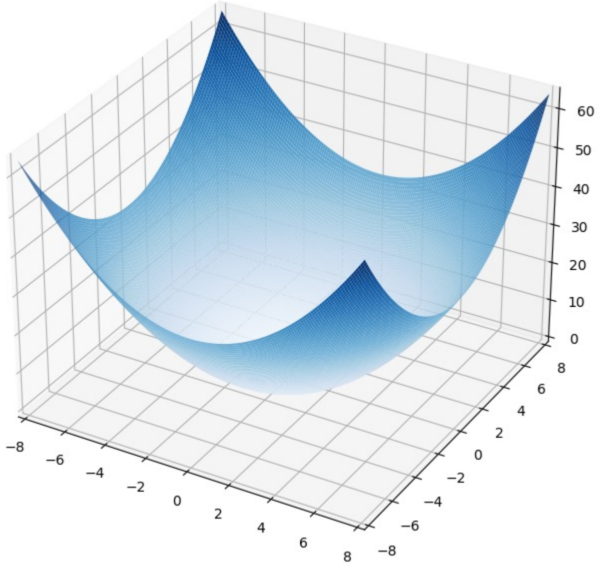
Hamiltonian Monte Carlo (HMC)

- In HMC, the potential energy is defined as the -log of the PDF.

Target probability density: $\pi(\boldsymbol{q})$

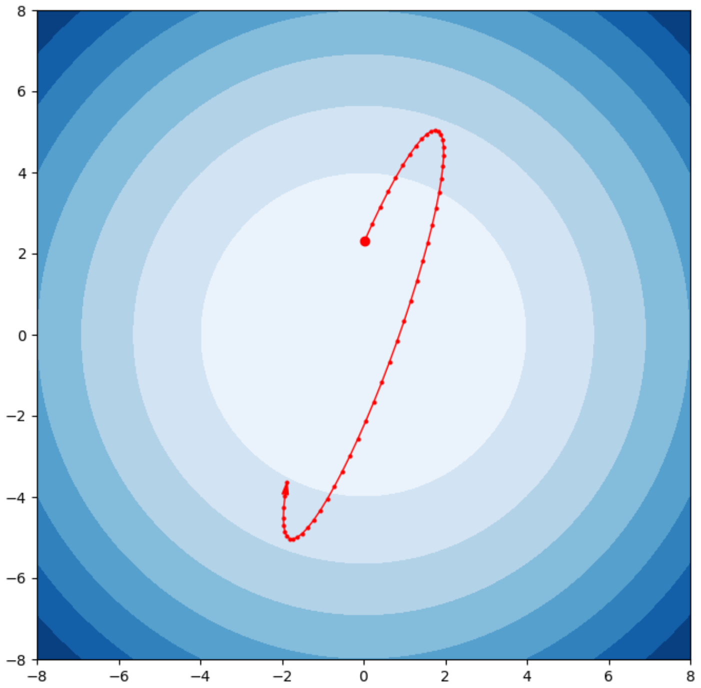


Potential energy: $\mathrm{U}(\boldsymbol{q}) = -\log \pi(\boldsymbol{q})$



Hamiltonian Monte Carlo (HMC)

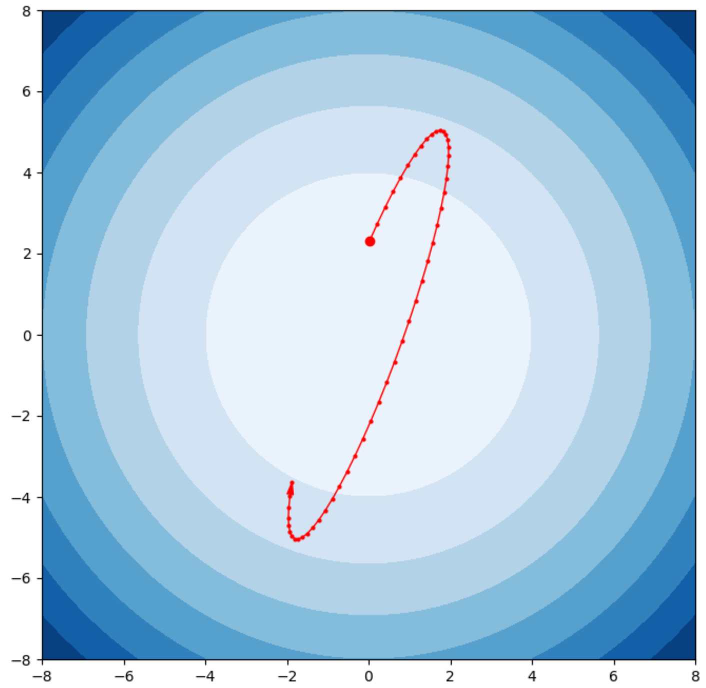
- In HMC, the potential energy is defined as the neg. log of the PDF.
- An auxiliary momentum $\boldsymbol{p} \sim N(0, I)$ is assigned to the current state. Then, it is evolved via the simulation of the equations of motion for a fixed time.



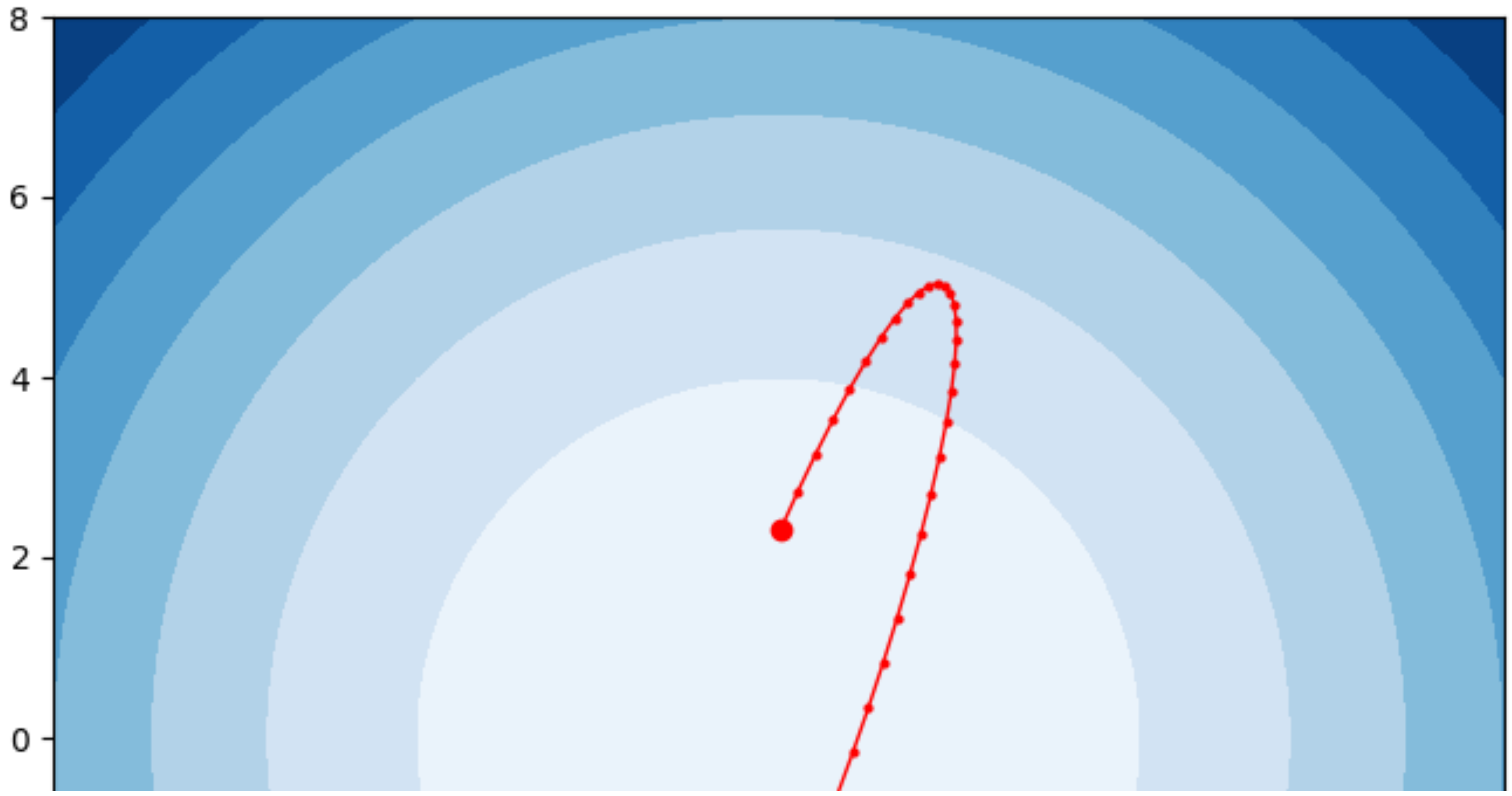
Hamiltonian Monte Carlo (HMC)

- In HMC, the potential energy is defined as the neg. log of PDF.

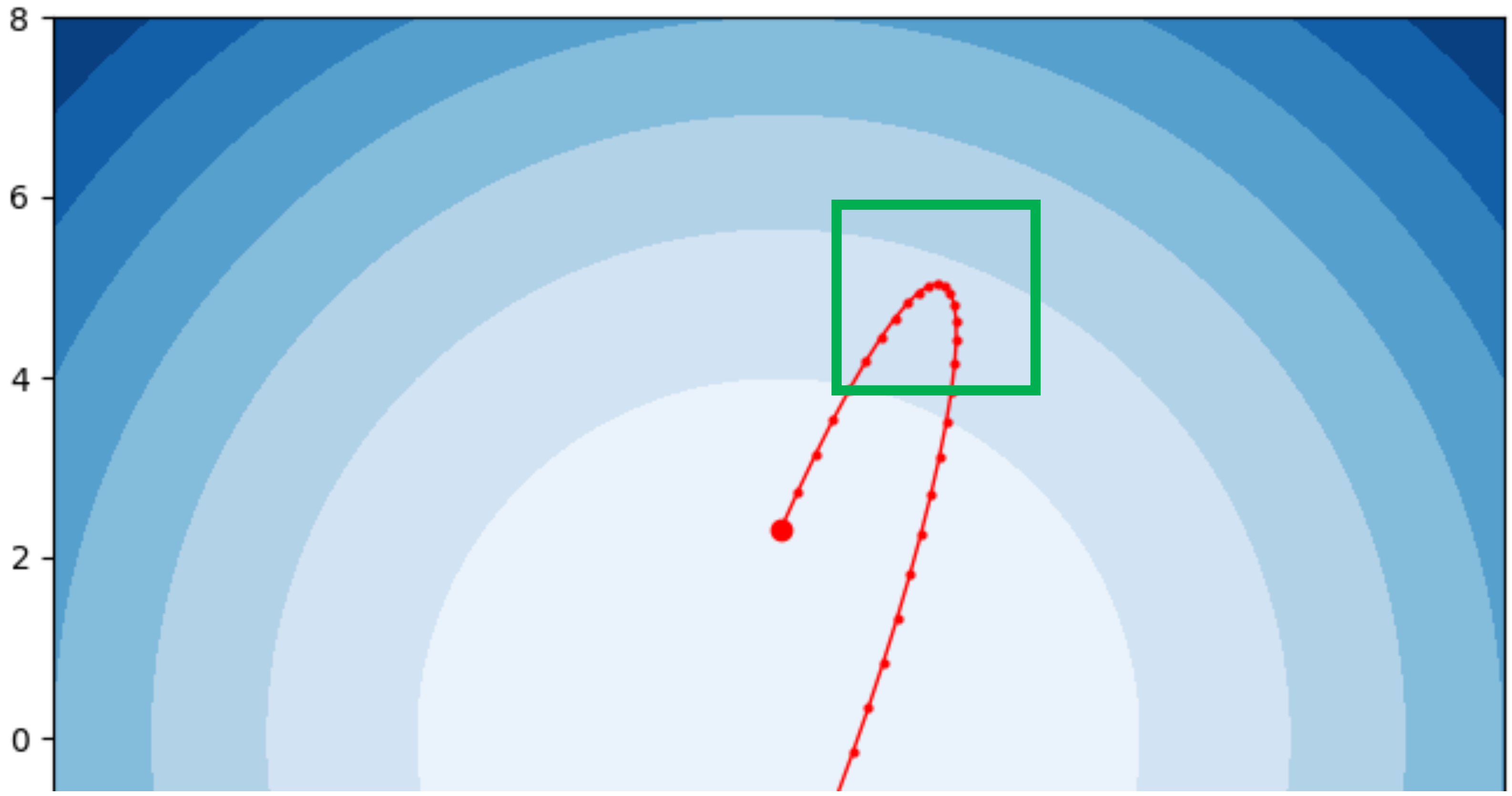
- An auxiliary momentum $\boldsymbol{p} \sim N(0, I)$ is assigned to the current state. Then, it is evolved via the simulation of the equations of motion for a fixed time.
- If the simulation is precise, the acceptance probability of the resulting proposal is high.



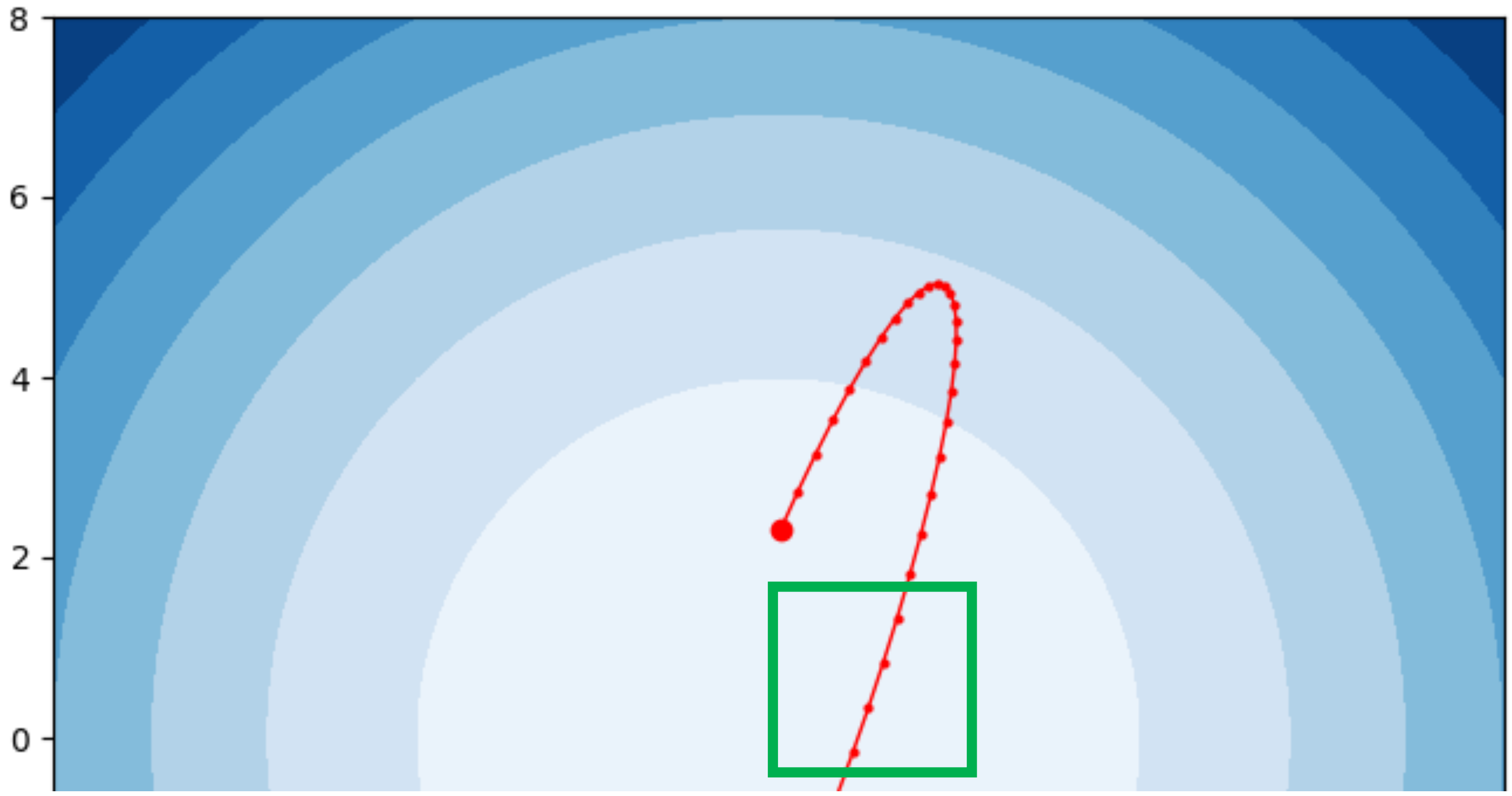
Note: In high potential (i.e. low probability) regions the velocity is low



More states are generated from these regions…

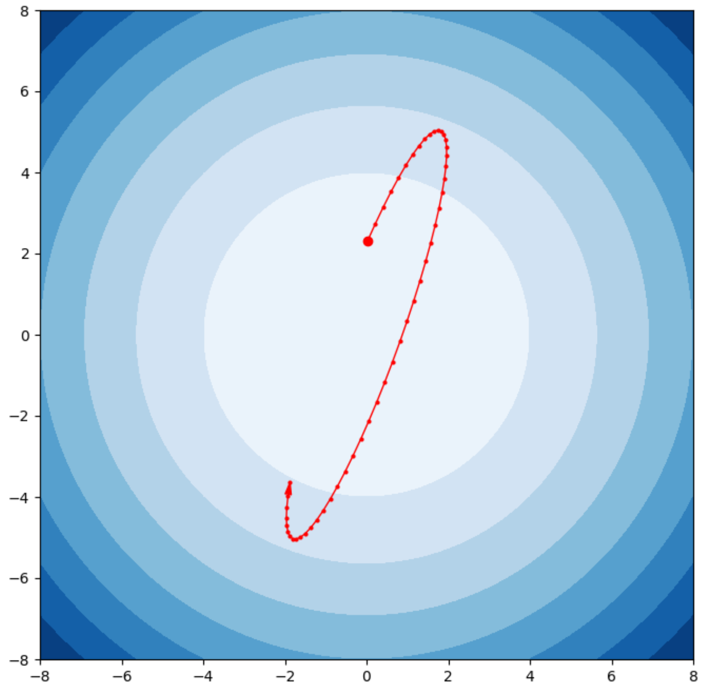


Conversely, less time is spent in the high-probability regions



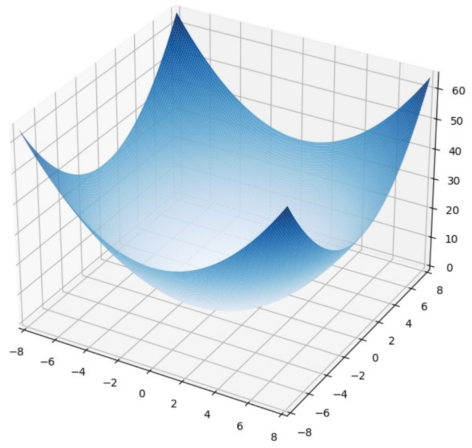
Evolution bias

- As such, there is bias towards proposing low-probability states which is counter-intuitive!



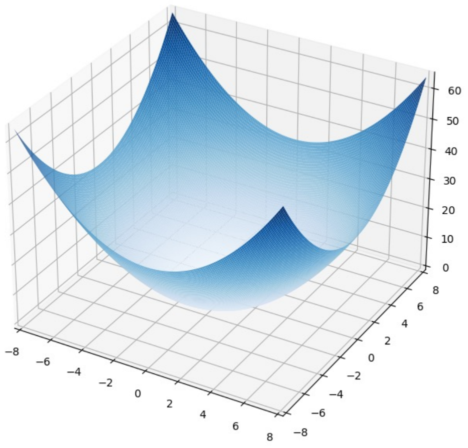
To verify …

Consider a potential energy function:



To verify …

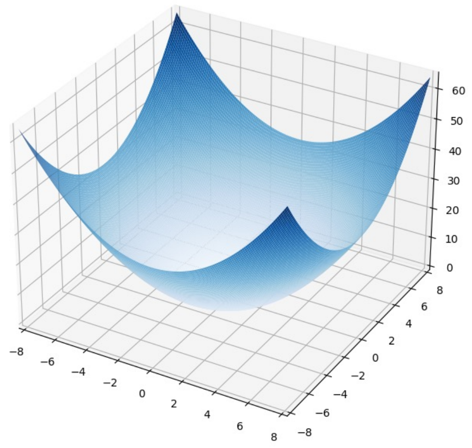
Consider a potential energy function:



We let the current state be on the unite circle

To verify …

Consider a potential energy function:

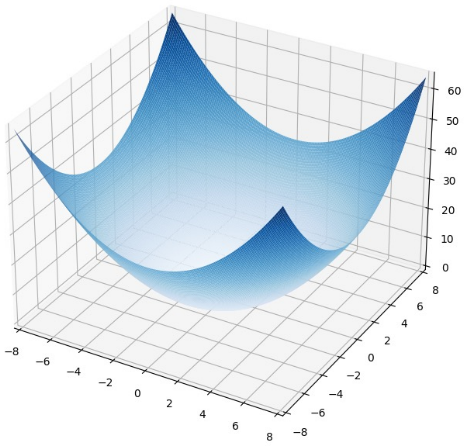


We let the current state be on the unit circle

We fix the momentum magnitude but randomly choose its direction

To verify …

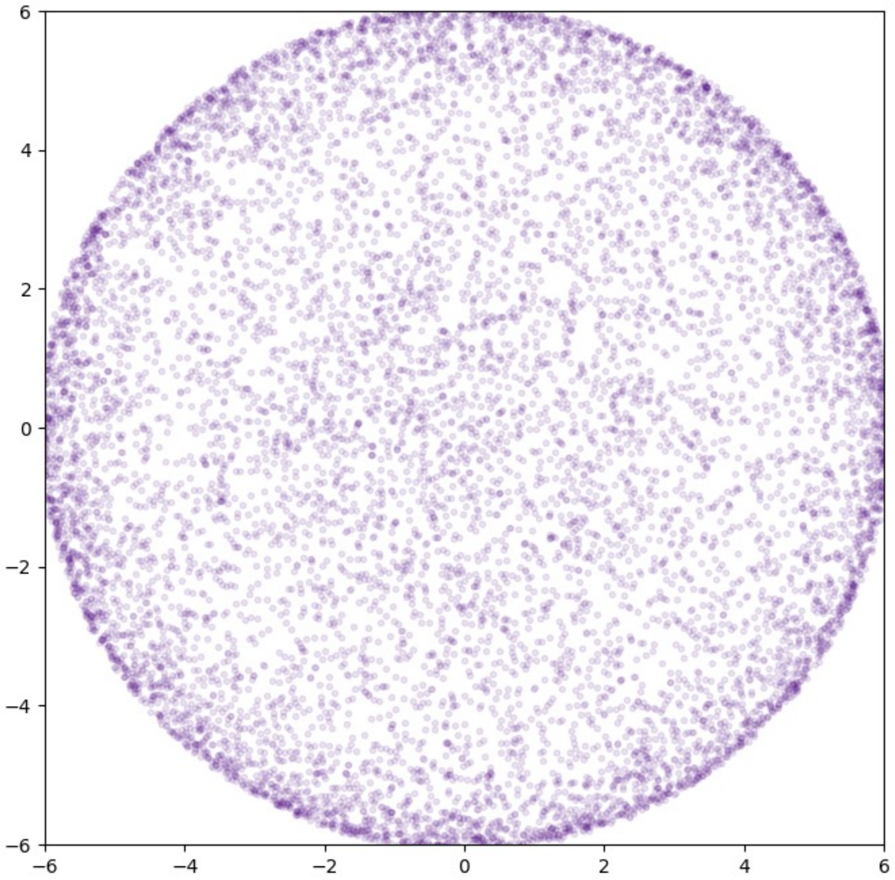
Consider a potential energy function:



We let the current state be on the unit circle

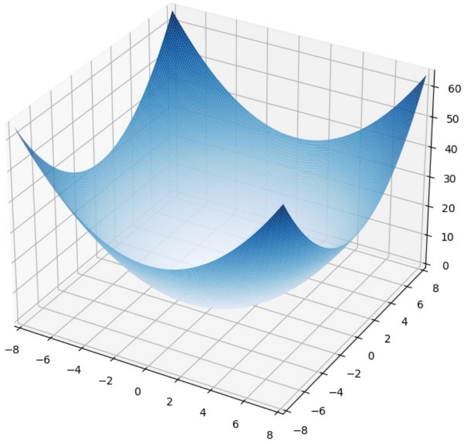
We fix the momentum magnitude but randomly choose its direction

… and plot the proposals:



To verify …

Consider a potential energy function:



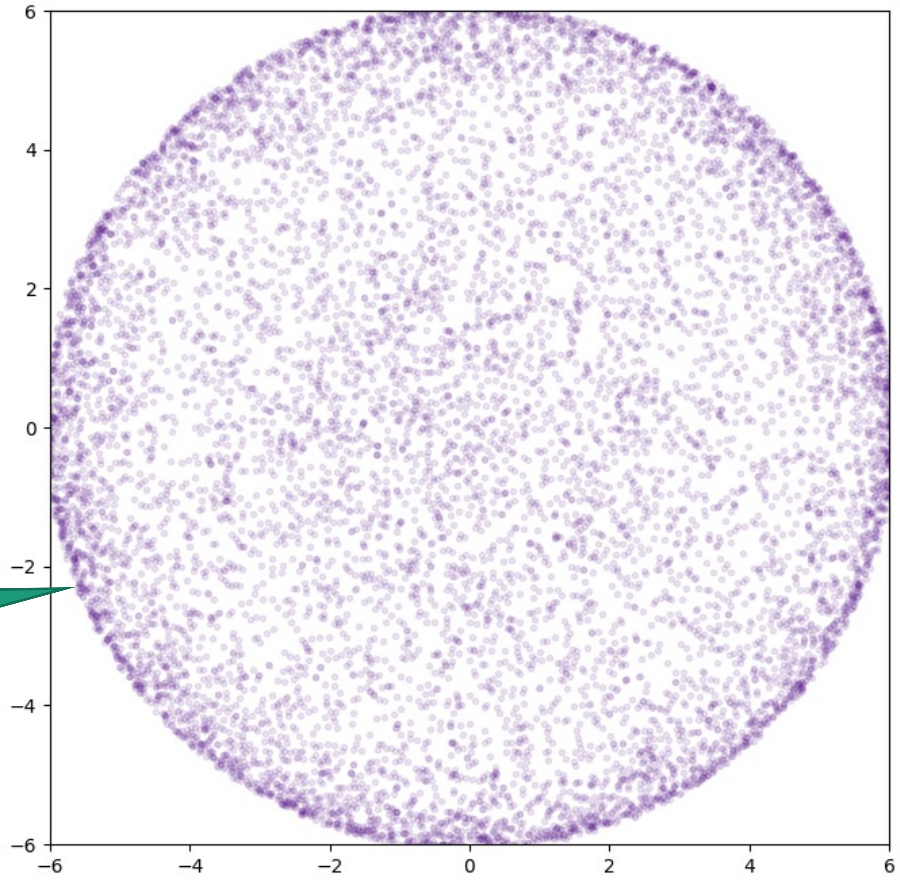
We let the current state be on the unit circle

We fix the m

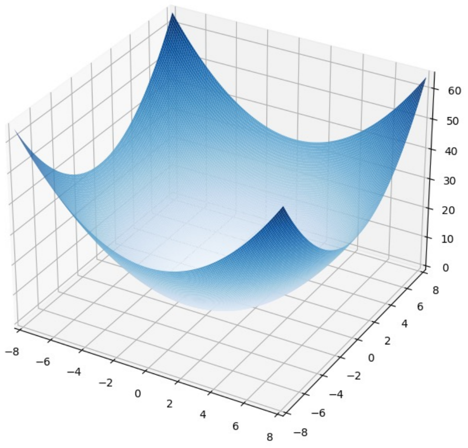
but randoml

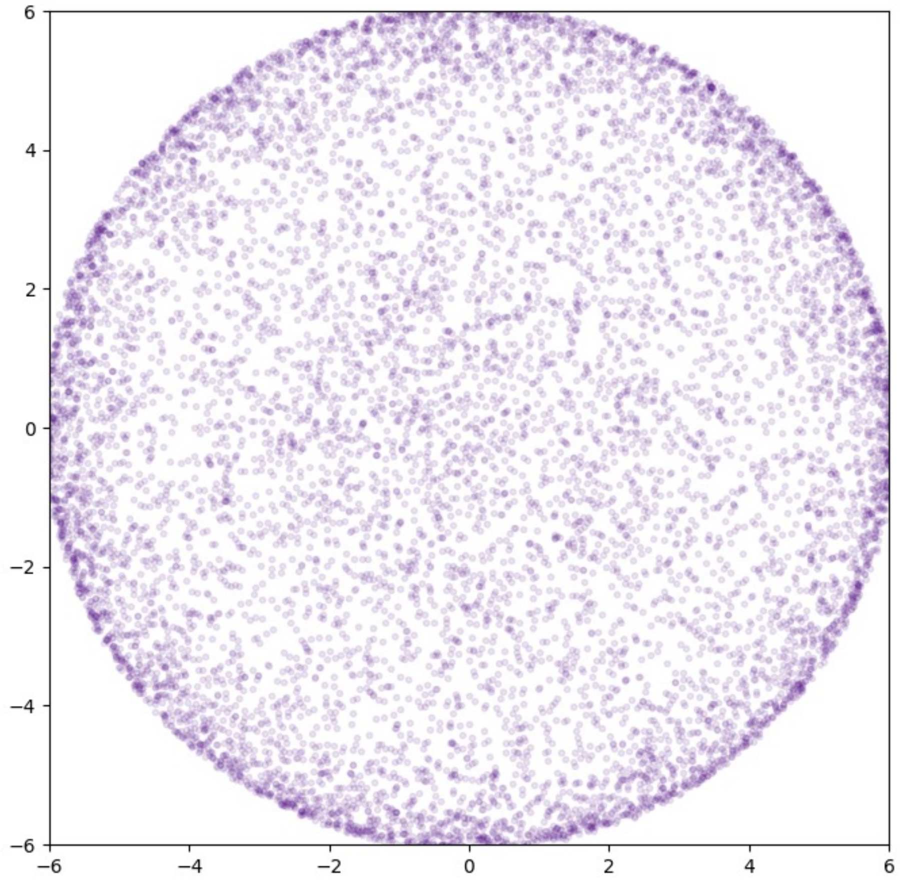
… and plot th

Regions of low probability are over-represented

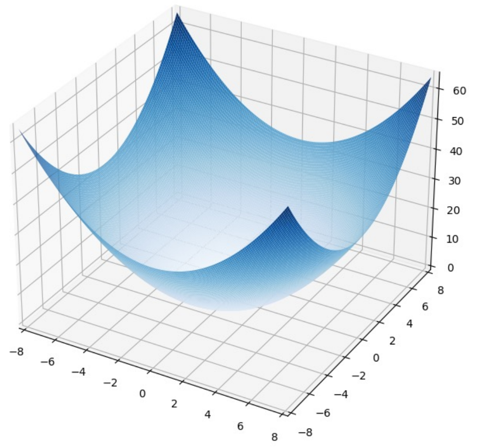


Then, what makes the total HMC sampling unbiased?

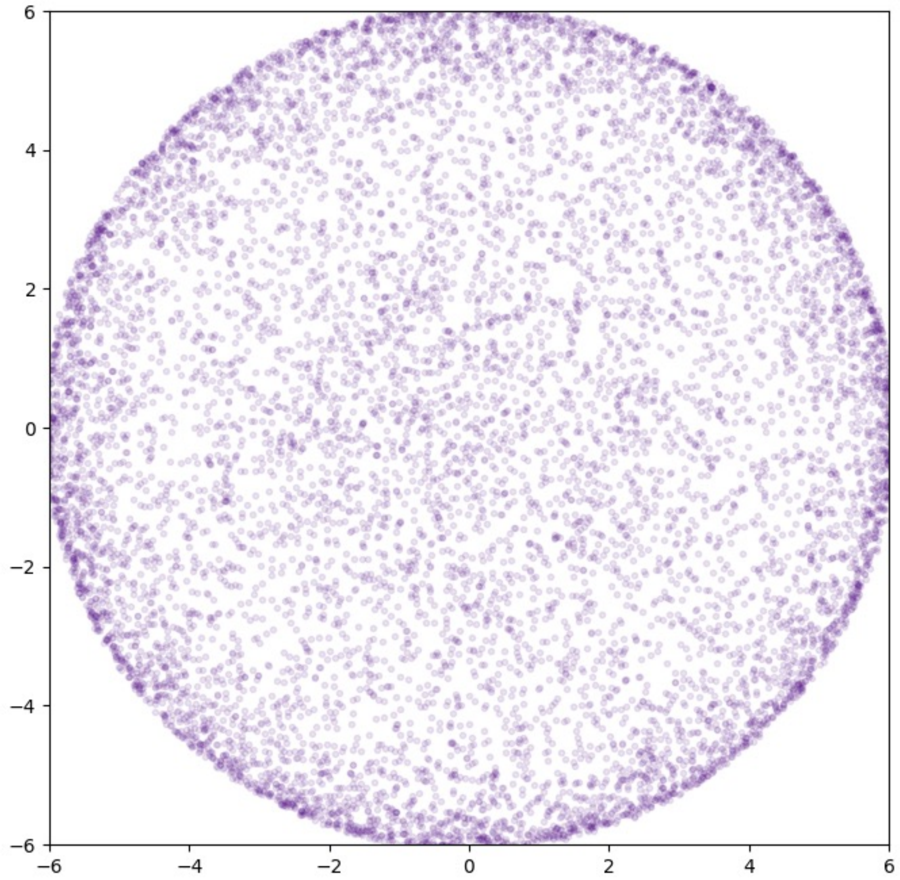




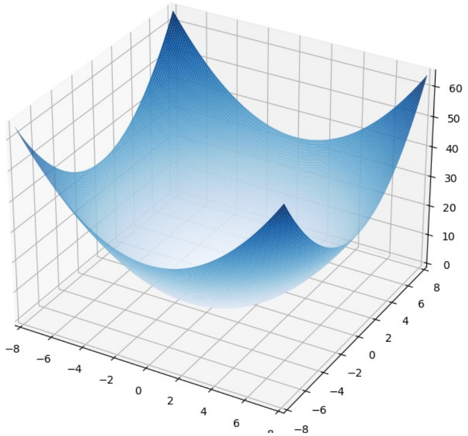
Then, what makes the total HMC sampling unbiased?



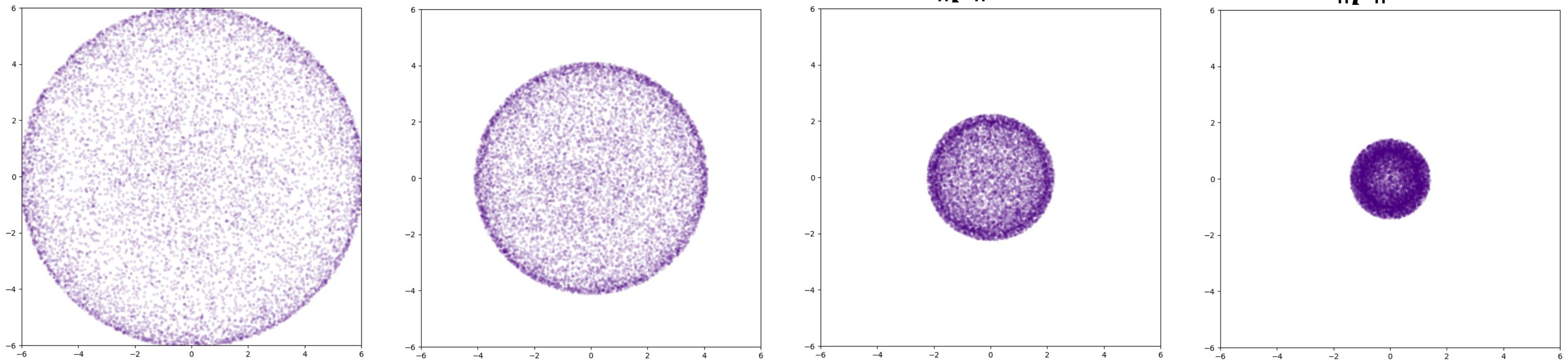
The momentum distribution…



Previous example with different momentum magnitudes…

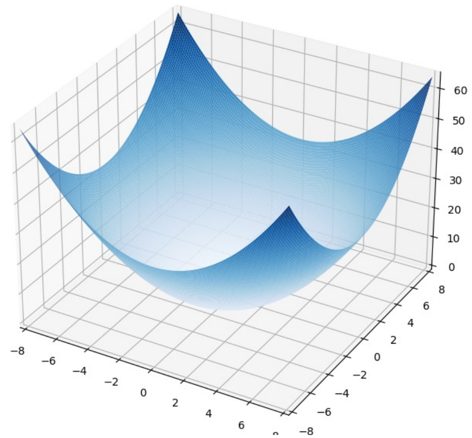


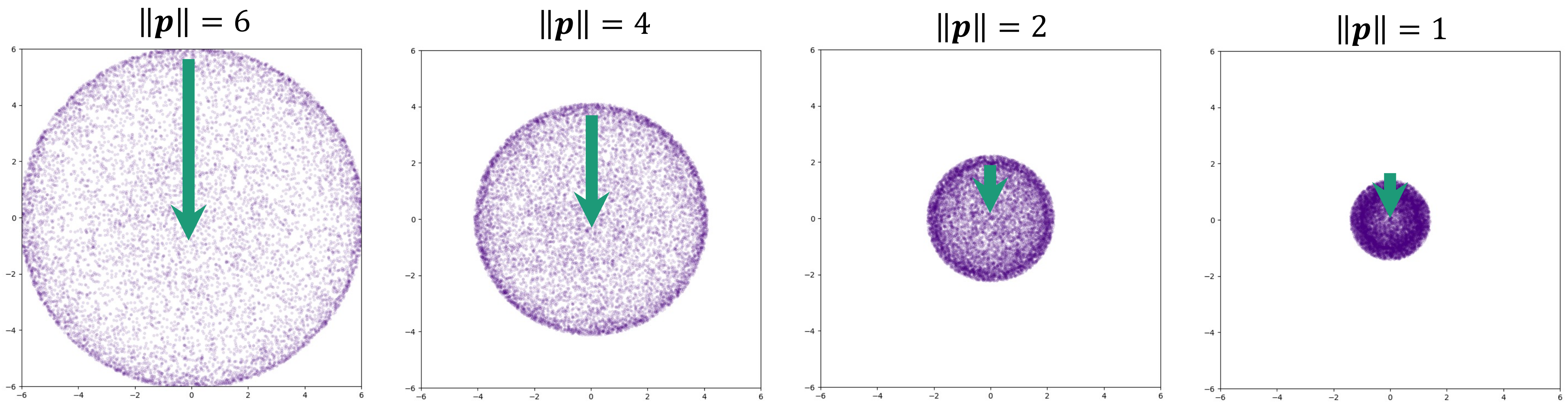
$\|\boldsymbol{p}\| = 6$ $\quad\|\boldsymbol{p}\| = 4$ $\quad\|\boldsymbol{p}\| = 2$ $\quad\|\boldsymbol{p}\| = 1$



Previous example with different momentum magnitudes…

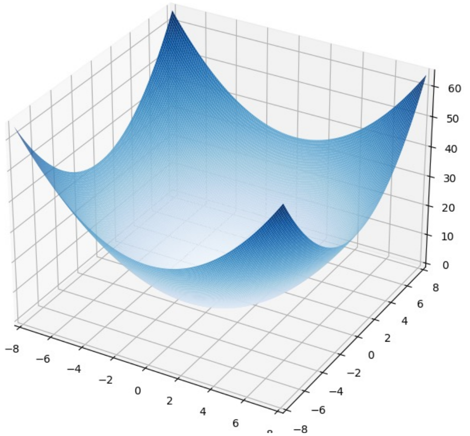
The transition from a low to high probability region is always possible



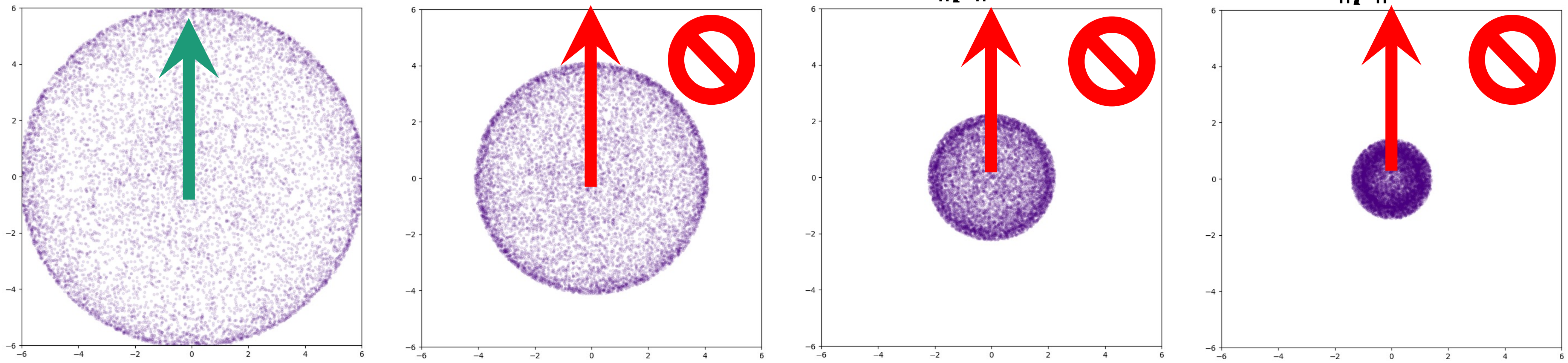


Previous example with different momentum magnitudes…

But if the momentum's magnitude is low, the transition from a high to low probability region may NOT be possible.



$\|\boldsymbol{p}\| = 6$ $\quad \|\boldsymbol{p}\| = 4$ $\quad \|\boldsymbol{p}\| = 2$ $\quad \|\boldsymbol{p}\| = 1$



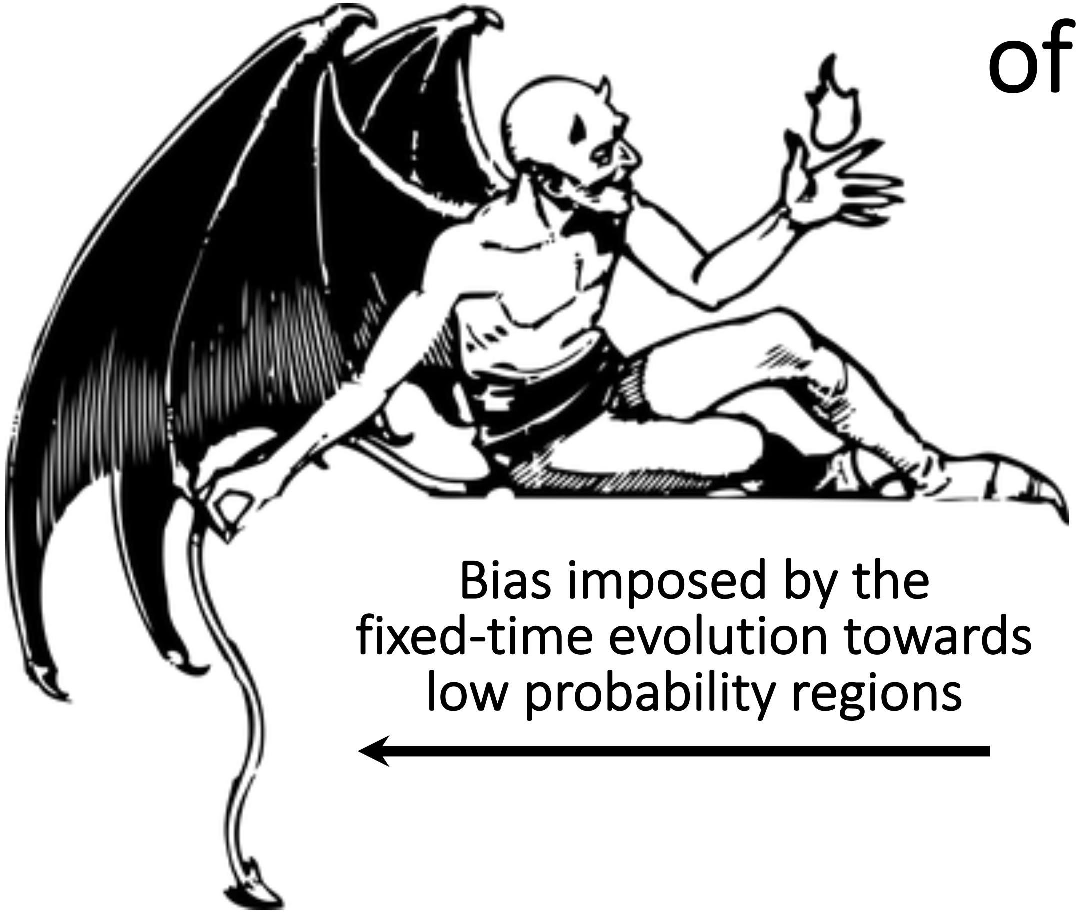
Two counter-balancing biases…

Convergence
of HMC

The Picture is downloaded from pixabay.com

Two counter-balancing biases…

Convergence of HMC

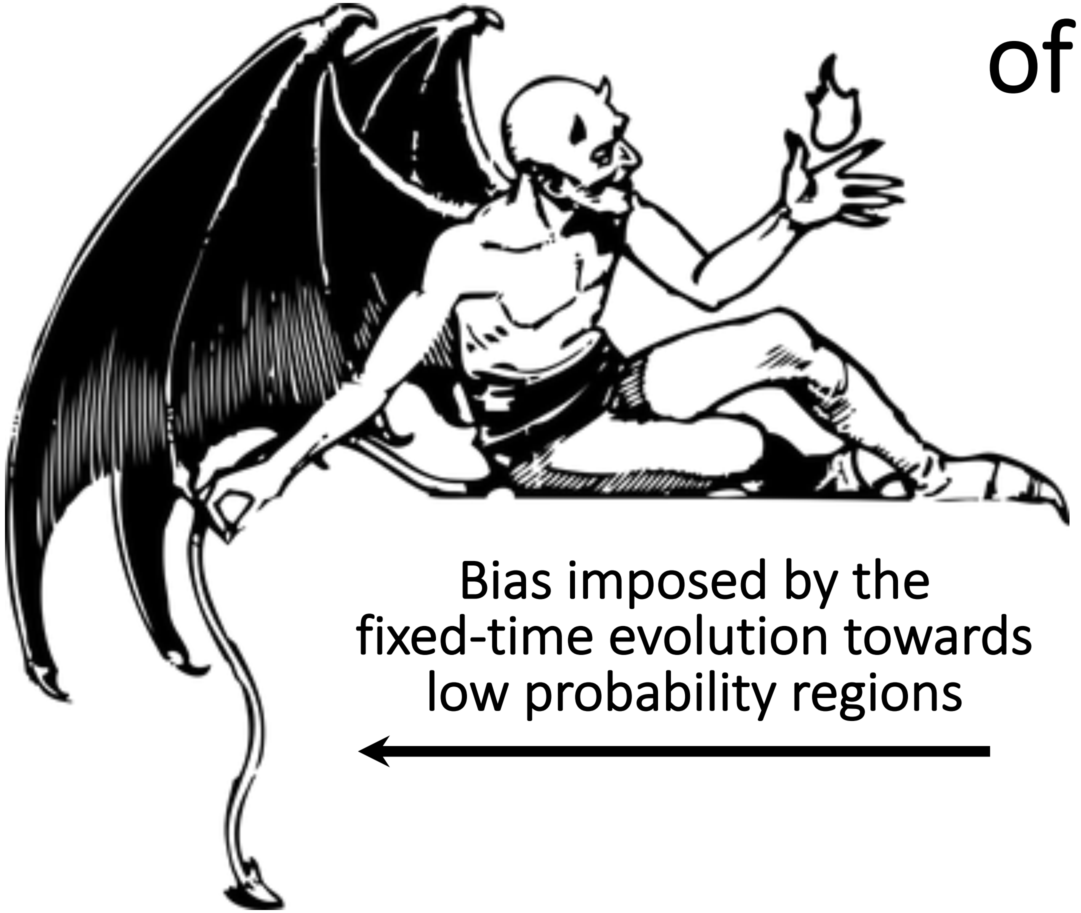


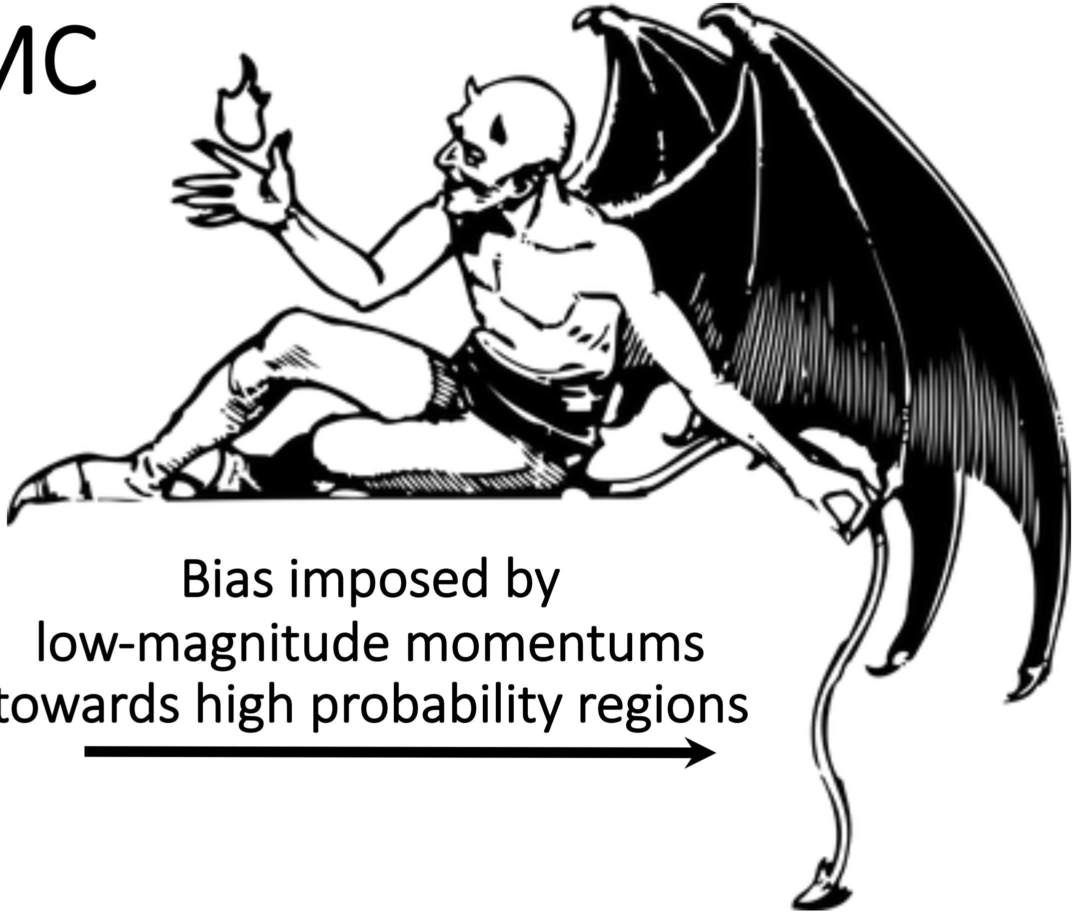
Bias imposed by the fixed-time evolution towards low probability regions

The Picture is downloaded from pixabay.com

Two counter-balancing biases…

Convergence
of HMC



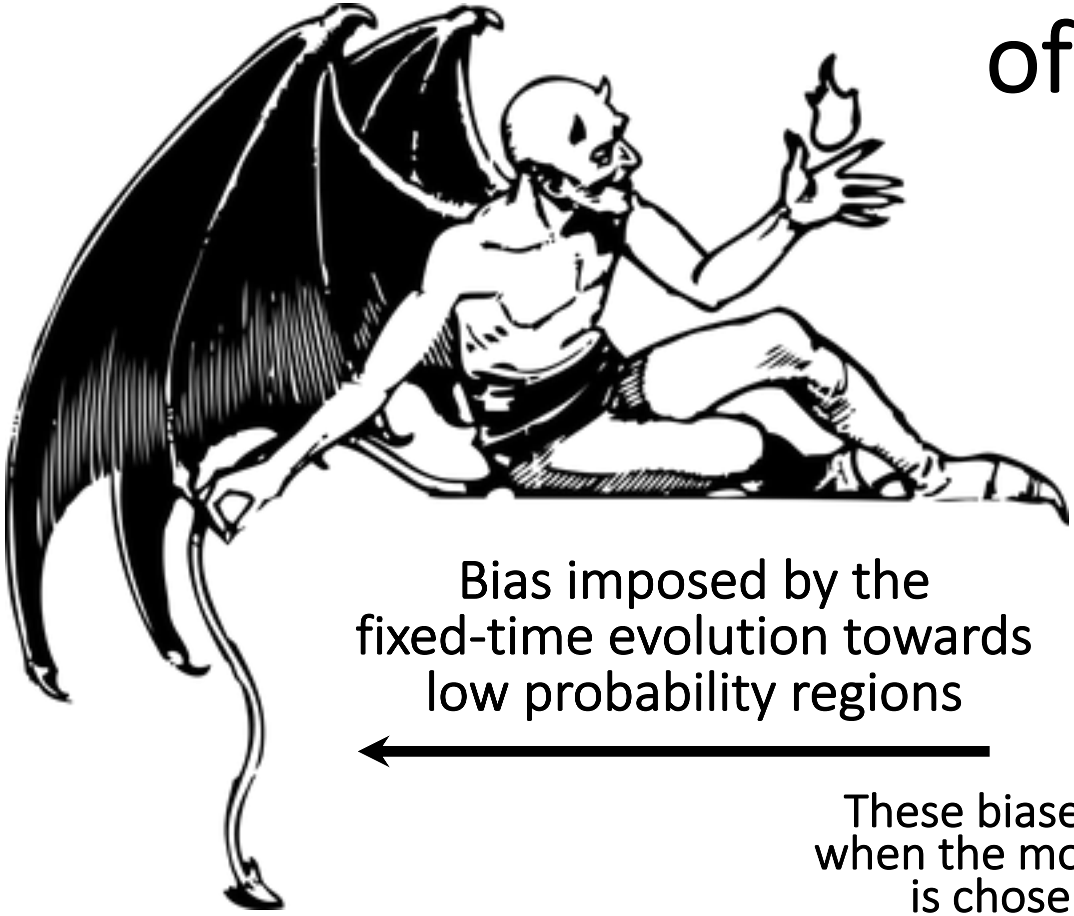


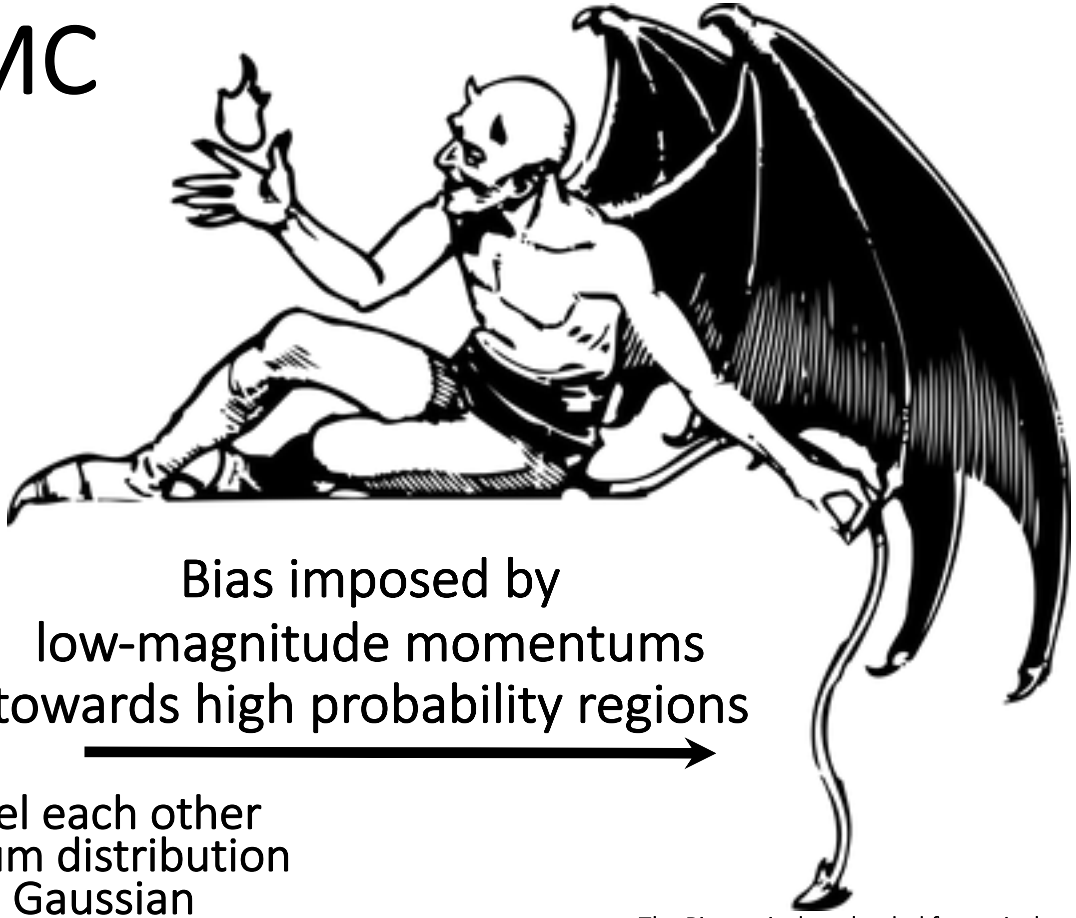
Bias imposed by the
fixed-time evolution towards
low probability regions

Bias imposed by
low-magnitude momentums
towards high probability regions

Two counter-balancing biases…

Convergence of HMC





Bias imposed by the fixed-time evolution towards low probability regions

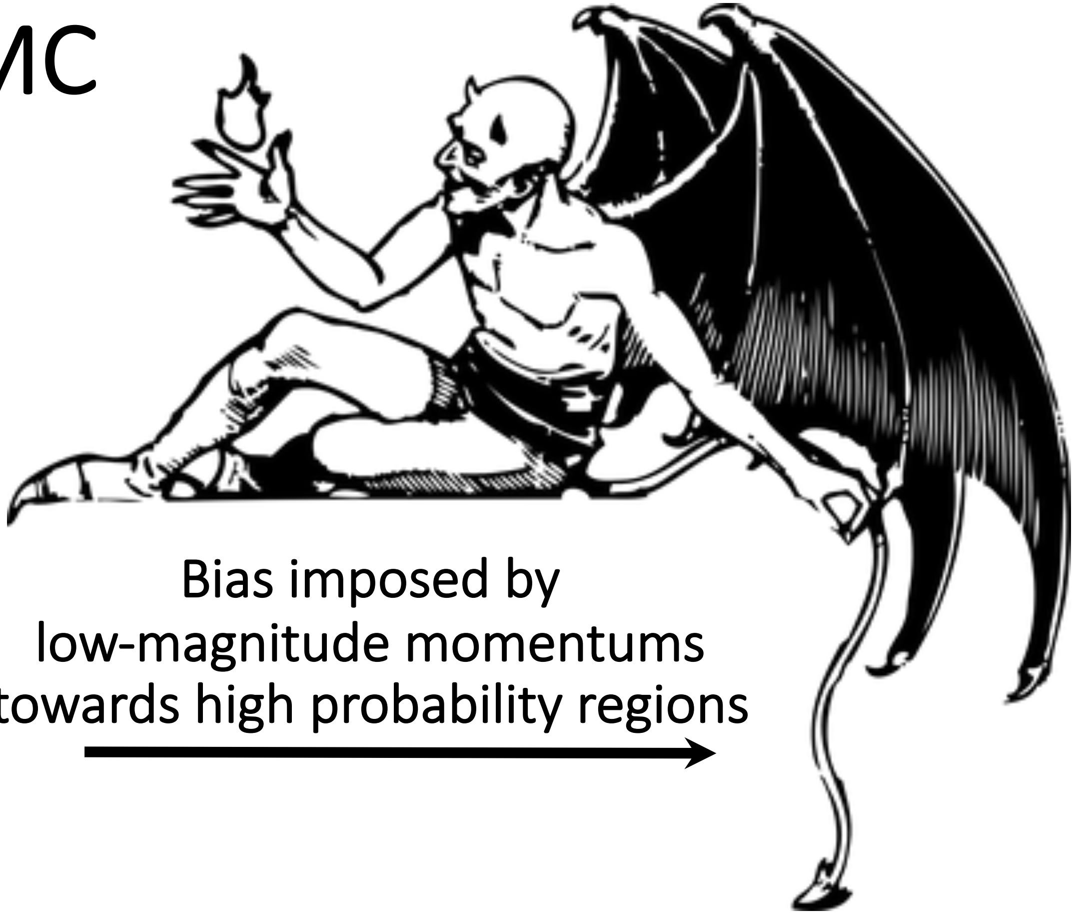
Bias imposed by low-magnitude momentums towards high probability regions

These biases cancel each other when the momentum distribution is chosen to be Gaussian

Resolving the first bias

Convergence
of HMC

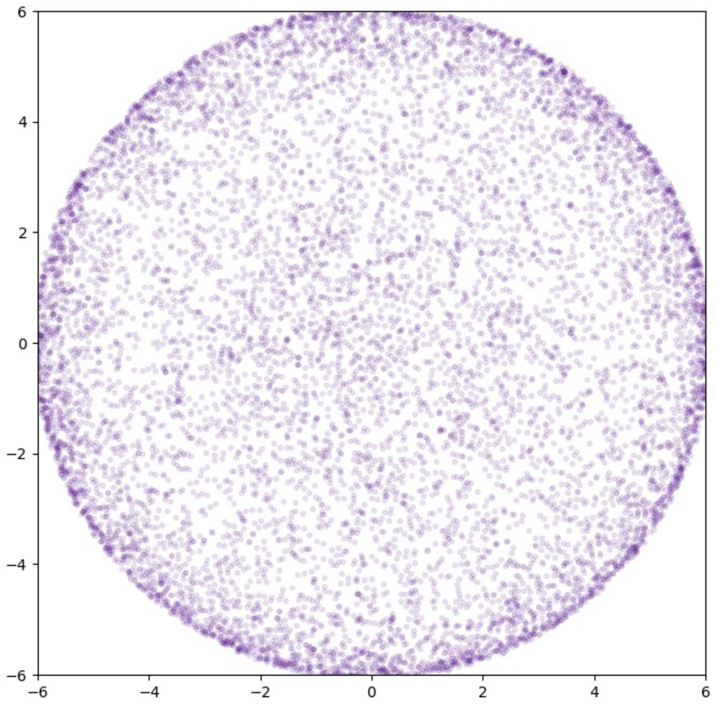
Fixed distance
(rather than fixed-time)
evolution



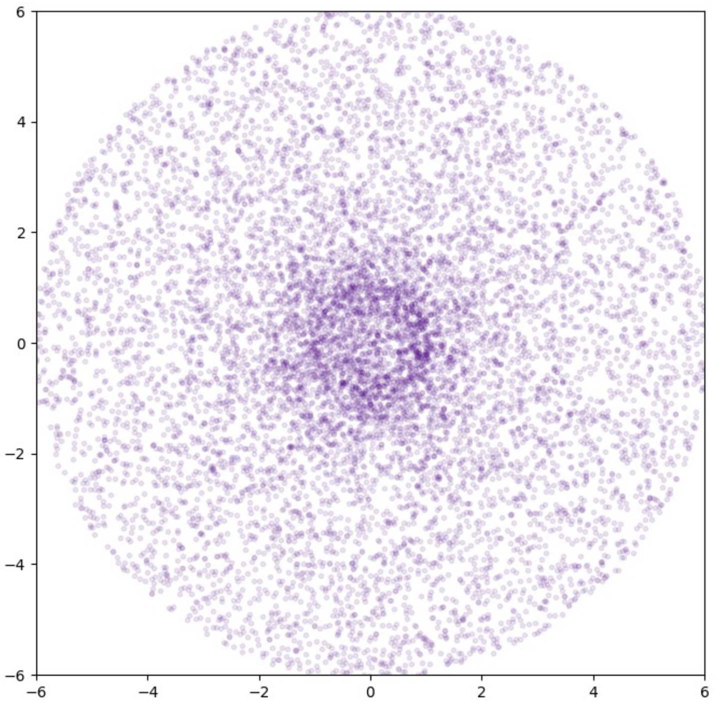
Bias imposed by
low-magnitude momentums
towards high probability regions

The Picture is downloaded from pixabay.com

Fixed-distance simulation of the equations of motion

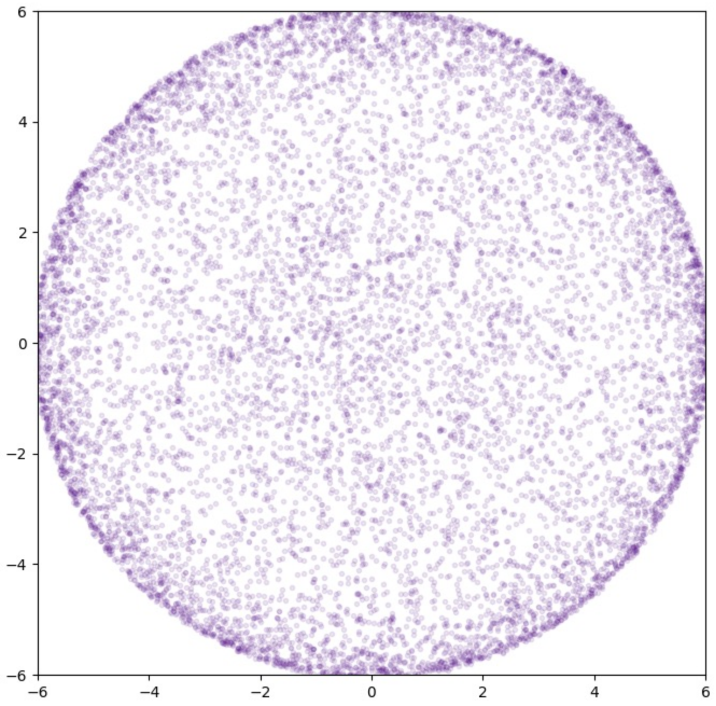


Fixed-time evolution

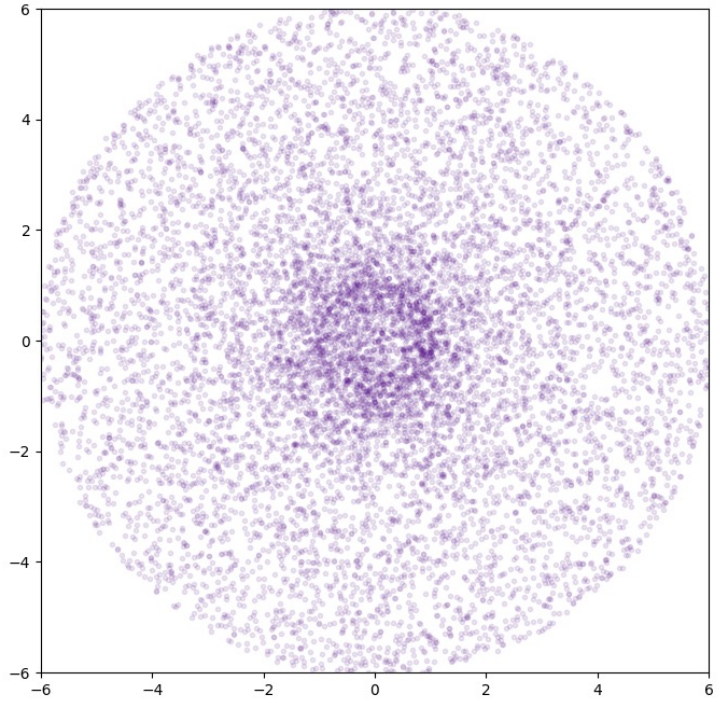


Fixed-distance evolution

Fixed-distance simulation of the equations of motion



Fixed-time evolution

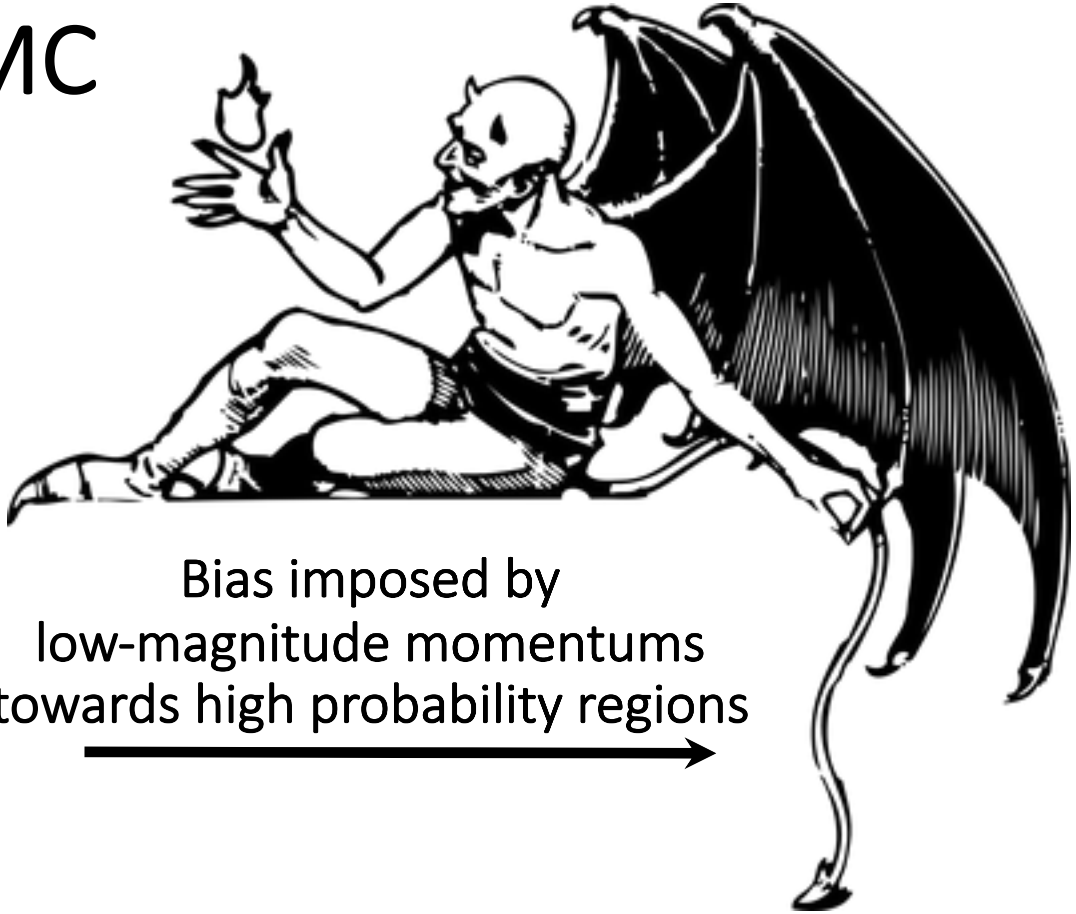


Fixed-distance evolution

Even though more time is spent in low-probability regions, more distance is not traversed there

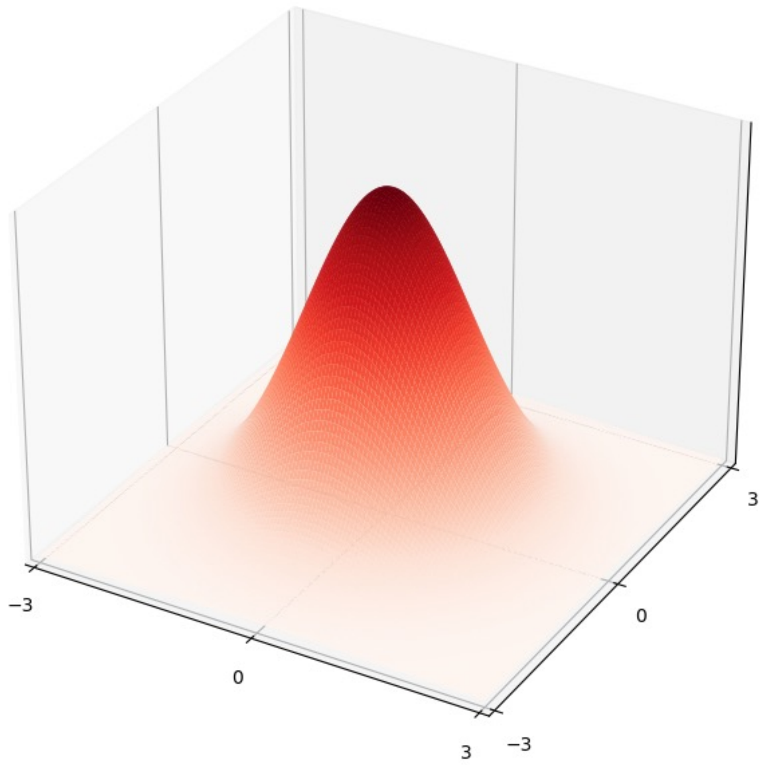
To resolve the second bias…

Convergence of HMC

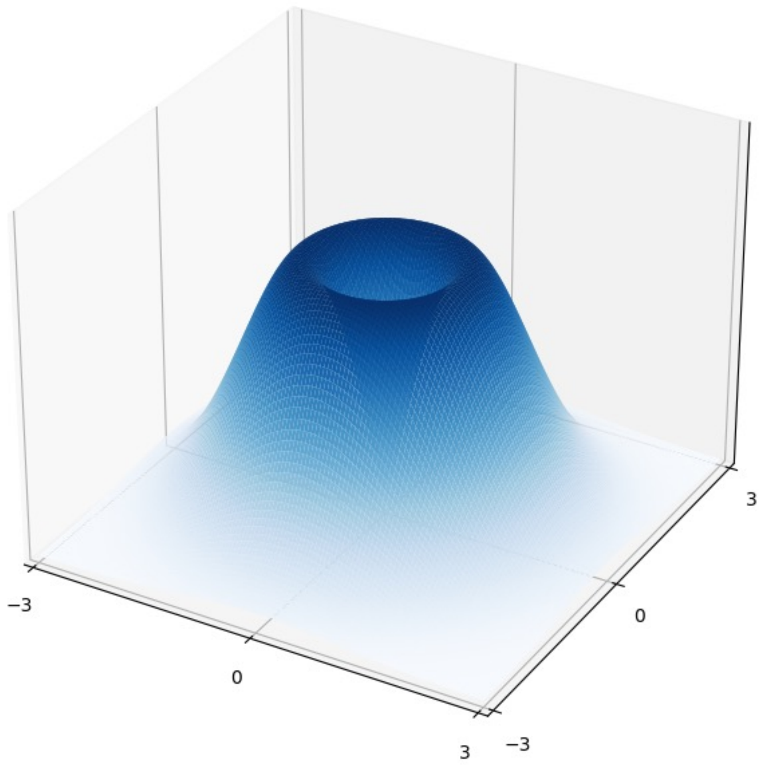


Bias imposed by
low-magnitude momentums
towards high probability regions

We prove...



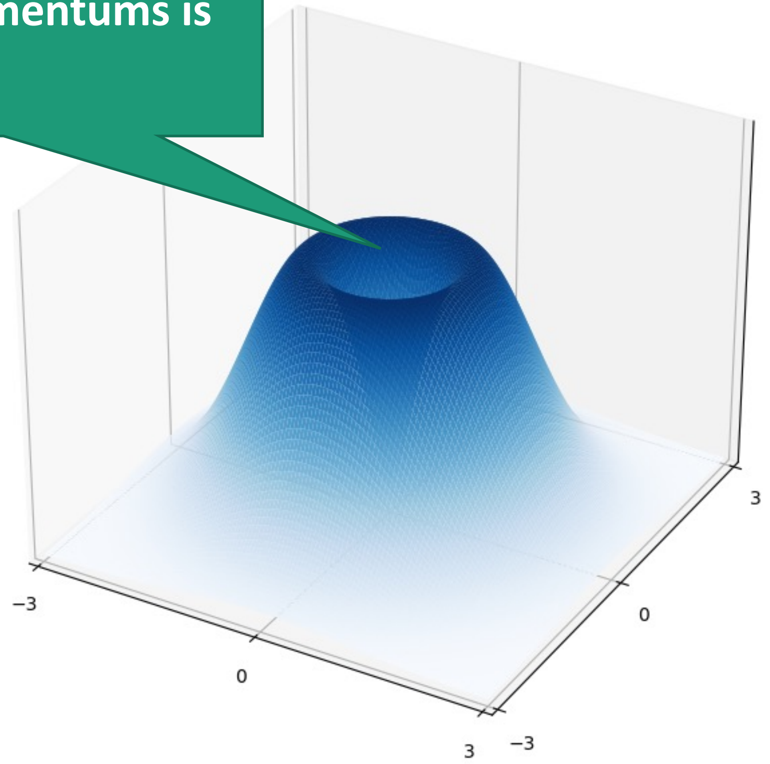
HMC's momentum distribution



FDHMC's momentum distribution

FDHMC’s momentum distribution

Probability assigned to low-magnitude momentums is low

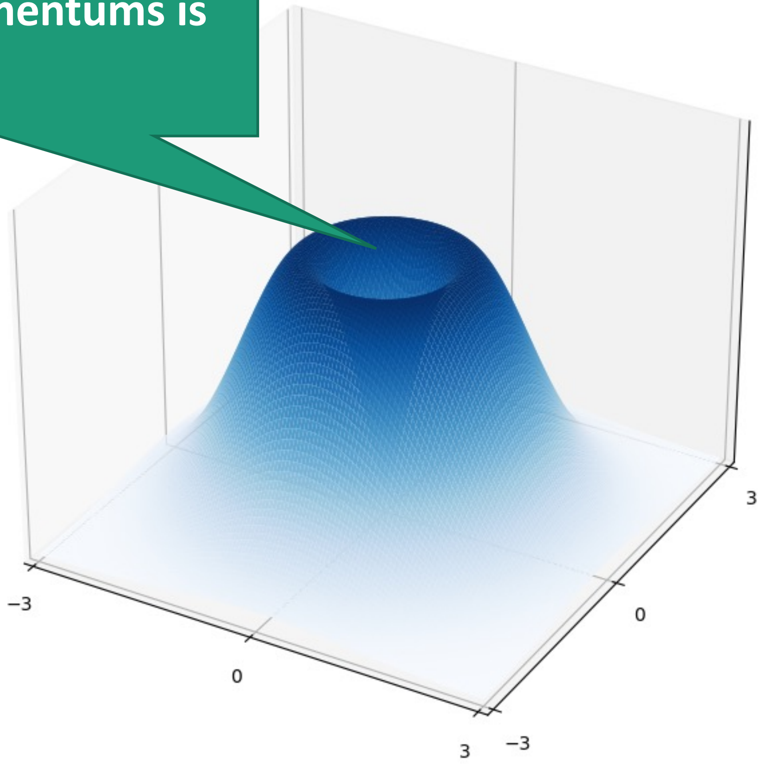


FDHMC’s momentum distribution

FDHMC’s momentum distribution

Probability assigned to low-magnitude momentums is low

The expected magnitude of FDHMC’s momentum vector is higher

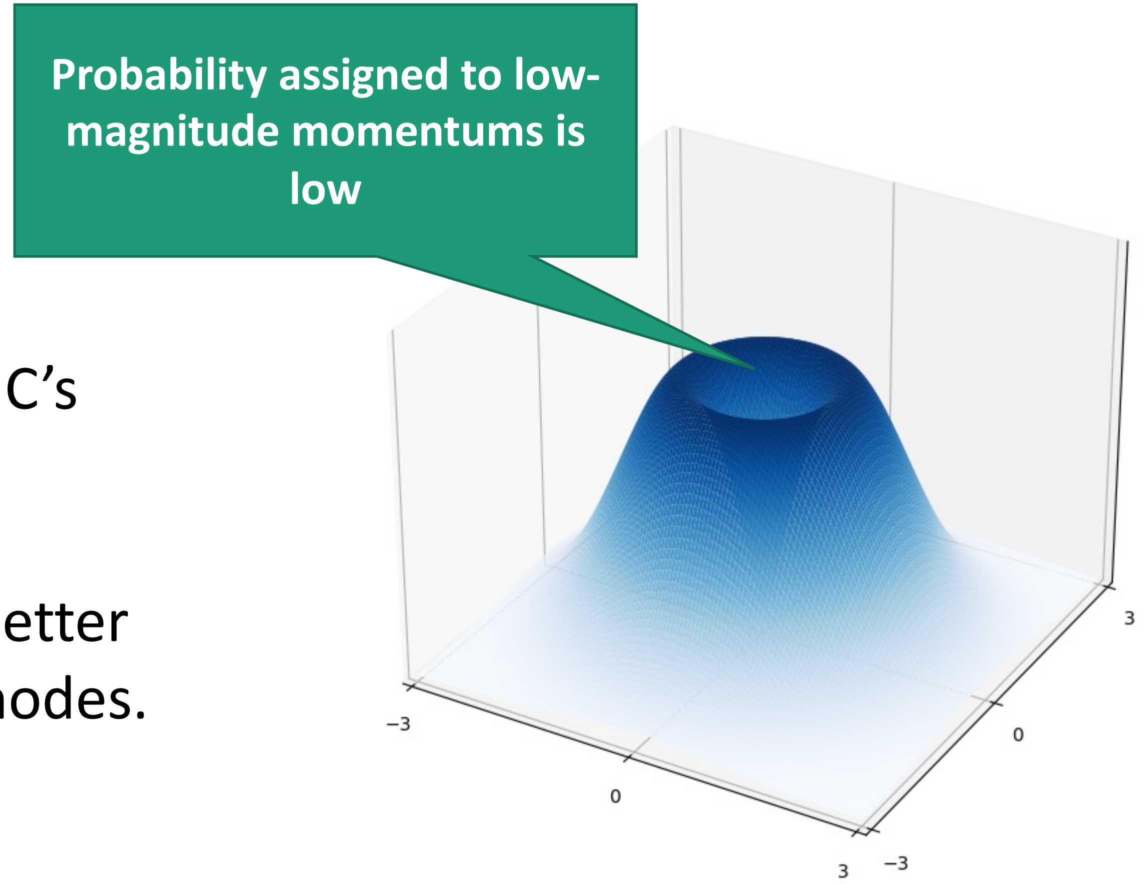


FDHMC’s momentum distribution

FDHMC's momentum distribution

The expected magnitude of FDHMC's momentum vector is higher.

This translates to higher ESS and better exploration of target probability modes.



FDHMC's momentum distribution

Conclusion

- With negligible computational overhead, we resolved two counter-balancing biases that exist in the core of HMC algorithms.

Conclusion

- With negligible computational overhead, we resolved two counter-balancing biases that exist in the core of HMC algorithms.
- Our experiments show that the resulting FDHMC has a higher ESS/grad.